Università di Genova

ISTITUTO ITALIANO DI TECNOLOGIA
ADVANCED ROBOTICS

UNIVERSITY OF GENOVA

PHD PROGRAM IN BIOENGINEERING AND ROBOTICS

Data Processing using Machine Learning techniques for Fall Detection and Prevention

by

Nicholas Cartocci

Thesis submitted for the degree of *Doctor of Philosophy* (38° cycle)

February 2026

Dr. Antonios E. Gkikakis	Supervisor
Dr. Jesús Ortiz	Supervisor
Prof. Darwin G. Caldwell	Supervisor
Prof. Paolo Massobrio	Coordinator of the PhD program

Thesis Jury:

Prof. Alessandro M. De Nunzio, *LUNEX University of Applied Sciences*	External examiner
Prof. Christopher P. Paolini, *San Diego State University*	External examiner

Department of Informatics, Bioengineering, Robotics and Systems Engineering

I would like to dedicate this thesis to my FAMILY.

Declaration

I hereby declare that, except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this or any other university. This dissertation is my own work and contains nothing that is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgments. During the research and drafting process of this thesis, generative AI tools were utilized for brainstorming, summarizing literature reviews, and initial outlining. All final content, analysis, and conclusions are the original work of the author and have been independently verified. This dissertation contains fewer than 100,000 words, including appendices, bibliography, footnotes, tables, and equations, and has fewer than 60 figures.

Nicholas Cartocci
February 2026

Abstract

Falls constitute a major global health and occupational safety crisis, resulting in millions of injuries and fatalities annually. Traditional fall safety measures are reactive, relying on kinematic data to detect an event only after it has occurred, which fails to enable timely injury mitigation. This PhD thesis develops a comprehensive, machine learning framework that transitions fall safety from reactive reporting to continuous, proactive prediction, leveraging multimodal biosensing and personalized data processing techniques.

The research establishes a three-part framework: Reactive Detection, Proactive Impact Prediction, and Predictive Risk Estimation. First, a Deep Neural Network (DNN) achieved a high True Negative Rate of 94.4% and a competitive True Positive Rate of 82.6% on the SisFall dataset, confirming robust fall detection capabilities while highlighting the limitations of kinematic data alone, particularly in detecting subtle falls.

Second, to enable injury mitigation systems, such as wearable airbags, the system integrated Proactive Impact Prediction. A novel Kolmogorov–Arnold Network (Fall-KAN) was introduced, achieving a superior mean Root Mean Squared Error of approximately 159 milliseconds for Time-to-Impact estimation, demonstrating the feasibility of millisecond-scale prediction. Alternatively, a Deep Neural Network architecture trained on an artificially enriched dataset generated using the Inverted Pendulum was tested, which forecasted the fall trajectory and provided a crucial preventive time window of approximately 200 milliseconds.

Third, the thesis addresses the root causes of falls through Predictive Risk Estimation, utilizing multimodal physiological data collected during Activities of Daily Living and high-stress Virtual Reality fall simulations. Crucially, analysis revealed that generalized Machine Learning models failed to reliably classify fall risk across different subjects due to high inter-subject variability. Consequently, the framework shifted to a personalized classification approach, which achieved highly robust performance, attaining a balanced accuracy of up to 90.39% on the test set for binary (Low vs. High) risk classification.

In conclusion, this work validates a holistic framework capable of reliable fall detection, high-speed impact forecasting, and robust personalized risk assessment. By fusing advanced deep learning architectures with individual physiological signatures, this research provides

the essential scientific foundation for the next generation of truly proactive, life-saving fall prevention and injury mitigation technologies.

Table of contents

List of figures

List of tables

Nomenclature

Acronyms / Abbreviations

ADL Activity of Daily Living

AMR Anisotropic magnetoresistive

ANS Autonomic Nervous System

AR Augmented Reality

ASI Anxiety Sensitivity Index

ASQ After-Scenario Questionnaire

BiLSTM Bidirectional Long Short-Term Memory

DL Deep Learning

ECG Electrocardiography

EDA ElectroDermal Activity

EDG ECG-Derived Respiration

EKF Extended Kalman Filter

EMG Electromyography

FC Fully Connected

FD Fall Detection

FDNN Fall Detection Neural Network

FF Fall Forecasting

FFNN Fall Forecasting Neural Network

FIE Fall Impact Estimation

FOG Fiber Optic Gyroscope

FRE Fall Risk Estimation

FR Fall Risk

GMR Giant magnetoresistive

GRU Gated recurrent unit

GSR Galvanic Skin Response

HCI Human–Computer Interaction

HD-EMG High-density Electromyography

HMD Head-Mounted Display

HRV Heart Rate Variability

ICA Independent Component Analysis

IMI Intrinsic Motivation Inventory

IMU Inertial Measurement Unit

IoT Internet of Things

IP Inverted Pendulum

IPQ iGroup Presence Questionnaire

KAN Kolmogorov–Arnold Network

LSTM Long Short-Term Memory

MEMS Micro-ElectroMechanical Systems

ML Machine Learning

PCA Principal Component Analysis

PPG Photoplethysmography

PTT Pulse transit time

RLG Ring Laser Gyroscope

RSA Respiratory Sinus Arrhythmia

SAM Self-Assessment Manikin

SCL Skin Conductance Level

SCR Skin Conductance Response

sEMG Surface Electromyography

SQUID Superconducting Quantum Interference Device

SSQ Subjective Symptoms Questionnaire

STF Slips, Trip, and Fall

SUM Single Usability Metric

SUS System Usability Scale

TEPR Task-evoked pupillary response

TLX Task Load Index

TMR Tunneling magnetoresistive

TNR True Negative Rate

TPR True Positive Rate

UMUX Usability Metric for User Experience

VCQ Visual Comfort Questionnaire

VEQ Virtual Embodiment Questionnaire

VRSUQ Virtual Reality System Usability Questionnaire

VR Virtual Reality

WHO World Health Organization

Part I

Introduction and Background

Chapter 1

Introduction

Slips, trips, and falls (STFs) remain one of the leading causes of severe injuries across both everyday activities and industrial environments, despite continuous advances in safety equipment and monitoring technologies. Current approaches focus predominantly on detecting falls after they occur or on modeling specific subsets of the population, most commonly older adults, leaving a critical gap in proactive fall-risk estimation and rapid injury mitigation for workers operating in hazardous settings. While wearable sensors and machine-learning models have improved the recognition of fall events, their predictive capabilities remain limited, especially under real-world conditions where human reactions, cognitive load, environmental variability, and individual biomechanics heavily influence fall dynamics.

To address these limitations, recent research has begun shifting toward fall impact prediction, multimodal sensing, and personalized modeling, yet no unified framework exists that links human biomechanics, physiological responses, and real-time predictive algorithms into a cohesive system deployable on wearable devices.

This thesis contributes to the evolving field by developing a comprehensive methodology for personalized, real-time fall forecasting and risk estimation. It combines inertial sensing, immersive simulation, physiological modeling, and advanced machine-learning architectures, including deep neural networks and Kolmogorov-Arnold Networks, to predict fall impact timing and fall-risk states in diverse contexts. The proposed framework aims to bridge the gap between laboratory-level accuracy and the stringent requirements of real-world deployment, particularly for protecting industrial workers exposed to fall hazards.

1.1 Topic

Slips, trips, and falls (STFs) represent a critical global health challenge with profound consequences for individuals, healthcare systems, and society. According to the World Health Organization (WHO), approximately 684,000 people die from STFs every year, and an additional 172 million individuals suffer STF-related disabilities annually. Over the past two decades, mortality rates due to STFs have increased more rapidly than those associated with other types of injuries, a trend driven mainly by demographic shifts such as population aging and changes in urbanization patterns. Furthermore, with the growing prevalence of chronic diseases and mobility impairments, the clinical and socioeconomic burden of falls is expected to continue rising over the following decades [10].

Beyond their immediate physical consequences, falls often initiate a cascade of long-term effects, including loss of independence, reduced quality of life, and psychological outcomes such as fear of falling, anxiety, and social withdrawal. These consequences are not limited to vulnerable individuals; falls also affect healthy adults, workers in high-risk environments, and children. Notably, many falls remain underreported, particularly non-fatal events, which means global estimates likely underestimate the true magnitude of the problem. Moreover, environmental conditions, such as slippery surfaces, poor lighting, and workspace hazards, combine with intrinsic factors like fatigue, stress, and impaired balance, making falls a multifaceted challenge that spans clinical, occupational, and public health domains [11].

The burden of STFs is evident across all age groups. For example, they are estimated to cause 32,000 deaths among children and adolescents under 15 years of age [10], 36,000 deaths among workers [12], and 39,000 deaths among people aged 65 years and older [13] in the United States. The unintentional STF death rate among adults aged 65 years and older reached 69.9 deaths per 100.000 population in 2023 in the US, showing a steady increase compared to previous years [14]. Although often overlooked, the injuries caused by STFs have many consequences, not only for the person who has the accident and those close to them, but also for the healthcare system. Specifically, the total medical cost for older people's STFs in the US in 2014 was more than 50 billion [15].

The growing burden of STFs is closely linked to the rapid demographic transition worldwide. In 2020, 9.3% of the global population was 65 years or older, which is projected to reach 16% by 2050 [16]. Notably, nearly 80% of older adults will be living in low- and middle-income countries, where healthcare infrastructure is often less equipped to cope with this challenge [17, 18]. With increasing life expectancy, chronic conditions such as diabetes, cardiovascular disease, and cognitive decline further elevate the risk of STFs and

STF-related disability [10]. Disabilities of any kind, which affect approximately 15% of the global population [19], exacerbate vulnerability to STFs and often necessitate external assistance, thus reinforcing the importance of preventive and assistive technologies.

Beyond the challenges in elder care, the occupational sector, especially high-risk environments such as construction, also faces a severe and costly burden from STFs. STF-related injuries in the workplace account for many injuries in all work sectors [10], including office work, and are the reason for significant absences from work of more than three days, especially in Small and Medium-sized Enterprises (SMEs). In Italy, falls from heights among workers result in an average of 47 days of absence from work and more than 2.5 million working days lost annually. The associated economic burden is substantial: direct costs exceed €90 million, while total expenses, including indirect consequences, are estimated at more than €370 million per year [20].

Taken together, epidemiological evidence and occupational safety statistics indicate that STFs are neither random nor unavoidable events. Instead, they are the outcome of recurrent and identifiable conditions arising from the interaction between human biomechanics, physiological state, task demands, and environmental constraints. Fatigue, cognitive load, reduced postural control, and momentary lapses in attention often coexist with extrinsic hazards, such as uneven surfaces, poor lighting, or limited footholds, particularly in dynamic, safety-critical work environments. Despite this, most existing prevention strategies remain predominantly reactive, relying on post-incident analysis or protective measures that are activated only after loss of balance has already occurred. This gap highlights a fundamental limitation in current approaches: the lack of systems capable of continuously monitoring the evolving state of the individual and anticipating hazardous situations before impact occurs. These observations naturally lead to the central research question of this thesis: can falls be modeled as predictable dynamical processes and, if so, can wearable sensing and intelligent algorithms be used to estimate fall risk and time-to-impact early enough to enable effective, personalized injury prevention? Addressing this question requires a paradigm shift from post-event detection to proactive, predictive intervention, which underpins the research presented in this work.

1.2 Aim

This thesis aims to demonstrate that current fall-prevention technologies underperform because they are based on incomplete biomechanical models and non-personalized machine-learning strategies, and to show that predictive, multimodal, and personalized modeling can

fundamentally improve real-time fall-risk estimation and mitigation. To this end, this work develops and evaluates a new computational framework that integrates wearable sensing, physiological monitoring, and predictive modeling to anticipate falls before impact and ultimately reduce fall-related injuries.

Although modern wearable systems offer accurate inertial sensors, physiological biosensors, compact computing units, and emerging mitigation technologies such as airbag vests, their performance in real-world fall prevention remains far below what these components could theoretically achieve. Why do existing systems still largely fail to anticipate falls, triggering too late or too often? Why are most fall-detection models built on datasets of simulated falls that do not reflect the real behaviors, compensatory reactions, and cognitive states of humans, especially workers operating in demanding environments? Furthermore, scientifically, what combination of data, modeling strategy, and representation is required to produce a reliable early warning system for humans at risk of falling?

These research questions lead to the central hypothesis of this thesis: that effective fall-injury prevention requires moving beyond reactive fall detection toward personalized, biomechanically informed, and multimodal predictive modeling that accounts for both physical dynamics and physiological state. In other words, preventing fall injuries is not simply a matter of recognizing a fall after it has begun, but of modeling the human as a dynamic, adaptive system whose stability can be predicted before impact.

State-of-the-art research primarily focuses on classifying a limited set of Activities of Daily Living (ADLs) and simulated STFs, without exploring how real human behavior, cognitive load, fatigue, and biomechanical variability influence fall dynamics. Moreover, many existing models overlooked the fundamental mechanical structure of the human body, treating fall prediction as a purely statistical problem, which results in algorithms that fail to generalize outside laboratory conditions. Finally, current systems rarely consider personalization, despite the well-established fact that balance, motor strategies, and physiological responses vary substantially between individuals.

For these reasons, the proposed work adopts a more scientific and integrative approach, modeling the physics of falling using simplified yet meaningful biomechanical representations, investigating multimodal physiological indicators, and developing machine-learning models capable of forecasting fall dynamics in real-time. By unifying these components, the thesis aims to contribute new knowledge at the intersection of biomechanics, physiological computing, and predictive modeling, while providing a foundation for next-generation wearable fall-prevention systems.

1.3 Terminology

This section introduces the key terminology used throughout the thesis. The definitions provided here establish a consistent conceptual framework for discussing falls, fall-related injuries, wearable sensing, predictive modeling, and the proposed methodology.

Slip, Trip, and Fall (STF) refers to a category of unintentional events/incidents in which a loss of balance occurs due to slipping, tripping, or other perturbations, potentially resulting in an injury. STF events constitute one of the leading causes of occupational injuries and fatalities across industrial sectors. In this thesis, STF represents the primary class of mechanisms targeted for detection and prevention.

A Fall is defined as an unexpected descent of the body to a lower level, typically due to loss of balance or external perturbation. Falls are modeled as dynamic transitions of the center of mass (CoM) relative to the base of support (BoS). Falls can be categorized into two types: falls to a lower level, which occur when there is a decrease in the base of support after the incident, and falls at the same level, which occur when the incident occurs without a decrease in the base of support. Falls from heights belong to falls to a lower level. In the context of this thesis, the term "fall" is sometimes used as an alternative to "STF" or to maintain the same terminology used in the literature. For example, the thesis will not refer to "STF detection" but rather to "fall detection," even though the study, analysis, and results include a wider range of cases than just falls.

Activities of Daily Living (ADLs) refer to routine, non-hazardous physical activities, such as walking, bending, sitting, standing, and turning, that must be distinguished from STF events by detection algorithms. They represent the primary source of false positives in real-world fall detection systems.

Fall Detection (FD) is the process of recognizing, in real-time, that a STF is occurring based on sensor data. In this thesis, Fall Detection is framed as the early identification of the onset of instability, prior to ground impact, but after the deviation from standard postural control.

Fall Impact Estimation (FIE) refers to the prediction of the time/trajectory remaining before the body reaches the ground once a fall has been detected.

Time-to-Impact Estimation is the temporal interval (typically measured in milliseconds) between the current moment and the estimated impact. Accurate time-to-impact estimation is essential for triggering protective devices within their mechanical activation constraints.

Fall Forecasting (FF) is the predictive estimation of the future trajectory of the body during a fall. In the context of this thesis, the research is based on dynamic modeling (such

as the inverted pendulum) and sensor observations. Unlike Fall Detection, which recognizes an ongoing fall, Fall Forecasting predicts the evolution of the fall, including estimating the future angle, orientation, and expected time-to-impact.

Fall Risk (FR) refers to the probability that a person will experience a fall in a given situation or over a specific time period. Fall risk is multifactorial, encompassing physiological conditions (such as fatigue), cognitive state (including attention), kinematic stability, and environmental context. In this thesis, Fall Risk Estimation (FRE) integrates the process of quantifying the likelihood of a fall based on multimodal sensor inputs.

Fall Detection is defined as Reactive, considering that any mitigation system would only respond reactively to the detection of a fall. Fall Impact Estimation is considered Proactive in the sense that any mitigation system can respond proactively by anticipating the moment of impact. Finally, the term Predictive refers to Fall Risk Estimation, meaning that by analyzing the estimated risk of a fall, it is possible to predict its occurrence before it happens and thus avoid or mitigate it predictively.

The Inverted Pendulum (IP) model is a simplified biomechanical model that represents the human body as a single rigid segment rotating around a pivot point (ground contact). It is widely used to study balance and prevent falls. In this thesis, the IP model is employed to simulate falls, generate synthetic datasets, and train predictive algorithms such as the Fall Forecasting Networks.

FDNN (Fall Detection Neural Network) is a deep learning model designed to classify whether a subject is currently falling or performing an ADL. The FDNN is trained on inertial data to detect deviations from stable movement patterns.

Fall-KAN refers to the Kolmogorov–Arnold Network developed in this thesis for Time-to-Impact estimation. KANs use learnable activation functions on edges rather than fixed nonlinearities in nodes, enabling higher accuracy with fewer parameters. Fall-KAN is introduced as a novel modeling paradigm for fall prediction.

FFNN (Fall Forecasting Neural Network) is a recurrent neural network, based on GRU cells, that predicts the future orientation of the body during a fall using signals derived from the inverted pendulum model. It outputs a prediction of the trajectory and assists in Time-to-Impact estimation.

In this thesis, Generalized models refer to algorithms trained on a small population and capable of delivering excellent results on a larger population. Personalized Models are subject-specific ML/DL algorithms that adapt to individual physiological and biomechanical characteristics.

1.4 Contributions

Unlike conventional approaches that treat fall detection as a standalone classification problem, typically limited to post-impact detection or pre-impact thresholding, this thesis presents a unified, multimodal framework that encompasses physiological fall-risk estimation, pre-impact fall detection, fall trajectory forecasting, and real-time mitigation decision-making. By integrating wearable biosensing, machine learning, and biomechanics, the framework advances beyond the fragmented approaches commonly found in the literature [21–30]. It contributes to the field by demonstrating that fall-related injury prevention can be viewed as a continuous, temporal process, rather than as separate, isolated tasks.

A recurrent neural network–based fall detection algorithm (FDNN) was designed, optimized, and validated on the SisFall dataset with newly curated fine-grained annotations. The contribution lies not only in the model itself, but also in demonstrating that temporal modeling of complete IMU sequences can significantly improve pre-impact recognition compared to threshold- or handcrafted-feature methods. It contributes to the scientific understanding of how sequence models behave under noisy, high-frequency fall dynamics.

This thesis is among the first to apply Kolmogorov-Arnold Networks (KANs) to inertial biomechanics for real-time fall forecasting. It also proposes a physics-informed ML strategy (FFNN) in which an Inverted Pendulum model is simulated to generate thousands of falls used for training. The scientific contribution is two-fold: demonstrating that KANs can outperform classical regression and neural architectures in estimating time-to-impact from IMU signals, and introducing a hybrid biomechanics approach with a machine learning pipeline that improves prediction robustness and generalization for fall-mitigation systems.

Two novel multimodal datasets were acquired using advanced biomedical sensors in combination with controlled and VR-based exposures to high-risk scenarios. These datasets fill a significant gap in existing public datasets [31–39], which typically lack physiological, cognitive, and emotional indicators preceding falls. By providing synchronized contextual information from ECG, EMG, EEG, respiration, EDA, IMU, and VR, this work lays the empirical foundation for research on fall risk modeling and expands understanding of how physiological and cognitive states can influence the probability of falling.

The thesis leverages pupillometry, EDA, EEG, and multimodal VR-based stressors to reveal how cognitive load, arousal, fatigue, and emotional responses shape fall risk. In contrast to prior literature [21–30] that has primarily focused on kinematics, this study argues that accurate fall-risk prediction depends on monitoring cognitive states, a finding of particular importance for high-risk professions such as construction work.

This thesis demonstrates that fall risk is inherently subject-specific, influenced by factors such as fatigue, physiological conditions, stress, and workload. The proposed multimodal machine learning models demonstrate that integrating cardiorespiratory signals, neuromuscular activation, cognitive load, and kinematics enables individualized monitoring that adapts to user variability. This contribution advances the field by demonstrating that future fall-prevention systems must move toward personalization, transcending population-level threshold models.

In summary, this thesis contributes new methods, datasets, and insights that collectively push the field beyond reactive fall detection toward proactive/predictive fall-injury prevention, guided by personalization, multimodal sensing, biomechanics, and next-generation neural architectures. These contributions provide a foundation for future research, real-world deployment, and eventual standardization of wearable predictive safety systems.

1.5 List of Publications

The following publications are part of the results of the work carried out during the studies:

- [40]: Cartocci, Nicholas, et al. "Real-time Fall Prevention system for the Next-generation of Workers". Workshop on Assistive Robotic Systems for Human Balancing and Walking: Emerging Trends and Perspectives @IROS 2022.
- [41]: Cartocci, Nicholas et al. "Artificial intelligence-based wearable solution to prevent fall from heights injuries for the next generation of workers". 2023 Slips, Trips, and Falls (STF) International Conference.
- [42]: Cartocci, Nicholas, et al. "Deep Learning-based Wearable Device to Prevent Fall from Height Injuries". 2023 Italian Conference on Robotics and Intelligent Machines (I-RIM).
- [43]: Cartocci, Nicholas, et al. "Fall-KAN: Fall impact time estimation Kolmogorov-Arnold Network". 2024 International Conference on Electrical, Computer, Communications, and Mechatronics Engineering (ICECCME).
- [44]: Cartocci, Nicholas, et al. "Recognition of Physiological Patterns during Activities of Daily Living Using Wearable Biosignal Sensors". 2024 Triennial Congress of the International Ergonomics Association (IEA).

- [45]: Cartocci, Nicholas, et al. "How can AI reduce fall injuries in the workplace?". 2025 International Conference on Safety & Innovation (ICSI).
- Cartocci et al. (2025b): Cartocci, Nicholas, et al. "Machine Learning Framework for Fall Detection and Time-of-Impact Estimation". 2025 Precision Health Day.
- [46]: Pitzalis, Roberto, et al.* "Evaluating Muscle Fatigue With Non-Invasive Approaches: A Review of Methods, Metrics, and Implications". IEEE Transactions on Neural Systems and Rehabilitation Engineering. *The PhD candidate is one of the co-first authors.
- [47]: Mariani, Giulia, et al.* "Physiological Measures of the Mental Workload in Users of a Lower Limb Exosuit: A Comparison of Subjective and Objective Metrics". 2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC). *The PhD candidate is one of the co-first authors.
- [48]: Moreno Franco, Olmo Alonso, et al. "Telepresence in Construction Sites: Usability Study of a VR-Based Environment for Working in Heights". 2025 IEEE Conference on Telepresence (TELEPRESENCE).

Most of the content related to Fall Risk Estimation will be considered for submission to a scientific journal. The complete list of dissertations can be found on the PhD candidate's *Google Scholar profile* or on *his website*.

1.6 Outline

This thesis is structured into three main parts, comprising a total of 12 chapters, which address the introduction, necessary background, developed framework, and concluding remarks.

- Part I, Introduction and Theoretical Background, comprises five chapters. Chapter 1 provides the motivation and urgency of the STF problem, highlighting its global health and economic burden. It outlines the topic, the limitations of current assessment methods, the overall aim of the research, and lists the main contributions. Chapter 2 discusses the global issue of STFs across different risk groups and the associated financial and social costs. It categorizes the diverse causes and risk factors, providing a critical overview of current commercial solutions. Chapter 3 introduces the basic principles of both inertial sensors and biomedical sensors, which are essential for continuous motion and physiological monitoring in this work. It also covers the

concept of Virtual Reality as an experimental platform. Chapter 4 details the self-report instruments used to evaluate human factors and user experience in the experiments, including metrics for perception, ergonomics, usability, and VR immersion. Chapter 5 reviews the foundational concepts of Machine Learning and Deep Learning, including model evaluation metrics for both regression and classification, as well as different learning methods utilized throughout the thesis. It also discusses strategies for handling class imbalance, a critical issue in fall-related data.

- Part II, Framework, is organized into four chapters. Chapter 6 provides an introduction to the experimental framework that has been developed. Chapter 7 summarizes the evolution of fall detection research, encompassing threshold-based, traditional ML, and deep learning methods. It critically reviews public datasets and presents the approach and results of the approach developed. Chapter 8 details the methodology for predicting impact time and Fall Forecasting and presents the corresponding results and discussion. Chapter 9 focuses on fall risk estimation and prevention by introducing a multimodal framework. The chapter details the collected datasets, the methodology for feature extraction and classification, and the study's results.

- Part III, Conclusions, comprises three chapters. Chapter 10 synthesizes the findings from all framework components, discussing the overall impact of the developed algorithms and the insights gained from the multimodal datasets. Chapter 11 suggests directions for future research, focusing on the deployment of real-world adaptive systems and personalized fall prevention strategies. Finally, Chapter 12 provides a summary of the thesis and reiterates the fulfillment of the research aims.

Chapter 2

Slips, Trips, and Falls (STFs)

Slip, Trip, and Fall (STF) constitutes a critical, multifaceted public health challenge worldwide, extending far beyond the immediate trauma of an injury. It is a leading cause of unintentional injury deaths and a main cause of morbidity and long-term disability across all age groups, with particularly devastating consequences for older adults and specific occupational sectors [10]. The global problem is exacerbated by large-scale demographic and societal trends, including rapid urbanization, increasingly sedentary lifestyles, and the progressive aging of the worldwide population. These factors have been consistently identified not only in global surveillance reports [49, 50] but also in recent epidemiological and public health reviews that highlight how demographic transitions and lifestyle changes systematically increase fall incidence and fall-related injuries across both occupational and non-occupational settings [51–53].

2.1 Statistics

Globally, over 684,000 people die each year from an STF, making STFs the second leading cause of unintentional injury deaths in the world. An estimated 172 million STFs occur yearly, resulting in short- or long-term disability [10, 12, 54]. This results in over 38 million Disability-Adjusted Life Years (DALYs) lost annually, a figure that has been consistently rising. From 2000 to 2019, there was a 53% increase in the number of total deaths due to STFs globally, vastly outpacing the increase in deaths from all other injuries combined during the same period. STFs result in more Years Lived with Disability (YLDs) than transport injury, poisoning, drowning, and burns combined. The vast majority (over 80%) of fatal STFs occur in low- and middle-income countries [10, 49].

The burden of STFs is not confined to a single demographic but affects three key risk groups: children and adolescents, workers, and older people.

2.1.1 Children and adolescents

While STF is a regular part of growth, severe STFs in young people pose a significant risk of injury. STFs are the leading cause of injury in young children and are estimated to account for 25–56% of all child injury hospital visits [55]. Most STFs among children and adolescents occur in or around the home. For infants, STFs most often result from falling from furniture or a caregiver's arms at home [56]. In 2019, STFs were responsible for an estimated 31.818 deaths among children and adolescents aged below 15 years globally [10, 49]. Child STF mortality rates are up to three times higher in low- and middle-income countries than in high-income countries [57].

2.1.2 Workers

The International Labour Organization (ILO) reports that occupational accidents are responsible for millions of injuries each year [58, 59]. STFs are consistently among the most frequent causes of serious or fatal workplace injuries. According to Eurostat data [1], STF is the second most common cause of non-fatal and fatal accidents in the workplace across the European Union. In 2023, nearly 2.825 million non-fatal workplace accidents occurred. Meanwhile, although fatalities are relatively rare (0.1% of total accidents), the human cost of those fatal accidents continues to draw attention. In 2017, the estimated number of global deaths from occupational STFs was 36,000, accounting for 12% of all occupational injury deaths [12].

The construction sector has the highest worldwide rate of fatal STFs [10], accounting for a disproportionate share of workplace accidents also in the US and EU. In the US, data from the Bureau of Labor Statistics reveal that nearly 1 in 5 workplace deaths in construction are due to STFs. More precisely, in 2022, 38.4% of fatal injuries in construction were attributed to STFs [60]. According to recent Eurostat analysis [1], a quarter (24.0%) of all fatal work accidents in the EU occur in the construction sector. Other high-risk sectors include Transportation and Storage (16.4%), Manufacturing (13.4%), and Agriculture (12.9%), as shown in Figure 2.1. The EU reported 3.298 fatal work-related accidents. It aligns with US trends, where STFs remain a leading cause of serious injuries and fatalities. While not all STFs are attributable to STFs from heights or on the same level, a significant fraction is [61].

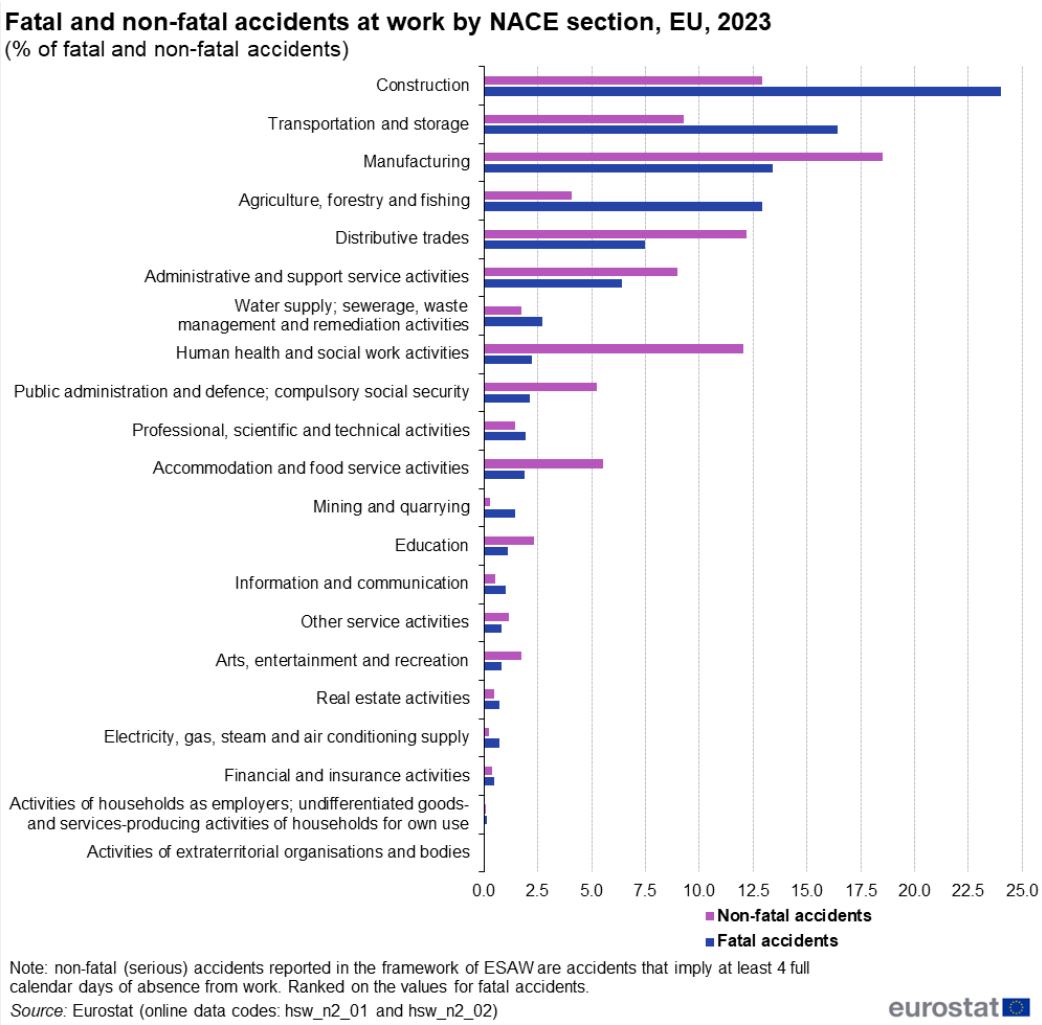


Figure 2.1 Fatal and non-fatal accidents at work by NACE section, EU, 2023 [1].

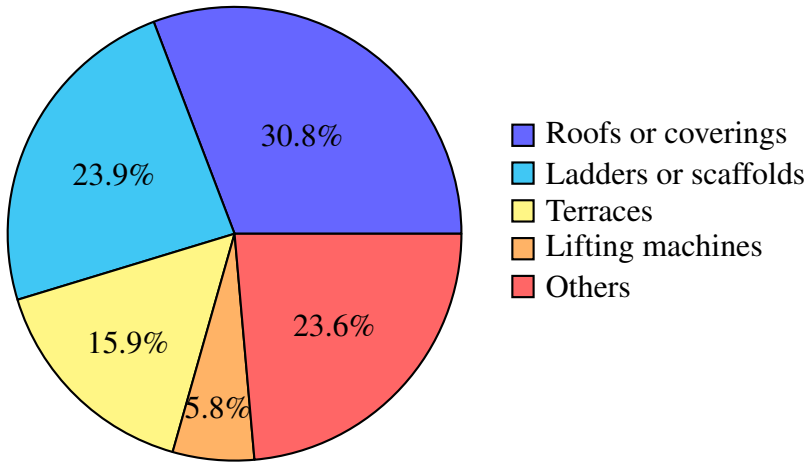


Figure 2.2 Fatal STFs from height by type of accident in Italy [2].

Also, the construction sector in Italy is the most heavily impacted by fatal falls from height [20, 2]. For the period 2008-2012, over 65% of fatal falls from height occurred in the construction sector, followed by agriculture at approximately 11%. For fatal falls from height, Fig 2.2, 30.8% of cases were from roofs or coverings, 23.9% were from work-at-height equipment (ladders, scaffolds), 15.9% were from fixed parts of buildings (such as terraces), and 5.8% were from lifting machines.

Sector-Specific Vulnerabilities

Although STFs occur in all sectors, as highlighted above, specific industries show higher rates due to task complexity, environmental conditions, and physical workload [1]:

- Construction: Highest rate of fatal STFs, elevated risk due to working at height, unstable surfaces, ladders, scaffolding, and exposure to weather.
- Manufacturing and Warehousing: Frequent non-fatal STFs caused by cluttered pathways, spills, carrying heavy loads, and exposure to fast-paced environments.
- Transportation and Utilities: Increased risk due to loading/unloading operations, uneven terrain, and vehicle ingress/egress.
- Healthcare: High rates of same-level STFs from rapid movement, patient handling, and long shift durations.

Systematic reviews and workplace audits [62–65] highlight several recurring contributors:

- Environmental hazards: height exposure, slippery floors, uneven ground, debris, and inadequate lighting.
- Task-related factors: manual handling, carrying tools, climbing, and time pressure.
- Human factors: repetitive or physically demanding work, fatigue, cognitive overload, inattention, and insufficient supervision.
- Weather and outdoor conditions: humidity, rain, ice, and wind significantly increase the risk of STF.
- Organizational issues include inadequate safety training, a lack of STF protection systems (such as harnesses, guardrails, or STF-arrest devices), and poor risk assessment.

Significantly, the mechanisms of STF differ across populations. While older adults typically STF due to physiological decline (such as balance deficits or muscle weakness), younger workers often STF during dynamic movements, while attempting corrective reactions, or while interacting with tools and equipment [66, 67]. A more detailed description is reported in Section 2.2. These differences make universal, one-size-fits-all fall detection systems ineffective and underscore the need for adaptive, context-aware, and user-specific STF prevention and mitigation technologies.

2.1.3 Older Adults (65+)

Age is the key risk factor for STFs [68]. The risk of injury or death from STFs increases exponentially with age globally [54]. Approximately one-third of people aged 65 and older experience a slip, trip, or fall at least once a year. This rate increases to 32–42% for those over 70 [50]. The average STF incidence in nursing homes is higher, estimated at 1.6 STFs per bed annually [69]. While older women are more likely than men to slip, trip, or fall and suffer fractures, STF-related mortality is higher among older men [70]. For those over 60, STF-related death rates are roughly similar for both sexes, but over 85% of STF-related deaths in women occur in those aged over 60, compared to just over 60% for men [50].

Slip, trip, and fall mortality among the elderly is a significant concern. For example, for people in Switzerland aged 65 and over, 2,044 deaths from non-occupational incidents were recorded in 2020, with 1,915 of those deaths occurring in the home and leisure environment [71]. The fatality rate (deaths per million inhabitants) for home and leisure incidents in this age group is extremely high (1,193 per million for 65+ in 2020). However, 75% of fatal

STFs among older people (aged 70 and over) occur in low- and middle-income countries [10].

2.1.4 Financial and social cost

The scale of the problem translates into enormous financial [72, 73] and social costs [74, 75], providing a strong rationale for preventive measures [10].

In the US, the annual medical costs for older adult STFs are roughly $50 billion [15]. For individuals aged 65 and over, the average cost of hospitalization for a STF-related injury ranges from $6.646 to $17.483 across different countries [76, 77]. These costs are projected to soar to $240 billion by 2040 [50, 78].

Beyond direct medical expenses, indirect costs encompass lost productivity for the injured individual and family caregivers, with average lost earnings estimated to be substantial in high-income countries. As a first approximation, indirect costs can be considered to be approximately three times higher than direct costs [20]. For example, the total societal cost of non-occupational injuries in Switzerland (circulation, sport, home, and leisure), where STFs are the leading home injury, reached about $14 billion in material costs alone in 2020 [71].

This substantial human and financial burden highlights a critical and urgent need for proactive detection and prevention solutions. Through intelligent data analysis and machine learning techniques, it is possible to move beyond passive reporting and detection towards anticipating and mitigating STF risk factors.

2.2 Causes and Risk Factors

Slips, trips, and falls are typically the result of a complex interplay between intrinsic (personal) factors and extrinsic (environmental) hazards [11, 79–82]. Simply put, no single cause is usually sufficient; the more risk factors a person accumulates, the higher their likelihood of experiencing a STF.

This section reviews the most salient causes and risk factors, grouped into intrinsic, extrinsic, and situational categories, highlighting key limitations in how the literature treats them.

2.2.1 Intrinsic factors

Intrinsic factors relate to the individual's health status, functional ability, sensory systems, cognition, and behaviors. Major intrinsic risk elements include:

- Age and physiological decline: Advancing age is among the strongest risk factors for STFs. Older adults experience decreased muscle strength (sarcopenia), slower reflexes, diminished proprioception, reduced vision/hearing, and slower recovery responses to perturbations [10, 83]
- Muscle weakness, balance, and gait disorders: Impaired strength, especially in the lower limbs, and deficits in balance or gait markedly increase STF risk. A history of gait instability or postural sway is consistently identified in reviews [52, 84]
- Sensory and cognitive impairments: Visual impairment (including low acuity and contrast sensitivity) and hearing loss contribute to STFs by reducing awareness of obstacles and environmental hazards. Cognitive decline, dementia, or attention deficits further compromise safe mobility [85, 86].
- Chronic medical conditions and medications: Conditions such as diabetes, stroke, Parkinson's disease, orthostatic hypotension, and cardiovascular disorders raise STF susceptibility. Polypharmacy, especially medications causing dizziness, sedation, or orthostasis, is likewise important [87, 88]
- History of prior STFs and fear of falling: Prior STFs are a strong predictor of future STFs; the psychological consequence is heightened fear of falling, often leading to reduced activity, muscle deconditioning, and increased STF risk [89].

While many studies enumerate these factors, fewer capture how they interact or examine the temporal sequence before a STF. Moreover, many datasets focus on older, healthy volunteers rather than high-risk populations, which limits the generalization of the findings.

2.2.2 Extrinsic factors

Extrinsic factors refer to the external environment and the individual's behavior. Key elements include:

- Home and community hazards: Slippery surfaces, poor lighting, uneven floors or rugs, lack of handrails, and clutter significantly elevate risk [79].

- Footwear and mobility aids: Inappropriate footwear (such as high heels or backless shoes), as well as misuse or lack of assistive devices like walkers or canes, contribute to STFs [83].
- Transition and postural challenges: Frequent antecedents include rising from a chair, climbing stairs, or walking on wet surfaces. Postural hypotension and rapid movements exacerbate risk [87].
- Behavioral factors: Risk-taking behaviors, hurried movements, distractions (such as using mobile devices while walking), or inadequate environmental awareness may trigger STFs.

Many studies assess extrinsic hazards in isolation, yet the STF event often emerges from the interaction of intrinsic vulnerability plus extrinsic trigger. Research that models this synergy remains limited.

2.2.3 Situational factors

The most insightful perspective may be that STFs result not from a single cause but from the intersection of vulnerabilities and situational triggers at a given moment [90, 87]:

- Multiple risk factors accumulate: The more risk factors present (primarily intrinsic), the higher the STF probability [91, 92].
- Acute changes or perturbations: Transient events (such as a sudden drop in blood pressure, an unexpected obstacle, a slippery ground, or a misstep) may precipitate a STF in a person otherwise stable [93].
- Activity context: STFs are more likely during transitional activities (such as changing rooms or reaching overhead) or dual-task movement when attention is split [94].

The literature often records the STF as the event, but less often the cause (such as fatigue, acute illness, or sensory overload). This gap limits prevention strategies.

2.3 Commercial solutions

Commercial technologies addressing STFs now encompass a range of functions, from detection (after a STF has occurred) to prevention (monitoring risk and intervening before a STF), and mitigation (reducing injury when a STF cannot be avoided). The following subsections

(a) Lifeline On the Go (b) RealSense D435i (c) Varyyar Care

Figure 2.3 Fall Detection solutions

cover each modality of deployment, providing detailed descriptions of representative solutions along with their benefits and limitations, to enable a critical assessment of the market landscape.

2.3.1 Wearable solutions

Wearable devices remain the most prevalent commercial approach for fall detection. These typically embed inertial sensors and connectivity (Bluetooth/cellular) to trigger alarms when STF are detected.

For example, smartwatches (such as the Apple Watch Series 4 and later models) incorporate automatic fall detection features. When sudden acceleration is followed by immobility, the device alerts an emergency contact and can initiate a call to emergency services. While this offers convenience and mobility for users, it remains reactive (such as the STF has already happened). Additionally, such solutions rely on continuous device wear and proper positioning; however, battery limitations and device removal (for example, when entering a bath) reduce their effectiveness in real-world applications.

Another example is the pendant-style medical alert systems, such as Lifeline On the Go (Figure 2.3a). These devices are well-established in the market and often meet the standards of assisted living. They offer reliability for users who are less inclined to manage smartphones or apps. However, these too are primarily post-fall alerts and do not actively monitor pre-fall risk or intervene in it.

Actual wearables excel in portability and connectivity, but compliance remains a significant bottleneck, as many users forget to wear or recharge their devices. Furthermore, while inertial thresholds can detect STFs, they rarely account for contextual risks (such as

slippery surfaces or fatigue) or provide preventive suggestions. Thus, while detection is robust, prevention remains under-addressed in this category.

2.3.2 Camera-based solutions

In assisted-living or hospital settings, camera-based systems (such as RGB or depth) operate to detect STFs and other contextual hazards. They provide rich visual data (body posture, trajectory, obstacle location) that supports more accurate discrimination between STFs and non-STFs.

For example, systems built around depth cameras (such as those using RealSense, Figure 2.3b) can detect body shape and movement without requiring color detail, which provides some privacy protection [95]. In such solutions, the system may detect a STF event, trigger alerts, and monitor gait or risky motion patterns to prevent them.

Due to privacy concerns and installation complexity, these systems are robust in supervised settings but less suited for home deployments. Vision-based monitoring still raises user acceptability issues, and its performance may degrade in dim light or when furniture is moved. Moreover, many commercial systems still prioritize detection over proactive prevention or mitigation.

2.3.3 Ambient systems

Ambient systems embed sensors into the environment, such as radar, pressure mats, and Wi-Fi motion sensing, to monitor users without requiring them to wear a device. These systems hold promise for detection and prevention because they can continuously monitor movement patterns, posture changes, and environmental hazards.

Examples include mmWave radar solutions, such as those from Vayyar Care (Figure 2.3c), which can detect STFs in bathrooms or bedrooms without requiring wearable devices for compliance. Innovative flooring systems (pressure-sensitive mats) detect deviations in gait or prolonged immobility, serving as early risk indicators. These sensors help reduce reliance on user behavior and increase passive coverage.

Ambient solutions take a step toward prevention by jointly monitoring the environment and user activity. However, they are constrained to fixed rooms (such as a bathroom or bedroom) and often cannot follow the individual outside that zone. Installation and maintenance costs can be high, and signal interference (multiple people, pets) may cause false alarms. Furthermore, many systems still focus on fall detection rather than risk modeling.

(a) D-Air Lab WorkAir

(b) Protechto MAD WORK

(c) Skylotec Skyvest

(d) Safeware C3

(e) Chutex Elder

(f) S-AIRBAG

Figure 2.4 Mitigation solutions

2.3.4 Commercial mitigation solutions

A growing subset of commercial solutions operates not only after or before STFs but also during STFs, aiming to reduce the severity of injuries. These are often marketed to high-risk workers (construction or industrial) but are increasingly targeted to seniors for fall mitigation.

D-Air Lab WorkAir

WorkAir (D-Air Lab Srl, Vicenza, Italy) [96], Figure 2.4a, is a certified Personal Protective Equipment (PPE) airbag vest designed for workers at height. According to the manufacturer, the vest inflates within just 40 milliseconds after fall detection is triggered by a tri-axial accelerometer and gyroscope embedded in the vest. The protection zone covers the chest, back, and vital organs with a reduced transmitted force of up to 60%.

The vest addresses the injury phase of falls, offering mitigation rather than purely detection. It is fully wearable and certified under PPE standards. It is designed primarily for falls from height ($\geq 1.2m$), typical of construction scenarios. The price is about 800€.

Protechto MAD WORK

The MAD WORK (Protechto Srl, Gallarate, Italy) [97] innovative airbag vest is a category II PPE designed specifically for fall protection in work environments, including falls from minimal heights (down to 2 meters). The vest shown in 2.4b uses an advanced electronic detection and control system with an AI algorithm to continuously monitor the wearer's stability. If the system detects an anomaly consistent with a fall, the airbag instantly activates, fully inflates, and provides protection in just 250 milliseconds.

The deployed airbag safeguards vital areas, including the back, chest, head, and coccyx, through an extended rear flap, significantly mitigating impact by 97%. Additionally, the vest features certified high-visibility materials and a dedicated app that automatically sends an emergency message with the worker's geolocation to pre-set contacts for rapid rescue intervention. The price is almost 1300€.

Skylotec Skyvest

The Skyvest (SKYLOTEC GmbH, Neuwied, Germany) [98], as illustrated in Figure 2.4c, is marketed as an airbag vest for falls from low heights (1.7–3.2 m), equipped with sensors that detect movement, acceleration, and position changes. It protects the chest, spine, back of the head, and tailbone region, is reusable after deployment, and is TÜV-certified category 2 PPE.

Covers a broader region (head/back/tailbone) and is suited for fall scenarios, not just free-fall but also "swinging" falls typical in scaffolding/low height work. Similar to WorkAir, the device is designed for use in occupational settings. The weight (2.39 kg) and price (900€) may limit use.

Safeware C3/C-light

Safeware (Safeware Inc., Gyeonggi, South Korea) [99] offers two smart wearable airbag vests, the C3 and C-Light, primarily targeting fall mitigation in occupational environments. Both models use the X-Motion AI algorithm for accident detection, enabling millisecond-level airbag deployment to protect vital body parts, including the head, cervical vertebrae, spine, and pelvis, thereby significantly reducing injury severity.

The C3 is an industrial-grade solution with a larger 70-liter airbag capacity and a suggested operating range of 1 to 5 meters (Figure 2.4d). Conversely, the C-Light is a machine-washable, lightweight alternative with a 45-liter capacity, designed for tasks at lower heights, ranging from 1 to 3 meters, such as indoor maintenance. Both systems are reusable by replacing the cartridge and, upon detecting a fall, automatically send an emergency alert with the accident location to designated contacts via a connected smartphone app. The prices are not publicly disclosed.

Chutex Elder

The Chutex Elder (Chutex SAS, Lyon, France) [100], as shown in Figure 2.4e, is specifically designed for older adults living independently. It incorporates fall detection and inflates airbags to protect the hips and back, and connects via a smartphone app to alert caregivers.

Bridges the gap between occupational protection and elder-care mitigation. As with other vests, wearing compliance remains challenging, and smartphone alerting assumes app usage and internet connectivity; published real-world data is limited. The price is about 900€.

S-AIRBAG

Marketed as full-body protection (head to pelvis) for elderly fall mitigation, S-AIRBAG (S-AIRBAG, Guangzhou, China) [101] integrates sensors and a smartphone app, Figure 2.4f.

Comprehensive protection of vulnerable anatomy is connected to the alerting network. The high cost and limited field validation make large-scale adoption uncertain. The price is almost 850€.

2.4 Summary

Commercial STF solutions exhibit varying trade-offs across detection, prevention, and mitigation categories.

Detection solutions (wearables and ambient/local sensors) are the most mature and widely adopted. However, they are reactive, offering alarms after the STF rather than preventing the event or reducing injury severity. They also depend heavily on user adherence, proper use, and charging.

Prevention-oriented systems (ambient sensors and smart homes) hold promise for continuous monitoring of risk patterns (such as gait changes, posture anomalies), but face installation,

cost, and contextual challenges. Few commercial systems fully integrate ML-based risk prediction yet.

Mitigation solutions (airbag vests) are emergent and valuable in high-risk settings, yet they are less widely used in everyday elder-care environments. The focus is on injury reduction rather than STF prevention; consequently, they do not capture the whole chain of STF risk, trigger, event, and outcome.

From an ML/data processing perspective, many commercial devices rely on threshold-based algorithms, which limit user and setting adaptability. There is a movement toward integrating advanced ML/DL to monitor pre-STF signatures (such as gait instability, fatigue) and to fuse data across sensors, but this has not yet dominated consumer solutions. The key challenge lies in transitioning from retrospective analysis (fall detection) to proactive, millisecond-scale prediction (fall impact prediction) and continuous fall risk modeling, which necessitates the robust, multimodal ML framework developed in Chapter 6.

Chapter 3

Fundamentals of Sensors

Understanding and accurately modeling STFs requires more than interpreting isolated movement patterns. STFs emerge from complex interactions between biomechanics, physiology, and the environment; therefore, any predictive or preventive framework must rely on high-quality sensor inputs that capture these multidimensional processes. This chapter provides the foundational knowledge necessary to understand the sensing technologies used throughout this thesis, explaining both what they can measure and their limitations.

Inertial and biomedical sensors have become essential tools in modern fall detection and fall-risk monitoring systems. Inertial sensors enable precise measurement of acceleration, rotation, and orientation, forming the backbone of most wearable fall-detection algorithms. Biomedical sensors, on the other hand, provide insight into internal physiological and cognitive states, such as muscle activity, autonomic arousal, cardiovascular dynamics, and visual–cognitive load, which often precede or contribute to a fall event. The complementary strengths of these sensor modalities allow a transition from purely reactive fall detection toward proactive and personalized fall-risk estimation.

However, each technology also comes with inherent limitations: MEMS sensors introduce noise and drift, biomedical signals are affected by motion artifacts and placement variability, and environmental factors can bias measurements. Understanding these capabilities and constraints is crucial for designing robust machine learning models, accurately interpreting results, and recognizing the limitations of what can be reliably inferred from sensor data.

For these reasons, this chapter reviews the fundamental operating principles of the inertial and physiological sensors used in this work. This knowledge forms the technical basis for the methodological choices made in the subsequent chapters, where sensor data are fused, processed, and modeled to develop systems for fall detection, fall impact prediction, and fall-risk estimation.

3.1 Inertial sensors

This section introduces the fundamentals of inertial sensors, with a focus on accelerometers, gyroscopes, and magnetometers, as these are the most widely employed sensing modalities in fall detection and prediction applications. When combined into an Inertial Measurement Unit (IMU), these sensors provide a three-dimensional characterization of human movement by measuring linear acceleration, angular velocity, and, when available, the local magnetic field for orientation estimation. Modern IMUs are typically manufactured using micro-electromechanical systems (MEMS) technology, which enables minimal form factors, low power consumption, and integration into wearable and mobile devices. Beyond their compactness, MEMS-based sensors are valued for their high sensitivity, low cost, and robustness, making them suitable for continuous monitoring in real-world conditions [102]. The fusion of signals from accelerometers, gyroscopes, and magnetometers, often achieved through algorithms such as complementary filters, Kalman filters, or machine learning-based sensor fusion techniques, enables the accurate estimation of posture, gait dynamics, and activity recognition. These capabilities form the technological foundation for developing reliable, unobtrusive systems to detect falls or assess fall risk.

3.1.1 Accelerometer

Accelerometers are the most widely adopted inertial sensors in commercial and research applications, primarily due to the maturity of MEMS technology, which enables the development of small, inexpensive, and low-power devices. Depending on their design and performance, accelerometers and IMUs are classified as low-, medium-, and high-grade.

- Low-grade sensors (costing as little as one dollar) are typically used in consumer electronics and automotive safety systems, such as crash airbags and pedestrian navigation [103]. They are also integrated into smartphones, wearable devices, and gaming controllers, where size, cost, and power efficiency are prioritized over absolute accuracy.
- Medium-grade sensors provide improved accuracy and stability, and are often employed in industrial robotics, drones, and pedestrian dead reckoning systems [104].
- High-grade sensors offer superior accuracy and minimal drift, making them suitable for aerospace and marine applications, including submarine navigation and aircraft inertial navigation systems. These sensors, however, can cost tens of thousands to millions of dollars [103, 105].

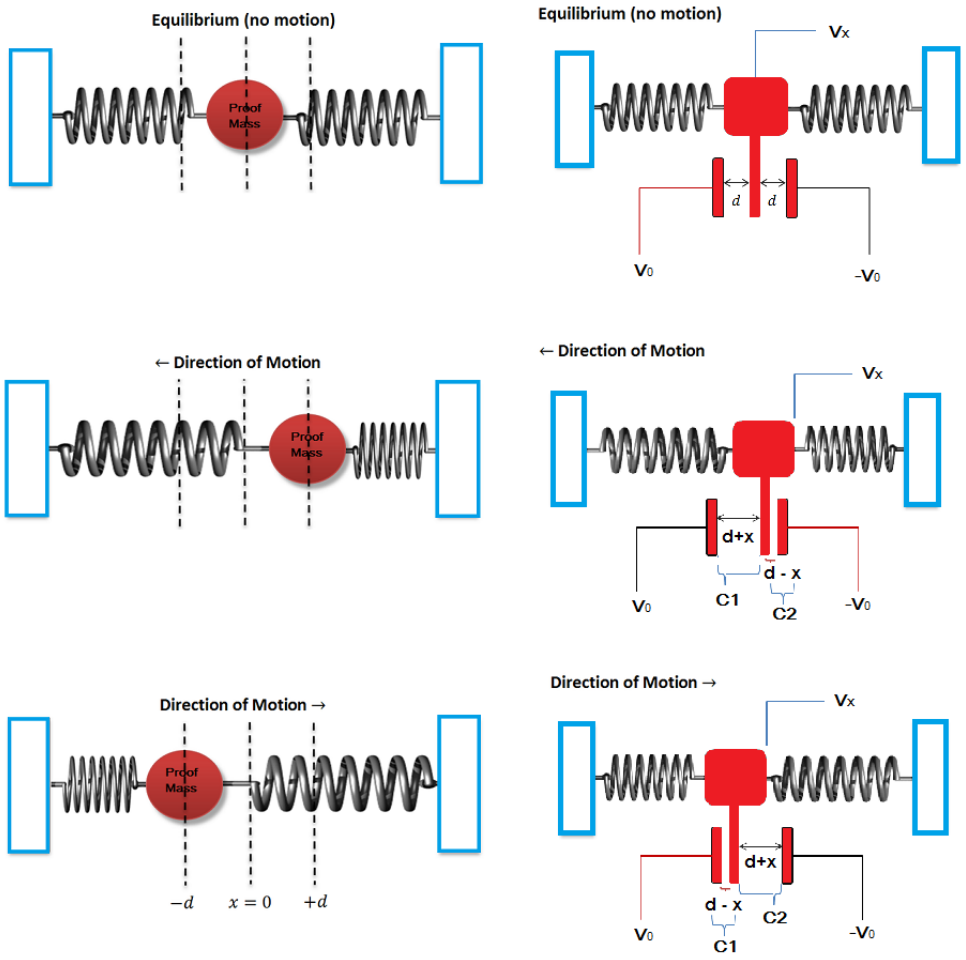


Figure 3.1 Capacitive accelerometer principle [3].

The transduction mechanisms of MEMS accelerometers include capacitive, piezo-resistive, electromagnetic, piezoelectric, ferroelectric, optical, and tunneling approaches [106]. Among these, capacitive accelerometers have become the most commercially successful due to their low power consumption, high sensitivity, robustness, and stability across a wide range of temperatures [107]. Their operating principle relies on a suspended proof mass attached to springs within the sensor package. When acceleration occurs along a sensitive axis, the mass displaces relative to the case, and this displacement is converted into an electrical signal by capacitive plates (Figure 3.1).

Accelerometers inherently measure static acceleration (such as gravity) and dynamic acceleration (such as movement). For instance, an accelerometer at rest will register the Earth's gravitational field ($g \approx 9.81 m/s^2$). In fall detection and prevention applications, this property is critical: sudden peaks in dynamic acceleration can indicate impact events, while

static acceleration values provide orientation cues (such as lying down versus standing). An another important design consideration when using accelerometers for fall detection is the selection of an appropriate full-scale range, defined as the maximum measurable acceleration that the sensor can represent without saturation. Several experimental studies report peak accelerations during falls ranging from approximately $3g$ up to $10-12g$, depending on sensor placement, fall direction, and surface compliance [108]. However, raw accelerometer signals often contain noise and gravitational components, requiring filtering or sensor fusion techniques to be applied before they can be used for movement or activity analysis.

The application landscape of accelerometers has expanded significantly over the past few decades. While in 2006, most accelerometers were still sold to the automotive market [106], they are now ubiquitous in consumer electronics. Smartphones, tablets, digital audio players, and mobile PCs incorporate accelerometers for orientation sensing, screen adjustment, and gesture recognition. Wearables and IoT devices rely on them for step counting, activity recognition, gait analysis, and fall detection [109]. In healthcare, accelerometer-based IMUs have become one of the leading tools for continuous and unobtrusive monitoring of physical activity and fall risk in elderly populations [110].

By combining their affordability, sensitivity, and low power consumption, accelerometers remain the cornerstone of IMUs. Their integration with gyroscopes, magnetometers, and advanced sensor fusion algorithms enables accurate motion tracking across a broad range of applications, from entertainment to critical healthcare monitoring.

3.1.2 Gyroscope

Gyroscopes are inertial sensors that measure angular velocity (expressed in rad/s or $^\circ/s$), providing critical information about rotational motion and orientation. Three main categories of gyroscopes exist: mechanical, optical, and vibratory [104].

- Mechanical gyroscopes are based on spinning rotors and were historically used in navigation systems, though solid-state solutions have mainly replaced them due to their bulk and maintenance requirements.

- Optical gyroscopes, such as Ring Laser Gyroscopes (RLGs) and Fiber Optic Gyroscopes (FOGs), exploit the interference of light beams to detect rotation via the Sagnac effect. These devices offer excellent stability and are commonly deployed in aerospace, marine, and defense applications, but their cost and size limit their use in consumer electronics [105].

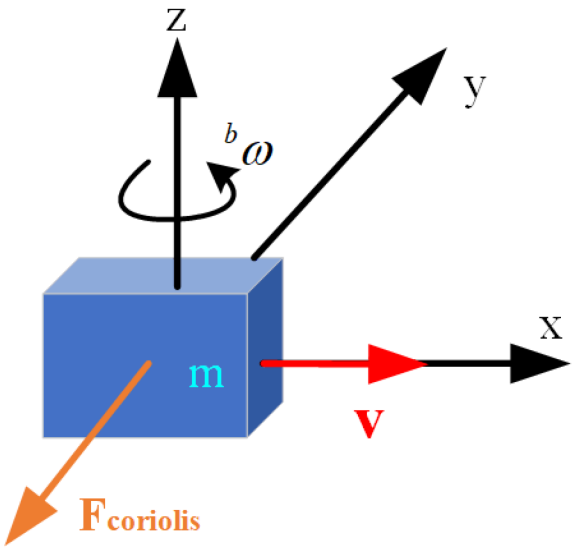


Figure 3.2 Coriolis effect illustration [4].

- Vibratory gyroscopes dominate the MEMS gyroscope market and operate based on the Coriolis effect. A vibrating mass within the sensor experiences a Coriolis acceleration when the device rotates, producing an orthogonal displacement that is orthogonal to both the velocity of the vibrating element and the angular velocity vector. This displacement is typically detected using capacitive sensing and converted into an angular velocity signal [111].

In MEMS gyroscopes, the Coriolis acceleration is expressed as: $a_c = 2(v \times \omega)$ where v is the velocity of the vibrating mass and ω is the angular velocity vector (Figure 3.2). The resulting physical displacement is proportional to the applied rotation rate, and the sensor outputs angular velocity in rad/s [4]. Gyroscope measurements are often expressed in terms of rotation about three orthogonal axes. The most common conventions are Euler angles and Tait–Bryan angles.

Despite their advantages, MEMS gyroscopes exhibit several limitations:

- Bias drift, where integration of minor measurement errors increases orientation error over time.
- Output offset when stationary, which can compromise accuracy in static conditions.
- Limited sensitivity and noise floor compared to higher-grade optical or mechanical gyroscopes.

To mitigate these issues, MEMS gyroscopes are often combined with accelerometers (and sometimes magnetometers) in IMUs. Through sensor fusion algorithms such as complementary filters, Extended Kalman Filters (EKF) [112], or, more recently, deep learning-based fusion methods, it is possible to compensate for drift and achieve accurate, long-term orientation estimates [113].

The proliferation of MEMS gyroscopes has been driven by their low cost, miniaturization, and integration into consumer electronics. They are now embedded in smartphones, wearables, drones, virtual/augmented reality headsets, and automotive safety systems, where they support applications ranging from gesture recognition and navigation to fall detection and gait analysis [109]. In healthcare, combining gyroscope signals with accelerometer data enables better discrimination between falls and activities of daily living, thereby improving the robustness of fall detection systems.

3.1.3 Magnetometer

Magnetometers are sensors that measure a magnetic field's strength and/or direction at a given point in space. Several physical principles can be exploited for magnetic field sensing, including the Hall effect, magnetoresistance, and fluxgate techniques [114]. Among them, Hall-effect magnetometers are the most widely used in MEMS devices because of their compact size, low cost, and compatibility with standard semiconductor processes.

The Hall effect occurs when a magnetic field with a component perpendicular to the current flow is applied to a conductor or semiconductor. The moving charge carriers (electrons or holes) experience a Lorentz force, which deflects them to one side of the material. This charge separation generates a measurable Hall voltage across the conductor, proportional to the magnetic field intensity. Figure 3.3 describes the principle of the single-axis Hall effect sensor.

Modern magnetometers are employed in a wide range of applications [115]:

- Geomagnetic field measurement: Used to determine orientation relative to the Earth's magnetic field, providing absolute heading information (compass functionality).
- Resource exploration: For mineral and oil prospecting and archaeological site surveying.
- Anomaly detection: To detect magnetic disturbances caused by ferromagnetic materials, which help locate pipelines, buried infrastructure, or lost metallic objects.

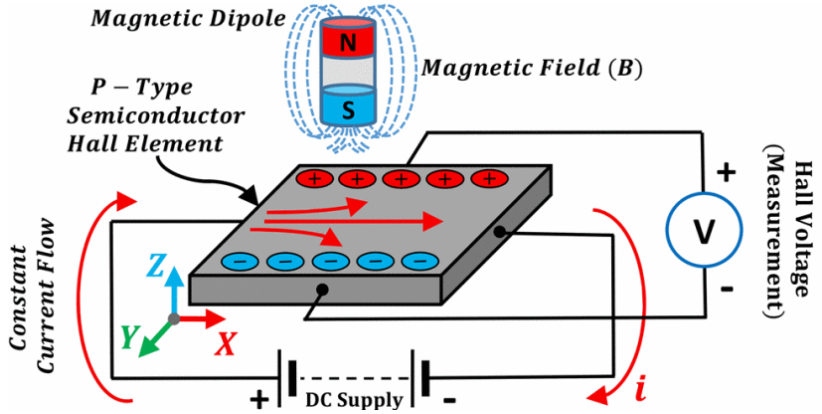


Figure 3.3 Principle of the single-axis Hall effect sensor [5]. The output signal from a Hall effect sensor is a function of the magnetic field density around the device. Three perpendicular axes are required to measure the magnetic field vector.

- Defense and security: In anti-submarine warfare, highly sensitive magnetometers (such as superconducting quantum interference devices, SQUIDs) detect the magnetic signatures of submarines.
- Consumer electronics: Integrated into smartphones, wearables, drones, and VR/AR devices to support navigation, indoor localization, and motion tracking when combined with accelerometers and gyroscopes in an IMU.

In smartphones, magnetometers typically measure the three orthogonal components of the geomagnetic field in microteslas (μT). These measurements are fused with accelerometer and gyroscope data to estimate device orientation. The azimuth angle, or heading relative to magnetic north, is particularly important for pedestrian navigation and augmented reality applications. However, magnetometer readings are influenced by several factors:

- Geographic location: The Earth's magnetic field strength varies globally, from $\sim 25\mu T$ near the equator to $\sim 65\mu T$ near the poles.
- Magnetic declination: The angle between geographic and magnetic north depends on the observer's position and changes over time.
- Environmental interference: Nearby ferromagnetic materials, permanent magnets, electric currents, and electronic devices can distort measurements, producing magnetic noise or hard/soft iron distortions.

Calibration and sensor fusion algorithms are used to overcome these limitations. Hard and soft iron compensation techniques reduce environmental distortions, while fusion with

inertial sensors (accelerometers and gyroscopes) provides robust orientation estimates even in magnetically disturbed environments [116].

Recent advances include the development of anisotropic magnetoresistive (AMR), giant magnetoresistive (GMR), and tunneling magnetoresistive (TMR) magnetometers, which offer improved sensitivity and are increasingly used in biomedical applications such as biomagnetic signal detection (such as magnetocardiography) and in emerging wearable healthcare systems [115, 117].

3.1.4 Summary

An IMU integrates three key sensors, an accelerometer, a gyroscope, and a magnetometer, to provide comprehensive information on movement and orientation. Thanks to their low cost, compact size, and robust performance, IMUs have become a standard component in most wearable and mobile devices we use daily, including smartphones, smartwatches, fitness trackers, personal assistants, mobile PCs, and smart home systems. Beyond consumer electronics, these sensors are increasingly employed in emerging research areas such as fall detection, time-to-impact assessment, and physical activity monitoring. When integrated into personal companion solutions, IMUs enable continuous and unobtrusive monitoring of potential fall events while offering valuable insights into movement patterns during everyday activities. Among the three sensors, the accelerometer is the most widely used for movement analysis, as it measures linear acceleration along three orthogonal axes when worn close to the body. However, since different body segments exhibit distinct motion signatures, many IMU-based solutions are designed for specific sensor placements (such as wrist, waist, ankle or shinbone) [118], which can limit flexibility and generalization in real-world use.

The research involved the interaction and processing of data from multiple inertial sensors. Two accelerometers: an Analog Devices ADXL345 (Analog Devices Inc., Wilmington, Massachusetts, US) and a Freescale MMA8451Q (NXP Semiconductors N.V., Eindhoven, Netherlands), and a Texas Instruments ITG3200 gyroscope (Texas Instruments Inc., Dallas, Texas, US) for analyzing the SisFall dataset [31], and Movella DOT (Movella Inc., El Segundo, California, US), and other wearable sensors for analyzing data collected in the laboratory. Most IMUs are positioned near the pelvis area to be as close as possible to the body's center of mass and balance. This thesis will investigate the feasibility of using inertial sensors in conjunction with signal processing and machine learning techniques to develop fall-detection and time-to-impact-prediction algorithms, and of analyzing signals from wearable devices to assess the risk of falling.

3.2 Biomedical sensors

In the context of fall risk assessment, biomedical sensors detect physiological and psychological states that may precede a fall or be related to a change in fall risk, providing complementary information that is fundamental to kinematic data. This section introduces the key techniques used in this work to characterize a subject's internal state: electromyography (EMG), electrocardiography (ECG), photoplethysmography (PPG), electrodermal activity (EDA), and pupillometry.

3.2.1 Electromyography (EMG)

Electromyography (EMG) is a technique used to record the electrical activity generated by skeletal muscles during contraction and relaxation. When a motor neuron activates a muscle fiber, it produces an action potential, and the summation of many such events can be measured as an EMG signal (Figure 3.4). Depending on the application, EMG can be recorded using either surface electrodes (sEMG), which are non-invasive and placed on the skin over the target muscle, or intramuscular electrodes, which provide higher selectivity but require invasive insertion [119]. The sEMG signal provides insight into muscle activation timing, intensity, and fatigue [46]. It reflects the neural drive to muscles and the mechanical aspects of muscle activity, making it a powerful tool for analyzing motor control, biomechanics, and neuromuscular disorders [120]. To ensure methodological rigor, reproducibility, and comparability with existing literature, the EMG methodologies adopted in this thesis are framed within the recommendations of the Consensus for Experimental Design in Electromyography (CEDE) developed by the International Society of Electrophysiology and Kinesiology (ISEK) [121].

In the context of wearable technologies and fall detection, EMG plays a dual role:

- Movement analysis and activity recognition: EMG helps characterize specific movement patterns, complementing inertial sensor data to improve recognition accuracy. For example, gait phases, sit-to-stand transitions, or compensatory movements can be more reliably identified by combining IMU and EMG features [122].

- Fall prevention and rehabilitation: Monitoring lower limb muscle activity provides insights into balance control and fatigue, two critical risk factors for falls in the elderly. EMG-based biofeedback systems have been proposed to train patients in maintaining postural stability [123].

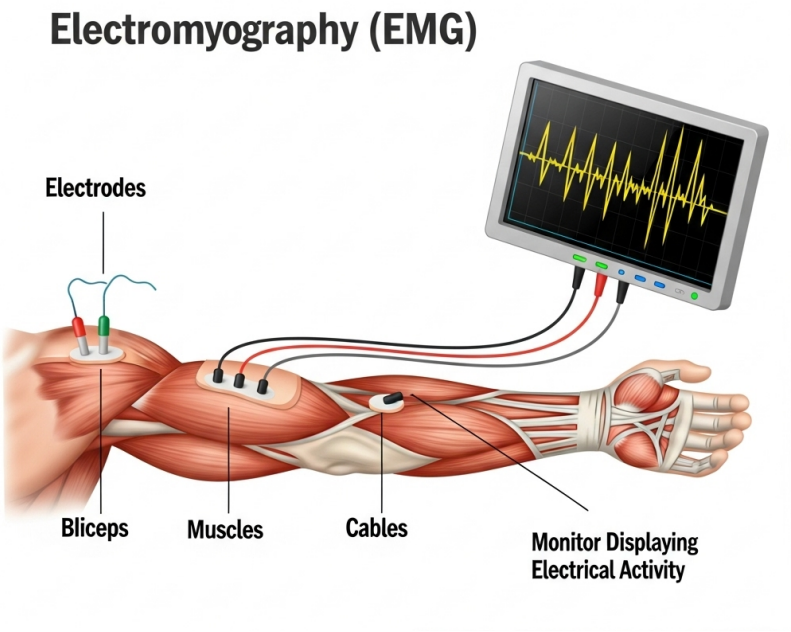


Figure 3.4 Principle of Electromyography (EMG).

Beyond fall-related research, EMG is widely applied in clinical diagnosis of neuromuscular disorders (such as amyotrophic lateral sclerosis, muscular dystrophy, neuropathies); prosthetics and exoskeletons, where EMG signals are used as control inputs for myoelectric devices [124]; sports science and ergonomics monitor muscle fatigue and workload and optimize training; human–computer interaction (HCI), enabling gesture recognition and intuitive control of digital systems.

Despite its versatility, EMG acquisition faces significant challenges. Surface EMG is highly sensitive to motion artifacts, electrode placement, crosstalk between adjacent muscles, and variations in skin impedance. Moreover, signal quality can degrade during dynamic activities, limiting its reliability in free-living environments. Recent advances address these issues through the development of:

- High-density EMG (HD-EMG) arrays provide spatially resolved information on muscle activation.
- Textile-based electrodes and wearable EMG systems improve user comfort and allow long-term monitoring [125].
- Advanced signal processing and machine learning, including deep learning methods for robust feature extraction and pattern recognition in noisy conditions [126].

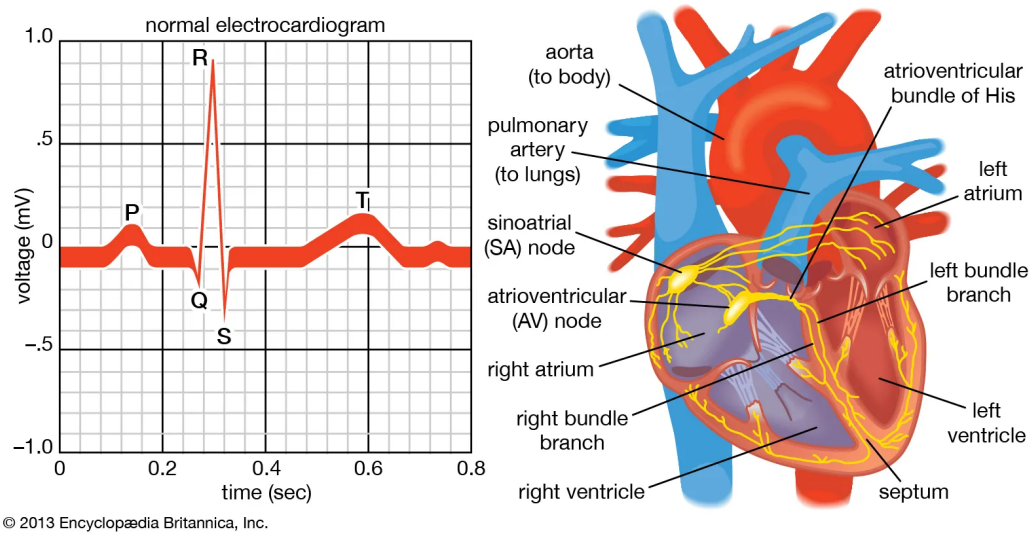


Figure 3.5 (Left) An electrocardiogram showing deflections that reflect the alternating contractions of the heart's atria and ventricles during a heartbeat. (Right) Atria, ventricles, and other components of the heart's impulse conduction system.

Overall, EMG represents a key complementary modality to inertial sensing, providing direct insight into the neuro-muscular mechanisms underlying movement. Its integration into wearable systems holds promise for next-generation fall detection, rehabilitation monitoring, and assistive technologies.

3.2.2 Electrocardiography (ECG)

Electrocardiography (ECG) is a non-invasive technique that records the heart's electrical activity over time by detecting voltage differences generated during cardiac depolarization and repolarization. Typically measured through electrodes placed on the skin, the ECG provides a time-series signal that reflects the cardiac cycle, including the P wave, QRS complex, and T wave, which correspond to atrial depolarization, ventricular depolarization, and ventricular repolarization, respectively [127] as shown in Figure 3.5.

Traditionally, the ECG has been the gold standard for diagnosing cardiovascular conditions, such as arrhythmias, myocardial infarctions, conduction abnormalities, and ischemic events. With the increasing prevalence of chronic cardiovascular diseases and the need for continuous health monitoring, ECG has transitioned from being confined to clinical and hospital environments to being widely used in ambulatory and wearable systems.

Recent advances in wearable technology and textile-based electrodes have enabled the integration of ECG monitoring into smartwatches, chest straps, adhesive patches, and smart

garments [128]. These wearable ECG systems allow long-term, unobtrusive monitoring, which is particularly valuable for:

- Cardiac arrhythmia detection, such as atrial fibrillation, where continuous monitoring significantly improves early diagnosis.
- Stress, fatigue, and activity monitoring, as heart rate variability (HRV) derived from ECG, strongly correlate with autonomic nervous system regulation and overall physical condition [129].
- In fall detection and prevention systems, ECG is often used alongside inertial sensors to provide physiological context (such as identifying syncope-related falls due to cardiac abnormalities) [130].

Despite its advantages, wearable ECG sensing presents several challenges. Motion artifacts, electrode–skin impedance variations, and noise can significantly degrade signal quality, especially during daily activities. To address these issues, research has focused on dry electrodes, conductive textiles, and advanced signal processing techniques, including adaptive filtering and machine learning-based denoising [131]. Recent advances in wearable sensing have led to the emergence of ring-based devices capable of continuously and unobtrusively acquiring ECG signals. Unlike traditional chest-mounted electrodes or adhesive patches, ECG rings typically use dry electrodes on the inner surface of the ring, enabling single-lead ECG acquisition via finger contact. These devices offer a favorable compromise between signal quality, user comfort, and long-term wearability, making them particularly attractive for ambulatory monitoring scenarios [132].

A key advantage of ECG compared to purely inertial measurements is its ability to capture physiological precursors to events such as syncope, dizziness, or cardiovascular stress, which may lead to falls. Combining ECG with inertial sensing in multimodal systems provides a more comprehensive view of biomechanical and physiological states, enabling more reliable fall risk prediction and monitoring solutions [133].

ECG-Derived Respiration (EDR)

The ECG-Derived Respiration (EDR) estimates respiratory activity from an electrocardiogram (ECG) signal, eliminating the need for additional respiratory sensors. Since ECG is already widely collected in clinical and wearable monitoring systems, EDR provides a cost-effective and unobtrusive way to obtain cardiac and respiratory information from a single physiological signal [134, 135].

The principle behind EDR lies in the fact that respiration influences ECG morphology through multiple mechanisms:

- Cardiogenic oscillations: As the thoracic cavity expands and contracts, the heart's and electrodes' relative position changes, modulating ECG amplitude and baseline.
- Respiratory sinus arrhythmia (RSA): Breathing affects autonomic nervous system activity, leading to cyclic variations in R-R intervals.
- Impedance changes: Lung volume variations alter thoracic impedance, indirectly affecting ECG waveforms.

Several computational methods have been developed to extract respiration from ECG [136], including:

- Amplitude-based methods: Tracking respiratory-induced modulations of the R-wave or QRS complex.
- Axis-based methods: Estimating changes in the cardiac electrical axis due to respiration-related electrode displacement.
- Interval-based methods: Using variability in R-R intervals (linked to RSA, Figure 3.6).
- Advanced signal processing and machine learning approaches, such as principal component analysis (PCA), independent component analysis (ICA), and deep learning models, combine multiple ECG features to improve accuracy.

The applications of EDR are diverse:

- Respiratory rate estimation is proper for continuous healthcare, fitness, and home monitoring.
- Sleep studies, where EDR can help identify sleep apnea, hypopnea, and sleep stages without requiring dedicated respiratory belts [137].
- Wearable health monitoring, where EDR enables dual cardiac–respiratory tracking from smartwatches, chest straps, or textile-based electrodes.
- In fall detection and prevention systems, respiration patterns may provide additional context (such as abnormal breathing after a fall or changes in respiratory rhythm due to syncope).

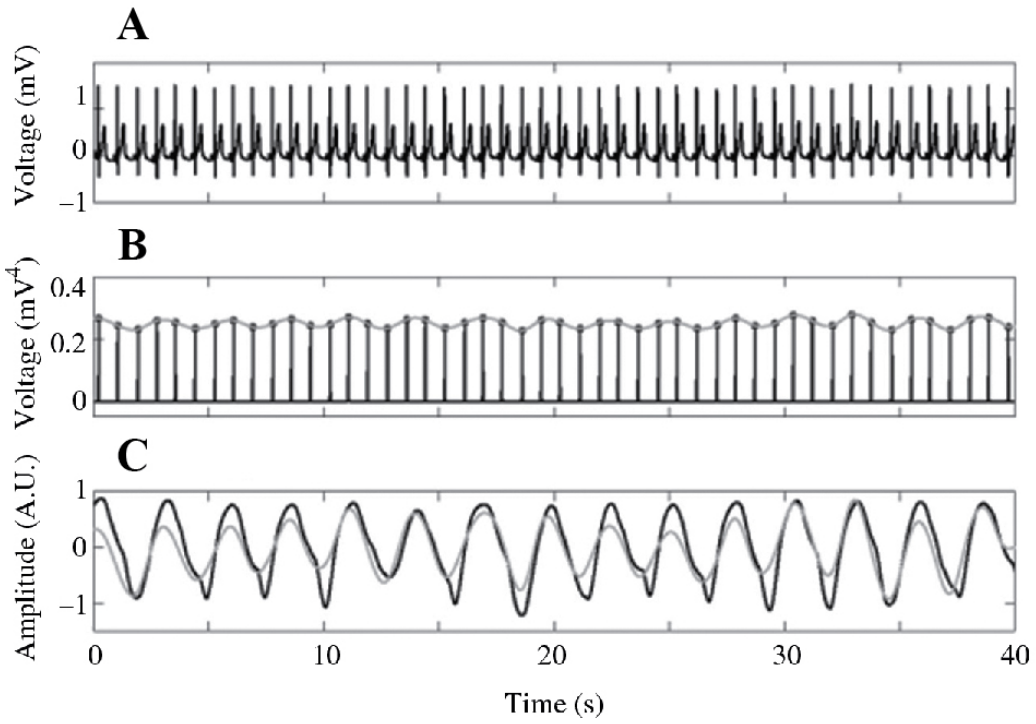


Figure 3.6 Example of ECG-derived respiration estimation (EDR) using central moments. A: ECG, B: fourth-order central moments in black with their maximum (black circles) interpolated using splines (grey curve), C: EDR curve (grey) and reference respiration signal (black) [6].

Despite its advantages, EDR has some limitations. Its accuracy strongly depends on signal quality, electrode placement, and the subject's level of activity. Motion artifacts and noise can significantly degrade the extraction of respiratory features, particularly in ambulatory or free-living conditions. Furthermore, EDR does not provide direct measures of tidal volume or airflow, limiting its use in detailed pulmonary function analysis.

Recent research focuses on enhancing robustness by combining EDR with other modalities (such as PPG-derived respiration) and by leveraging deep learning models trained on multimodal datasets for improved generalization across subjects and activities [138]. These advances suggest that EDR will become a key component in next-generation wearable health systems, enabling continuous, non-invasive, and unobtrusive monitoring of both cardiac and respiratory health.

3.2.3 Photoplethysmography (PPG)

Photoplethysmography (PPG) is an optical technique used to measure changes in blood volume in the microvascular bed of tissue. It emits light (commonly from an LED) into the skin and detects variations in light absorption or reflection using a photodetector. These variations correspond to the pulsatile nature of blood flow, enabling the estimation of heart rate, respiratory rate, blood oxygen saturation (SpO_2), and vascular health indicators [139].

The PPG waveform consists of two primary components [7]. The AC component, a pulsatile signal synchronous with the heartbeat, represents cardiac cycles, while the DC component, a slowly varying baseline influenced by respiration, sympathetic nervous activity, thermoregulation, and vascular tone, Figure 3.7.

PPG is widely used in wearable devices and healthcare monitoring due to its non-invasive, low-cost, and compact nature. Modern smartwatches, fitness trackers, and smartphones commonly incorporate PPG sensors for continuously monitoring cardiovascular and respiratory parameters. In clinical contexts, PPG is routinely applied in pulse oximeters, intensive care units, and sleep monitoring systems [140].

In the context of fall detection and prevention, PPG provides valuable complementary information to inertial sensors and EMG/ECG:

- Physiological monitoring: Sudden drops in heart rate, blood pressure surrogates, or perfusion may precede syncope-related falls.
- Activity and stress assessment: PPG-derived metrics such as heart rate variability (HRV) and pulse transit time (PTT) can reveal autonomic nervous system activity, stress levels, and fatigue, which are relevant fall risk factors [141].

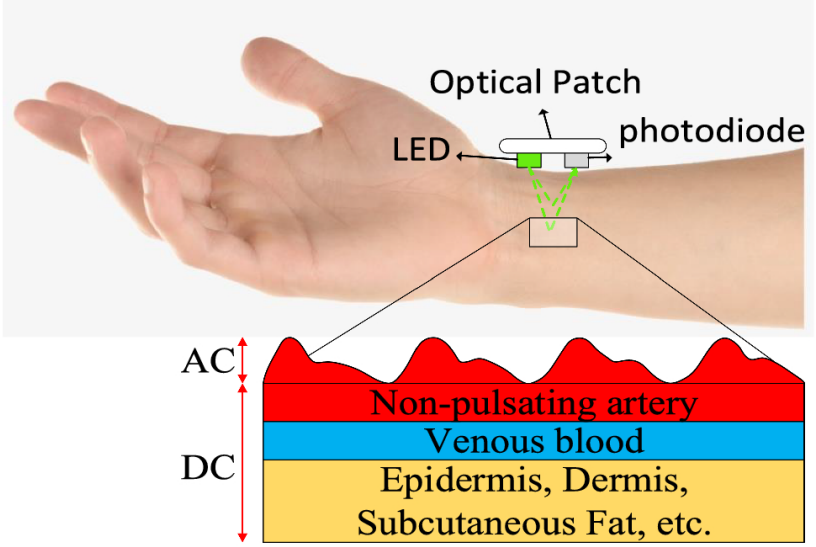


Figure 3.7 Principle of the PPG signal using a photodetector and an LED [7].

- Multimodal integration: Combining PPG with IMUs enhances detection accuracy by correlating hemodynamic changes with movement patterns, providing a more holistic view of the fall event.

Despite its versatility, PPG signals are highly sensitive to motion artifacts, ambient light interference, and the quality of sensor-skin contact. This makes reliable acquisition challenging in free-living environments or during vigorous activity. To address these limitations, recent research has focused on:

- Motion artifact reduction using adaptive filtering, blind source separation, and deep learning approaches [142].
- Multichannel PPG and imaging PPG techniques improve robustness by capturing spatially distributed signals.
- Flexible and textile-integrated PPG sensors, enabling comfortable long-term monitoring in wearable systems [143].

With ongoing advances in sensor design and signal processing, PPG is evolving from a simple pulse monitoring tool into a powerful modality for continuous cardiovascular health assessment and fall-risk prediction. Its seamless integration into consumer wearables underscores its potential as a cornerstone technology in personalized healthcare and ambient assisted living systems.

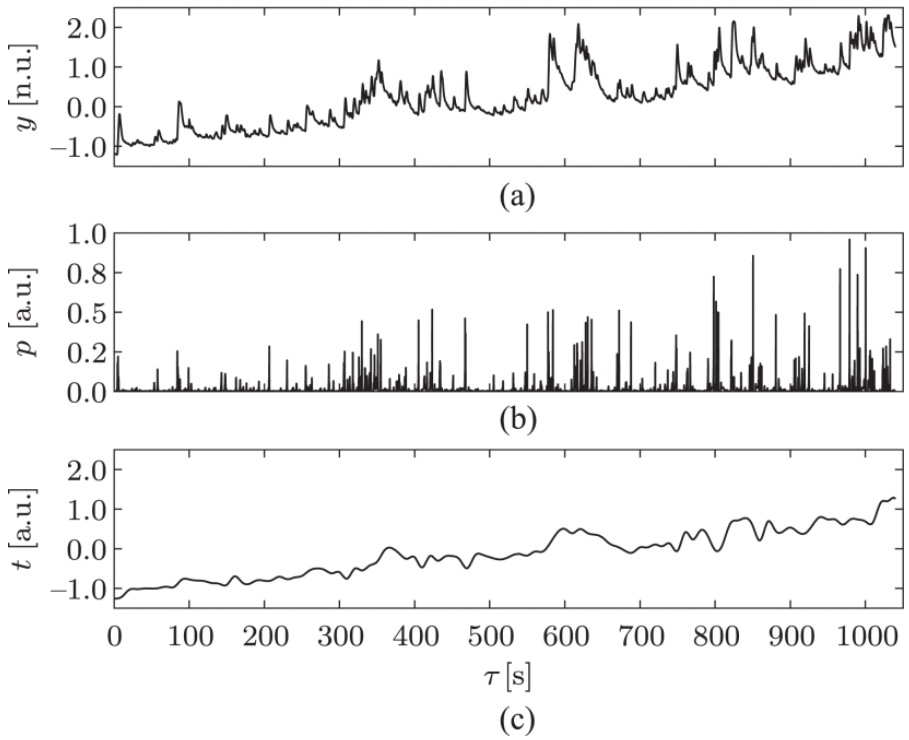


Figure 3.8 (a) Raw EDA signal y, normalized with Z score. (b) Sparse phasic driver component p. (c) Slow tonic component t. [8].

3.2.4 Electrodermal activity (EDA)

Electrodermal activity (EDA), also known as galvanic skin response (GSR) or skin conductance, measures changes in the skin's electrical properties resulting from variations in sweat gland activity. These changes are primarily driven by the sympathetic branch of the autonomic nervous system (ANS), making EDA a valuable non-invasive proxy for physiological arousal, stress, emotional states, and cognitive workload [144].

EDA is typically measured by applying a small, imperceptible electrical current across two electrodes placed on the skin, most commonly on the fingers, palms, or soles, where sweat glands are highly concentrated. The signal reflects variations in skin conductance resulting from eccrine sweat gland activity, which increases during sympathetic nerve activation. The EDA signal is generally decomposed into two components: Tonic component (Skin Conductance Level, SCL), Figure 3.8(c), a slow-varying baseline that reflects general arousal or stress level; and Phasic component (Skin Conductance Response, SCR), Figure 3.8(b), rapid fluctuations in conductance associated with discrete stimuli, stressors, or emotional events [145].

With the rise of wearable devices, EDA sensors have been miniaturized and embedded into wristbands, smartwatches, and finger-based devices (such as Empatica E4, Fitbit Sense). These systems enable continuous monitoring in daily life settings and have been applied to [8]:

- Stress and mental health monitoring: tracking physiological responses to psychological stress, anxiety, or depression.
- Human–computer interaction (HCI): adapting systems based on user arousal or engagement levels.
- Sleep studies: assessing nocturnal sympathetic activity and sleep quality.
- Fall detection and prevention: while less explored than IMUs or ECG, EDA can provide valuable context by identifying stress-related precursors to falls (such as panic, anxiety, or physiological overactivation) or characterizing post-fall states such as distress.

Despite its potential, EDA acquisition faces several challenges. The signal is susceptible to motion artifacts, electrode placement, environmental humidity, and temperature. Moreover, inter-individual variability in sweat gland density and responsiveness complicates interpretation. Recent research addresses these issues by applying machine learning models for signal cleaning and interpretation, as well as multimodal sensor fusion (EDA combined with PPG or accelerometry). Additionally, novel electrode materials, such as dry, textile-integrated, or flexible conductive polymers, are being explored to improve long-term wearability [146].

Overall, EDA is a crucial modality in the expanding ecosystem of wearable biosensors. While not directly measuring biomechanical activity, such as IMUs or EMG, it provides unique insights into psychophysiological states. Its integration with inertial and cardiovascular sensors provides a more comprehensive understanding of the physical and emotional factors contributing to falls and activity-related risks, supporting the development of next-generation personalized healthcare and preventive systems.

3.2.5 Pupillometry

Pupillometry measures pupil size and reactivity over time, providing a non-invasive window into the functions of the neurological and autonomic nervous system (ANS). The pupil is controlled by two antagonistic muscles: the sphincter pupillae, innervated by the parasympathetic system and responsible for constriction, and the dilator pupillae, innervated by the sympathetic system and responsible for dilation. Changes in pupil diameter occur in response

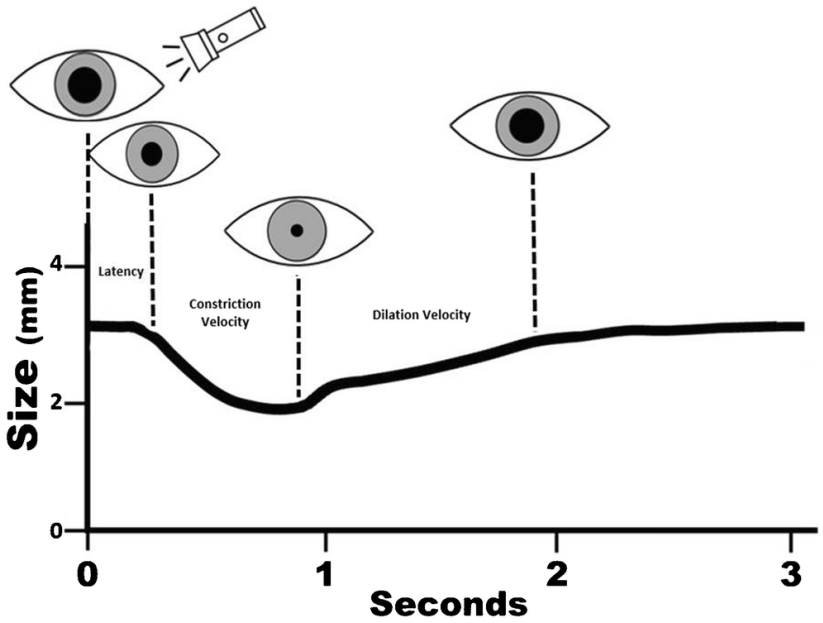


Figure 3.9 Key components of the pupillary light reflex [9].

to light intensity (the pupillary light reflex, Figure 3.9) and cognitive load, emotional arousal, and stress [147].

Modern pupillometry relies on infrared eye-tracking systems that enable precise and unobtrusive measurement of pupil diameter, often with sampling rates ranging from 30 Hz to over 1000 Hz. Unlike earlier clinical methods based on manual observation or simple penlights, infrared pupillometry enables continuous and high-resolution monitoring of dynamic pupil changes. These advances have led to broad adoption in neuroscience, psychology, and human–computer interaction research.

Pupillometric features can be categorized into [148]:

- Static measures: baseline pupil diameter, often linked to tonic arousal or cognitive effort.
- Dynamic measures: latency, amplitude, and velocity of constriction/dilation responses to stimuli.
- Task-evoked pupillary responses (TEPRs): transient changes in pupil size reflecting mental effort, memory load, or emotional states.

Applications of pupillometry span both clinical and non-clinical domains:

- Neurology and critical care: automated pupillometers are used in intensive care units to assess brainstem function, detect early signs of neurological deterioration, and guide prognosis after brain injury [149].
- Cognitive workload and fatigue assessment: pupillometry is a reliable marker of mental effort. It has been applied to monitor driver fatigue, pilot workload, and operator performance in safety-critical environments [150, 47].
- Affective computing and stress detection: pupil dilation correlates with emotional arousal, making pupillometry useful in human–computer interaction, adaptive systems, and psychological research.
- Fall detection and prevention research: while less direct than IMUs or cardiovascular signals, pupillometry can provide complementary information on cognitive load, fatigue, or stress states that increase fall risk, especially in elderly populations or patients with neurodegenerative diseases.

Recent advances focus on wearable and mobile pupillometry systems. Miniaturized infrared sensors and camera-based approaches are being integrated into smart glasses, VR/AR headsets, and smartphones, enabling the tracking of pupil dynamics outside laboratory environments [151]. In parallel, machine learning models are being increasingly used to process raw pupil signals, correct for artifacts such as blinks or head motion, and classify cognitive or affective states in real-time.

Despite its promise, pupillometry faces several challenges: the signal is sensitive to ambient lighting, gaze direction, and individual variability. Moreover, interpreting pupil changes requires a careful experimental design to disentangle the overlapping influences of light, cognition, and emotion. Nonetheless, the integration of pupillometry with other physiological sensors is emerging as a powerful multimodal strategy for holistic monitoring of both physical and cognitive contributors to fall risk and broader health outcomes.

Virtual Reality (VR)

Virtual Reality (VR) refers to computer-generated, immersive environments that simulate real or imagined scenarios, enabling users to interact with and experience a digital world in three dimensions. Typically delivered through head-mounted displays (HMDs), VR systems integrate visual, auditory, and sometimes haptic feedback to create a sense of presence, often described as the subjective feeling of "being there" [152]. The recent proliferation of consumer-grade devices such as the Meta Quest, HTC Vive, and Sony PlayStation VR,

alongside lightweight mobile-based solutions, has significantly expanded VR's accessibility beyond research laboratories into healthcare, education, and rehabilitation. VR has become a robust simulation, training, and rehabilitation tool in biomedical engineering and healthcare. Its applications span:

- Neurorehabilitation: VR-based therapy is used to improve motor function, balance, and coordination in patients recovering from stroke, Parkinson's disease, or traumatic brain injury [153].
- Cognitive and psychological therapy: VR has been successfully employed to reduce anxiety, manage chronic pain, and treat phobias by providing controlled exposure in a safe environment [154].
- Surgical and clinical training: VR enables healthcare professionals to practice complex procedures in risk-free simulated environments, enhancing skill acquisition and patient safety.
- Fall prevention and balance training: immersive environments can safely expose users to challenging postural scenarios, such as obstacle negotiation or uneven terrains, which enhance stability and adaptability without real-world risk.

VR's relevance to fall detection and prevention lies in its capacity to simulate controlled perturbations and complex mobility scenarios that would be unsafe to replicate in real life. By integrating VR with wearable inertial sensors, motion capture systems, or physiological monitoring, researchers can analyze motor responses, balance control, and stress reactivity in ecologically valid yet controlled conditions. This combination provides a valuable framework for assessing individual fall risk and testing personalized preventive interventions. Recent advances in immersive technologies are further expanding the role of VR in healthcare. For example:

- Augmented Reality (AR) and Mixed Reality (MR) blend virtual stimuli with real-world environments, enhancing ecological validity in rehabilitation and monitoring tasks.
- Haptic feedback systems provide tactile stimulation to improve motor learning and balance training.
- VR combined with artificial intelligence (AI) enables adaptive, real-time adjustment of training environments based on a user's physiological state or performance [155].

Despite its promise, the adoption of VR in clinical and daily life applications faces several challenges. Extended use of HMDs can induce cybersickness (nausea, dizziness, or disorientation) and may limit applicability for older adults or patients with vestibular dysfunction. Moreover, cost, accessibility, and the need for technical expertise continue to be barriers to large-scale deployment in community or home-based settings. Nevertheless, as hardware becomes more affordable and ergonomically optimized, VR is increasingly positioned as a transformative tool for rehabilitation, fall risk assessment, and the development of prevention strategies.

3.2.6 Summary

An integrated approach combining these biomedical sensors is crucial for a comprehensive assessment of fall risk. While IMUs provide biomechanical (kinematic) data related to the fall event itself, biosensors capture the physiological and cognitive states (such as fatigue, cardiovascular stress, and cognitive overload) that often precede the loss of balance. This multimodal data fusion enhances predictive accuracy and is essential for developing the next generation of personalized and predictive fall prevention systems.

The project involved the interaction and processing of data from various wearable biomedical sensors. Four devices worn on different parts of the body are used to collect the data. The Biosignalsplux platform (PLUX Wireless Biosignals S.A., Lisbon, Portugal) is positioned behind the subject to collect EMGs, ECGs, and breathing data. The Empatica E4 smartwatch (Empatica Inc., Cambridge, MA, USA), worn on the wrist, records blood volume pulse (BVP), electrodermal activity (EDA), temperature, and acceleration (ACC). The Polar H10 chest strap (Polar Electro Oy, Kempele, Finland) records ECG and acceleration (ACC), while the Vive Pro 2 headset (HTC Corp., Taoyuan, Taiwan) records pupillometry data. A more detailed description is available in Section 9.3.

3.3 Conclusions

The sensors presented in this chapter were selected because each contributes a unique and necessary perspective on the mechanisms underlying falls. Inertial sensors describe how the body moves, capturing kinematic patterns associated with balance loss, gait instability, or impact. Biomedical sensors monitor the body's internal functions, revealing physiological states, such as fatigue, stress, reduced attention, or cardiovascular instability, that often precede or accompany a fall event. Together, these modalities form a comprehensive and

multidimensional representation of both movement and underlying condition, enabling machine learning models to infer not only whether a fall is occurring but also the risk that a fall will occur.

The philosophy guiding their use in this thesis is grounded in three principles:

- Wearability: Sensors must be feasible for real-world deployment, being wearable, lightweight, unobtrusive, and capable of operating in dynamic environments.
- Complementarity: No single sensor is sufficient to capture the complexity of STF. Combining biomechanical and physiological information leads to more robust and interpretable models.
- Personalization: Because physiological and behavioral signatures vary across individuals, sensors should capture subject-specific patterns that can be leveraged to build personalized models.

By integrating these diverse sensing technologies, the thesis moves from traditional and reactive fall detection toward proactive prediction and individualized risk estimation.

The next part of the thesis builds upon the technological foundations established in this section. Chapter 4 introduces the questionnaires used to capture subjective human factors, such as workload, comfort, presence, and perceived exertion, which complement the sensor-based measurements. Together, these objective and subjective data streams form the basis for the multimodal framework developed in subsequent chapters, enabling a holistic and human-centered approach to fall detection, forecasting, and prevention.

Chapter 4

Fundamentals of Questionnaires

In the comprehensive assessment of human factors, usability, and user experience in complex systems, particularly those involving advanced technology like virtual reality (VR) and wearable sensors, objective performance metrics are often insufficient. While measurements such as task completion time, classification accuracy, or physiological signals provide critical quantitative data on performance and internal state, they fail to capture the user's subjective experience.

Questionnaires are thus essential complementary tools, offering a window into the user's perception and mental state. They provide reliable, standardized methods for quantifying typically non-observable human factors, which are crucial for assessing the viability and impact of the developed fall safety framework.

The questionnaires utilized throughout this work are broadly categorized to evaluate four key aspects:

- Perception Evaluation: To quantify the user's subjective emotional, motivational, and cognitive responses to tasks and stimuli (such as effort, frustration, and intrinsic motivation).
- Ergonomics Evaluation: To assess human factors, such as anxiety, stress, and physical/visual discomfort, which are vital for determining the long-term wearability and safety of the systems.
- Usability Evaluation: To measure how easy and efficient the virtual reality environment and sensor systems are to use from the user's perspective.
- Virtual Reality Evaluation: To specifically assess the quality of the immersive experience, including the sense of presence and embodiment within the VR environment.

By integrating subjective data from these sources with objective data from kinematic and biosensor measurements, this thesis provides a comprehensive and user-centric evaluation of fall prevention and risk estimation methodologies.

4.1 Perception Evaluation

The following questionnaires are considered for evaluating the subject perception: the After-Scenario Questionnaire (ASQ, Section 4.1.1), the Intrinsic Motivation Inventory (IMI, Section 4.1.2), the NASA-TLX (Section 4.1.3), and the Self-Assessment Manikin (SAM, Section 4.1.4)

4.1.1 After-Scenario Questionnaire (ASQ)

The After-Scenario Questionnaire (ASQ) [156] is one of the most widely adopted instruments for capturing post-task usability feedback in human–computer interaction. Unlike global usability measures administered only once at the end of an evaluation, the ASQ is completed immediately after each task, allowing researchers to capture users' task-specific impressions with minimal recall bias. The questionnaire consists of just three items, each rated on a 7-point Likert scale ranging from "strongly agree" to "strongly disagree," with an additional "not applicable" option (Figure A.2). These items assess three critical dimensions of user satisfaction: the ease of completing the task, the adequacy of the time required, and the quality of the support information (such as instructions, prompts, or feedback from the system).

The ASQ's brevity and simplicity are key to its enduring relevance. It places a minimal cognitive burden on participants while providing diagnostic insights into task-level usability issues. Psychometric evaluations have demonstrated that the ASQ exhibits strong reliability, sensitivity, and validity [157]. Importantly, its scores have been shown to correlate with both objective performance metrics (such as task completion time and error rates) and broader usability instruments, such as the System Usability Scale (SUS, see Section 4.3.2) [158]. The ability of the ASQ to highlight scenario-dependent variations makes it particularly valuable in multi-task usability studies, where user satisfaction may differ across task types or levels of complexity.

Over the years, the ASQ has been applied beyond its original focus on software evaluation. It has been successfully used in the assessment of mobile applications, wearable devices, and healthcare technologies, as well as in virtual and augmented reality environments.

For example, in healthcare applications, the ASQ can provide insights into how users perceive system clarity and task efficiency in safety-critical contexts, such as fall detection or rehabilitation training. Similarly, in immersive VR applications, the ASQ has been adapted to evaluate task-specific satisfaction during activities such as navigation, calibration, or interaction with virtual objects, where "support information" may include real-time feedback from sensors or haptic guidance, rather than traditional help menus.

Despite its strengths, the ASQ's brevity also constitutes a limitation, as it cannot capture broader dimensions of the user experience, such as emotional engagement, cognitive workload, or trust in automation. Consequently, it is often used with complementary measures, such as NASA-TLX for workload (see Section 4.1.3) or presence questionnaires for immersive environments [159]. Still, its efficiency, diagnostic precision, and adaptability ensure that the ASQ remains a cornerstone of usability evaluation, particularly when researchers need to identify task-level bottlenecks and collect user feedback quickly across multiple scenarios.

4.1.2 Intrinsic Motivation Inventory (IMI)

The Intrinsic Motivation Inventory (IMI) [160] is a multidimensional self-report questionnaire developed within the framework of Self-Determination Theory (SDT) [161] to assess individuals' subjective experiences during a specific activity. Unlike global motivation measures, the IMI captures task-specific constructs such as Interest/Enjoyment, Perceived Competence, Effort/Importance, Value/Usefulness, Perceived Choice, and Pressure/Tension; some versions also include Relatedness [162].

The Interest/Enjoyment subscale is often considered the primary indicator of intrinsic motivation, while the others assess complementary dimensions of task engagement. Items are rated on a Likert scale and scored by averaging across subscales, with some items reverse-scored. Psychometric studies confirm the IMI's reliability and validity across various domains, including sports, education, clinical research, and interactive technologies [160]. Internal consistency is generally high, although some subscales (such as Interest/Enjoyment) may vary depending on context [163]. Shorter versions, such as the IMI-SR for schizophrenia research, have also shown strong reliability and test–retest stability [164].

In human-computer interaction and VR studies, the IMI is widely used to complement objective performance measures by capturing motivational quality and user experience. For instance, it helps evaluate whether participants felt competent, autonomous, or pressured during system interaction, offering insight into long-term adoption potential [165].

In summary, the IMI is a validated, flexible tool for measuring intrinsic motivation across diverse experimental settings, including sensor-based and VR studies, where understanding performance and subjective engagement is essential.

4.1.3 NASA Task Load Index (NASA-TLX)

The NASA Task Load Index (NASA-TLX) [166] is one of the most widely used subjective tools for assessing perceived workload across different domains, including aviation, healthcare, human–computer interaction, and virtual reality. Unlike performance-based measures, the NASA-TLX captures the cognitive, physical, and emotional demands experienced by participants during task execution, providing insights into how users perceive and manage workload.

The instrument evaluates workload across six subscales (Figure A.3): Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. Participants rate each subscale on a 20-point scale, typically presented as a continuous line or Likert-type slider. The standard method involves a pairwise weighting procedure, where participants indicate which dimensions contributed more to their workload, enabling the computation of a weighted global score. However, in many studies, researchers use the simpler unweighted average across subscales, which has been shown to provide comparable validity [167, 168].

The psychometric properties of NASA-TLX have been extensively validated. It demonstrates good construct validity, consistently distinguishing between tasks of varying complexity and demand [169, 170]. The tool is flexible, as subscales can be administered individually if only specific workload dimensions are relevant, and adaptations are available for computerized, paper-based, and mobile formats.

In human-computer interaction and VR contexts, NASA-TLX is particularly valuable because it complements usability and performance metrics by revealing cognitive strain, frustration, or effort that may not be observable in task outcomes. For instance, two VR interaction techniques might yield similar task completion times. However, one may impose significantly higher mental demands, as reflected in NASA-TLX scores, which are critical for system design and evaluation.

In summary, NASA-TLX remains a robust, flexible, and widely validated tool for capturing subjective workload. Its ability to provide a multidimensional profile of task demands makes it especially useful in experiments involving complex systems, sensor-based applications, and immersive environments.

4.1.4 Self-Assessment Manikin (SAM)

The Self-Assessment Manikin (SAM) [171] is a widely used, non-verbal pictorial instrument for measuring an individual's emotional response to stimuli along three core affective dimensions: valence (pleasure), arousal, and dominance (often called the PAD model), see Figure A.1. SAM was developed to provide a quick, culture-free means of capturing emotional reactions without relying on language-based scales, making it useful in cross-cultural or non-verbal settings.

SAM presents respondents with three rows of stylized figures (manikins) for each dimension. For valence, the sequence ranges from a smiling, happy figure to a frowning, unhappy figure; for arousal, from a sleepy or calm figure to an excited or alert figure; and for dominance, from a small, submissive figure to a significant, in-control figure. Participants select the figure that best matches their subjective emotional feeling on a 9-point scale for each dimension. The resulting scores yield a three-dimensional affective profile [172, 171].

One of SAM's major strengths is its efficiency, which can typically be completed in under 15 seconds, thereby minimizing participant fatigue, especially in experiments involving multiple stimuli [172]. Because it does not depend on verbal descriptors, SAM reduces linguistic and translation biases, making it suitable for diverse populations, including children or individuals with limited language proficiency. Empirically, SAM's valence and arousal scores correlate strongly with scores from the semantic differential method. At the same time, dominance shows a somewhat lower but acceptable correspondence [172, 173].

SAM has been applied in various domains, including psychophysiology, advertising research, and media studies, and is increasingly used in virtual reality and affective computing. SAM is often used after exposure to visual, auditory, or multisensory stimuli in VR or interactive systems experiments to assess users' emotional reactions to environments or system components [174].

However, SAM is primarily a short-term, post-stimulus measure, capturing the immediate affective response. However, it does not probe deeper constructs such as emotional regulation, long-term valence changes, or complex cognitive–emotional interactions. Also, although the pictorial format helps minimize language bias, interpreting dominance or subtle arousal differences can be more subjective and less intuitive for participants.

In the design of studies involving sensors, VR, or fall detection systems, SAM can assess how users feel after interacting with a device or environment (such as how pleasant, exciting, or controlling the experience was). Combined with performance metrics (accuracy, response time) and usability/motivation scales, it enriches your understanding by providing emotional context to quantify the affective dimension of user experience.

4.2 Ergonomics Evaluation

The Anxiety Sensitivity Index (ASI, Section 4.2.1), Subjective Symptoms Questionnaire (SSQ, Section 4.2.2), and Visual Comfort Questionnaire (VCQ, Section 4.2.3) are used to evaluate the human factors and ergonomics.

4.2.1 Anxiety Sensitivity Index (ASI)

The Anxiety Sensitivity Index (ASI), initially developed in 1986 [175], is a widely used self-report measure designed to assess the extent to which individuals fear anxiety-related sensations due to beliefs that these sensations have harmful consequences. The concept of anxiety sensitivity (AS) reflects a dispositional tendency to interpret physiological symptoms of anxiety, such as palpitations, dizziness, or shortness of breath, as signals of imminent physical, psychological, or social harm. Unlike general trait anxiety, which reflects the overall tendency to experience anxious affect, anxiety sensitivity is considered a specific risk factor for the development and maintenance of anxiety disorders, particularly panic disorder [176].

The original ASI consisted of 16 items, each rated on a Likert-type scale, covering three main domains: physical concerns, such as fear of somatic symptoms like chest pain; cognitive concerns, including fear of losing mental control; and social concerns, including fear of publicly observable anxiety symptoms. Respondents indicate how much they agree with statements about their reactions to anxiety-related sensations, producing a total score that reflects overall anxiety sensitivity as well as subscale scores that highlight specific concern domains.

Research has shown that high ASI scores are associated with an increased likelihood of panic attacks, phobic avoidance, and broader anxiety pathology [177, 178]. Furthermore, anxiety sensitivity has been linked to heightened emotional reactivity in stressful contexts and to the maintenance of maladaptive coping strategies. Due to these associations, the ASI and its subsequent revisions (ASI-3 [179], Figure A.4) have been widely utilized in clinical research and applied settings to identify individuals at risk, inform treatment planning, and assess intervention outcomes.

The strength of the ASI lies in its ability to capture a specific vulnerability construct that predicts the onset and severity of anxiety-related problems beyond general negative affectivity. However, its reliance on self-report makes it sensitive to response biases, and cross-cultural studies have highlighted the need to consider cultural interpretations of bodily sensations in assessing anxiety sensitivity [180]. Nonetheless, the ASI remains a cornerstone

instrument in the assessment of anxiety vulnerability, widely applied in both research and clinical practice.

4.2.2 Subjective Symptoms Questionnaire (SSQ)

The Subjective Symptoms Questionnaire (SSQ) [181, 182], also referred to as the User Fatigue Survey (UFS) in [48], is a self-report instrument designed to assess the subjective experience of visual fatigue and associated discomfort during or after visually demanding tasks. Unlike broader measures of cybersickness or usability, this questionnaire specifically targets the ocular and systemic symptoms that arise when binocular vision is strained, such as during stereoscopic viewing, VR exposure, or intensive near-work tasks.

The SSQ comprises seven items divided into two domains: the first addresses eye-related symptoms (such as tiredness, clarity of vision, and sensations of pain or dryness), while the second captures physical and psychological discomfort (such as back and neck fatigue, headache, and sleepiness). Each symptom is rated on a five-point Likert scale, ranging from "none" to "severe," allowing researchers to calculate subscores for eye symptoms and general discomfort, as well as an overall symptom score. By administering the questionnaire immediately before and after a visual task, the change in score provides a sensitive measure of the incremental fatigue induced by the activity.

This instrument has proven particularly useful in clinical and experimental contexts where objective measures of binocular function, such as fusion maintenance or vergence stability, are complemented by participants' perceptions of fatigue. Demonstrating that an increase in subjective symptom scores correlated with reduced binocular fusion maintenance in individuals with intermittent exotropia, highlighting the questionnaire's ability to link subjective discomfort with functional impairment. Beyond ophthalmology, the SSQ has been applied in virtual reality and vision science research to evaluate the tolerability and usability of immersive systems, making it a valuable tool for patient studies and technology assessment [181].

While the SSQ is concise and task-specific, its reliance on self-reported perceptions can be influenced by expectation effects and individual differences in symptom awareness. Nevertheless, its efficiency and sensitivity to short-term changes in ocular strain make it a reliable complement to objective measurements in studies of visual workload, fatigue, and user comfort.

4.2.3 Visual Comfort Questionnaire (VCQ)

The Visual Comfort Questionnaire (VCQ) [183] is a subjective and self-report instrument used to measure the immediate visual and systemic discomfort experienced by a user during or after exposure to demanding visual tasks, such as those encountered in Virtual Reality or Head-Mounted Display (HMD) environments.

While similar instruments focus on general usability or motion sickness, the VCQ targets the unique sensory and cognitive symptoms associated with sustained or challenging visual input in immersive systems explicitly. The typical version of the VCQ employs a 10-point rating scale, ranging from 'None' to 'Severe' with scores of 0-9, to assess the perceived intensity of a specific set of symptoms. The symptoms evaluated typically encompass both local (ocular) and systemic discomfort, including: Visual Discomfort, dry eyes, eye irritation, Difficulty focusing, Visual fatigue, Headache, Dizziness, Nausea, and General Tiredness.

By quantifying the perceived severity of these symptoms, the VCQ offers developers and researchers an essential tool for assessing the tolerability and ergonomic quality of VR systems and HMDs. High scores on the VCQ often indicate potential issues related to display quality, latency, visual-vestibular mismatch, or overall system design, which can significantly impact user performance, long-term adoption, and, in the context of this work, the ecological validity of the VR environment for fall-risk assessment.

4.3 Usability Evaluation

The Single Usability Metric (SUM, Section 4.3.1), the System Usability Scale (SUS, Section 4.3.2), and the Usability Metric for User Experience (UMUX, Section 4.3.3) are used to evaluate the usability of the virtual reality environment.

4.3.1 Single Usability Metric (SUM)

The Single Usability Metric (SUM) [184] was developed as a standardized method to consolidate multiple usability aspects into a single, interpretable score. Usability evaluation traditionally involves collecting diverse metrics, such as task completion rates, error frequencies, task times, and subjective satisfaction, which provide valuable but often fragmented insights. SUM addresses this challenge by integrating these measures into one composite index, facilitating more direct comparisons across systems, iterations, or design alternatives.

The methodology behind SUM is straightforward yet rigorous. The four component measures (completion rate, error rate, time, and satisfaction) are first standardized by converting

them into a z-score relative to a defined benchmark or specification limit. These standardized values are then averaged to produce a single number that reflects the system's overall usability. This approach enables evaluators to capture the effectiveness, efficiency, and satisfaction dimensions of usability (as defined by ISO 9241-11) within a unified framework. Higher SUM values indicate better performance across these dimensions, providing stakeholders with a concise and easily interpretable summary of the results.

One of the main strengths of SUM is its communicative power. While detailed analyses remain crucial for identifying usability problems, decision-makers and non-specialist audiences often prefer a single metric summarizing the user experience. SUM therefore serves as an effective tool in benchmarking studies, competitive product evaluations, and iterative design processes, where a precise measure of progress is essential.

However, this simplification comes with limitations. Because SUM collapses multiple metrics into a single score, it can sometimes mask important details. For example, a system might achieve a high SUM even if user satisfaction is low, or vice versa. Moreover, the method assumes equal weighting of its four components, which may not be appropriate in domains where certain factors (such as error prevention in healthcare applications) carry more importance. The reliance on specification limits and assumptions about data distribution further highlights the need for careful application of the method.

In summary, SUM should be viewed as a complementary measure rather than a replacement for detailed usability analysis. It provides a powerful way to communicate overall usability performance and compare systems at a glance. However, it is most effective when used in conjunction with a deeper examination of the individual metrics it aggregates.

4.3.2 System Usability Scale (SUS)

The System Usability Scale (SUS) [185] is a widely used, reliable, and quick-to-administer tool designed to assess the usability of interactive systems from the user's perspective. It consists of 10 items, alternating between positive and negative statements, rated on a five-point Likert scale from "Strongly Disagree" to "Strongly Agree." Responses are converted to a 0–100 scale, where higher scores indicate better perceived usability. SUS provides a holistic measure of usability, capturing aspects such as learnability, efficiency, and satisfaction, without requiring task-specific questionnaires.

The SUS is valued for its simplicity, versatility, and broad applicability. It has been used extensively across various systems, including web applications, mobile apps, software interfaces, and wearable devices [186]. Its standardized scoring allows for benchmarking

across studies and technologies, making it particularly useful for comparing design alternatives or tracking usability improvements over iterative development cycles. Studies have demonstrated its strong reliability and validity across different domains, confirming that SUS scores correlate well with other usability measures while remaining quick and easy to administer [187].

Despite its strengths, SUS is subjective and does not provide diagnostic information about specific usability problems. Consequently, it is often combined with qualitative methods, such as user observations, think-aloud protocols, or semi-structured interviews, to gain a deeper understanding of user experience [157]. The balance of brevity, robustness, and interpretability has made SUS a gold standard in academic research and industry usability testing. Its widespread adoption allows for meaningful comparisons and meta-analyses across studies.

4.3.3 Usability Metric for User Experience (UMUX)

The Usability Metric for User Experience (UMUX) [188] is a brief, reliable questionnaire designed to assess the perceived usability of interactive systems from the user's perspective. UMUX was developed as a shorter alternative to the System Usability Scale (SUS), retaining high correlation with SUS scores while reducing the number of items from ten to just four. Each item is rated on a seven-point Likert scale, with statements covering the effectiveness, efficiency, and satisfaction of system use. Scores are transformed into a 0–100 scale, facilitating comparison with other usability metrics and benchmarking against established norms.

UMUX is valued for its simplicity, brevity, and strong psychometric properties. It has been shown to produce reliable usability assessments across a variety of systems, including web applications, mobile apps, and interactive devices, making it especially suitable for contexts where time constraints or user fatigue limit the use of longer questionnaires [188, 189]. Despite its brevity, UMUX effectively captures the essential elements of user experience and perceived system usability, enabling effective comparisons across different systems or iterations of a design.

While UMUX provides a rapid and efficient usability measure, it remains subjective. As with other self-report scales, scores may be influenced by the user's expectations, prior experience, or the task context. Therefore, UMUX is often combined with qualitative methods, such as observation or interviews, and objective performance metrics to obtain a more comprehensive understanding of system usability [189]. Its combination of brevity,

reliability, and interpretability has made UMUX a valuable tool in academic research and practical usability testing, particularly when repeated or large-scale evaluations are needed.

4.4 Virtual Reality Evaluation

The virtual reality environment is evaluated with the iGroup Presence Questionnaire (IPQ), Virtual Embodiment Questionnaire (VEQ), and Virtual Reality System Usability Questionnaire (VRSUQ).

4.4.1 iGroup Presence Questionnaire (IPQ)

The iGroup Presence Questionnaire (IPQ) [190] is a widely used self-report instrument designed to measure the subjective sense of presence experienced by users in virtual environments. Presence is defined as the psychological sensation of "being there" in a mediated or virtual environment, and it is considered a central factor in the effectiveness and realism of immersive systems such as virtual reality, augmented reality, and 3D simulations. The IPQ provides a standardized framework to quantify this subjective experience, enabling researchers to assess and compare the sense of presence across different systems or experimental conditions.

The IPQ consists of several items grouped into subscales that capture key dimensions of presence: Spatial Presence, which reflects the feeling of being physically located within the virtual environment; Involvement, which measures attention, engagement, and absorption in the virtual experience; Experienced Realism, which assesses the perceived realism of the environment and events; and a General Presence item representing an overall global assessment of the sense of "being there." Participants rate each item on a Likert-type scale, allowing quantitative analysis of subjective presence experiences.

The IPQ has been validated in numerous studies and correlates with behavioral and physiological indicators of immersion, such as navigation performance, gaze patterns, and heart rate variability [191]. Its concise format, strong psychometric properties, and multidimensional structure make it particularly suitable for research in virtual reality, training simulations, and telepresence applications. Although it is a subjective measure, the IPQ provides critical insights into user experience that cannot be fully captured by objective performance metrics, offering guidance for system design and the evaluation of immersive environments.

4.4.2 Virtual Embodiment Questionnaire (VEQ)

The Virtual Embodiment Questionnaire (VEQ) [192] is a self-report instrument designed to assess the subjective experience of embodiment within virtual environments, particularly in contexts involving avatars or full-body representations. Embodiment refers to the sense of owning and controlling a virtual body, and it is a critical factor in immersive virtual reality experiences, influencing presence, engagement, and behavioral responses. The VEQ provides a standardized method for quantifying these experiences, enabling researchers to investigate how different design choices, such as avatar appearance, motion tracking, or perspective, affect the user's sense of embodiment.

The questionnaire evaluates multiple dimensions of virtual embodiment. Key subscales include Body Ownership, which measures the feeling that the virtual body is one's own; Agency, which assesses the perception of controlling the virtual body's actions; and Self-Location, which captures the experience of being spatially located within the virtual body or environment. Participants rate their experiences on Likert-type scales, allowing quantitative comparisons across conditions, avatars, or VR systems.

The VEQ has been applied in studies exploring the illusion of body ownership, full-body VR interactions, and avatar-mediated social experiences. Its psychometric validity has been supported by correlations with behavioral and physiological measures, such as movement synchrony, reaction times, and skin conductance responses during virtual interactions [193, 194]. By capturing subjective aspects of embodiment that are not observable through objective performance metrics alone, the VEQ provides essential insight into user experience, offering guidance for designing more immersive and persuasive virtual environments.

4.4.3 Virtual Reality System Usability Questionnaire (VRSUQ)

The Virtual Reality System Usability Questionnaire (VRSUQ) [195] is a self-report instrument specifically designed to evaluate the usability of virtual reality systems from the user's perspective. While general usability scales, such as the System Usability Scale (SUS), provide broad insights, the VRSUQ addresses VR-specific interaction challenges, including navigation, control devices, and feedback on immersive environments. It enables researchers to assess how intuitive, efficient, and satisfactory a VR system is, providing valuable guidance for system design and improvement.

The questionnaire consists of several items rated on a Likert-type scale, covering key dimensions of VR usability. These include system responsiveness, ease of interaction, clarity of feedback, and overall user satisfaction. Participants' responses are combined to generate an

overall usability score, which can be used to compare different VR systems or configurations. The VRSUQ has been applied in diverse contexts, from immersive training simulations and rehabilitation platforms to entertainment and educational VR applications.

The VRSUQ focuses on VR-specific usability aspects, which are often overlooked in generic usability instruments [196]. It captures challenges unique to immersive environments, such as motion control, spatial navigation, and perceptual feedback, complementing objective performance measures and broader subjective usability questionnaires. While self-reported measures may be influenced by individual experience or prior exposure to VR, the VRSUQ provides essential insights into the user-centered evaluation of immersive systems, supporting iterative design and optimization of user experience.

4.5 Conclusions

The questionnaires presented in this chapter serve as essential complements to the objective kinematic and physiological measurements introduced earlier. While sensors capture how the body moves and responds physiologically, questionnaires provide insight into how participants perceive, experience, and cognitively interpret the tasks and environments encountered throughout the experiments. This dual perspective, both objective and subjective, is crucial for developing a comprehensive understanding of human behavior in fall-related scenarios.

Each group of questionnaires contributes a distinct dimension of human-centered evaluation. Perception-related instruments reveal subjective workload, emotional states, and motivational factors that may shape user engagement and performance. Ergonomics-focused questionnaires identify discomfort, anxiety, visual strain, and potential barriers to prolonged or realistic system use. Usability measures offer a high-level assessment of how intuitive and efficient the technological setup is, while VR-specific evaluations quantify critical aspects such as presence, embodiment, and immersion, which directly influence the ecological validity of the fall-risk simulations.

The philosophical rationale for including these instruments aligns with the broader framework of this thesis: falls are not merely biomechanical events but are deeply influenced by cognitive, emotional, and perceptual factors. Understanding user experience ensures that developed systems are not only technically accurate but also comfortable, acceptable, and safe for real-world applications. Moreover, subjective responses help interpret physiological and behavioral data, contextualizing whether observed patterns reflect genuine task demands, stress responses, or usability constraints.

By integrating questionnaire data with multimodal sensor information, this thesis provides a comprehensive understanding of each participant's physical, physiological, and psychological state. This integration supports the transition to the next part of the work, where both objective signals and subjective experiences are combined within the proposed experimental framework. The following chapters apply these tools to evaluate fall detection algorithms, impact prediction models, and generalized/personalized fall-risk estimation methods, ensuring that system performance is interpreted within a comprehensive human-centered context.

Chapter 5

Fundamentals of Machine Learning

Machine Learning (ML) provides the computational foundations for the entire framework developed in this thesis. From reactive fall detection to proactive impact prediction and generalized/personalized fall-risk estimation, each component relies on ML models capable of interpreting complex, noisy, and multimodal data. Inertial and physiological signals, introduced in Chapters 3 and 4, are high-dimensional, heterogeneous, and often exhibit strong temporal dependencies, making traditional rule-based or threshold-based approaches insufficient for robust generalization in real-world conditions.

This chapter introduces the fundamental ML concepts required to understand the methods used throughout the thesis. It begins with a comprehensive overview of model evaluation metrics for both regression and classification tasks, emphasizing the importance of selecting appropriate metrics when dealing with imbalanced datasets, rare events (such as falls), and continuous prediction horizons (such as time-to-impact). Next, it reviews a broad range of machine learning methods, from traditional statistical models to deep learning architectures. Each method is presented together with its advantages and limitations, providing rationale for why different approaches are later compared, combined, or replaced based on empirical evidence obtained in Chapters 7-9.

Finally, the chapter discusses strategies for imbalanced learning, a critical aspect of fall-related datasets where non-fall activities vastly outnumber fall events and where high-risk physiological states may be underrepresented. These foundations prepare the reader for the methodological choices made in the development, benchmarking, and interpretation of the ML models deployed in subsequent chapters.

5.1 Overview

In Artificial Intelligence (AI), Machine Learning (ML) represents a core set of techniques that enable systems to learn autonomously from data without requiring explicit programming. ML addresses induction problems in which an underlying model is discovered for predictive or descriptive purposes, based on patterns learned from a training set [197, 198]. Over the past decade, ML has emerged as a cornerstone of data-driven healthcare, robotics, and wearable sensing research, offering tools capable of uncovering complex and often non-linear relationships within large datasets [199, 200].

Machine Learning approaches are typically divided into categories based on the nature of the learning process:

- Supervised learning uses labeled data, where both input features and target outputs are provided, allowing the algorithm to learn a mapping between them [201]. Tasks can be formulated as classification problems, when the output is categorical (such as "fall" vs. "no fall"), or as regression problems, when the output is continuous (such as predicting gait speed).

- Unsupervised learning relies on unlabeled data to identify hidden structures and patterns within the dataset. Typical approaches include clustering, dimensionality reduction, and anomaly detection, which can reveal meaningful groupings and reduce complexity for downstream tasks [198, 202].

- Semi-supervised learning combines small amounts of labeled data with larger amounts of unlabeled data, falling between supervised and unsupervised learning. This approach is advantageous in domains such as healthcare, where collecting annotated data can be costly and time-consuming [203].

- Reinforcement learning, increasingly applied in robotics and human–machine interaction, is another paradigm in which agents learn by interacting with their environment and optimizing long-term rewards [204].

A particularly transformative branch of ML is Deep Learning (DL), which leverages artificial neural networks (ANNs) with multiple layers to automatically learn hierarchical feature representations directly from raw data [199]. Inspired by the structure of biological neurons, artificial neurons compute weighted sums of inputs and apply non-linear activation functions. By stacking many such layers, deep networks can capture increasingly abstract features, often surpassing traditional ML in speech recognition, computer vision, and biomedical

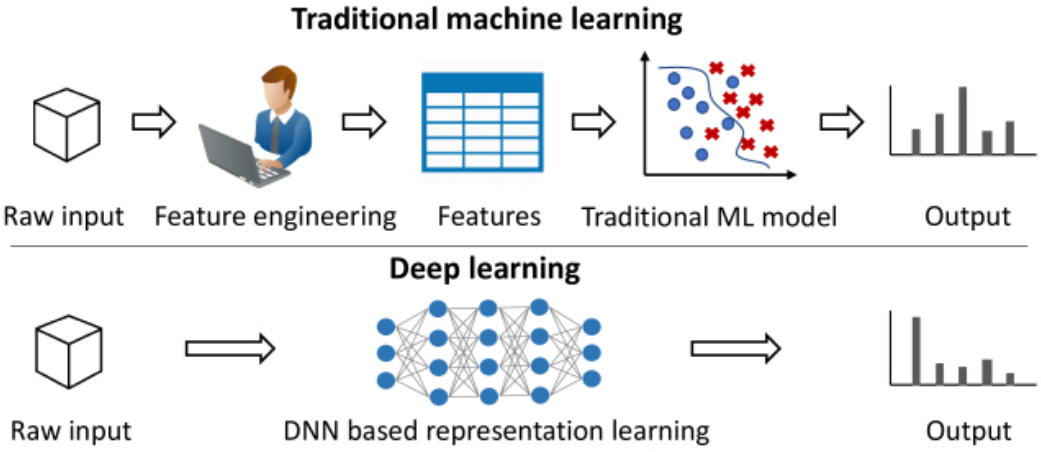


Figure 5.1 Difference between traditional "machine learning" (above) and "deep learning" (below).

signal analysis tasks. Unlike conventional ML algorithms, which require manual feature engineering, DL models perform feature extraction automatically, end-to-endly.

As illustrated in Figure 5.1, a typical ML pipeline begins with data collection and preprocessing, followed by feature extraction, which transforms raw data into informative attributes. An ideal feature set maximizes inter-class variability while minimizing intra-class variability [205]. In DL-based methods, this stage is often replaced or supplemented by automated feature learning. Depending on the task, the extracted features are fed into classification or regression algorithms.

Evaluating ML models requires careful consideration of generalization performance and potential pitfalls such as overfitting, where a model performs well on the training data but fails to generalize to unseen data. Validation strategies, such as k-fold cross-validation (CV), are widely employed to address this issue. In k-fold CV, the dataset is partitioned into k subsets; the model is trained on k–1 subsets and tested on the remaining one, iterating until each subset has been used for testing. Results are averaged across folds to provide a robust estimate of performance [206, 207].

In contexts with multiple samples per subject, Leave-One-Subject-Out (LOSO) cross-validation is often applied to mitigate subject bias, ensuring that models generalize to new individuals rather than memorizing subject-specific patterns. Furthermore, nested cross-validation has been proposed as a rigorous method for hyperparameter tuning and performance estimation, as it reduces bias in performance evaluation and allows statistical testing of differences between models [208, 209]. Although nested CV is computationally

expensive and less critical for large homogeneous datasets, it is particularly valuable when data are limited, noisy, or contain outliers.

Overall, ML provides a robust framework for predictive modeling and knowledge discovery. Its integration with wearable sensing, signal processing, and healthcare has enabled the development of increasingly accurate and adaptive systems for applications such as fall detection and prevention, where robust generalization across diverse populations and contexts remains a key challenge.

5.2 Performance Evaluation

After the validation process, the performance of a regressor/classifier can be assessed using a variety of metrics, each providing a different perspective on predictive ability.

5.2.1 Regression

In regression problems, the goal is to predict a continuous outcome rather than a discrete class label. Evaluating the performance of regression models requires metrics that capture the accuracy of predictions and the magnitude and distribution of prediction errors. Unlike classification tasks, where confusion matrices and derived measures are central, regression relies on error-based and correlation-based indices.

One of the most widely used metrics is the Mean Absolute Error (MAE), which calculates the average magnitude of prediction errors, disregarding their direction. Eq. 5.1 reports the mathematical equation, where x_i is the true value, y_i is the prediction, $e_i = x_i - y_i$ is the prediction error, and n is the number of samples. MAE provides an intuitive measure of prediction accuracy, expressed in the same units as the target variable [210].

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i| = \frac{1}{n}\sum_{i=1}^{n}|e_i|. \tag{5.1}$$

The Mean Squared Error (MSE) is closely related, which penalizes larger errors more heavily due to the squaring operation (Eq. 5.2). The square root of MSE Eq. 5.3, known as the Root Mean Squared Error (RMSE), is often preferred for interpretability, as it shares the same scale as the dependent variable [211].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2, \tag{5.2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2}. \tag{5.3}$$

Another important family of metrics focuses on relative performance. The Mean Absolute Percentage Error (MAPE) expresses errors as percentages of the true values, providing scale-independent interpretability Eq. 5.4.

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{x_i - y_i}{x_i}\right|. \tag{5.4}$$

However, it can be unstable when true values are close to zero. To overcome such limitations, alternative measures such as the Symmetric Mean Absolute Percentage Error (sMAPE) and the Mean Absolute Scaled Error (MASE) have been proposed, particularly in time-series forecasting, where comparing across datasets of different scales is critical [212].

$$sMAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{2|x_i - y_i|}{|x_i| + |y_i|}. \tag{5.5}$$

Correlation-based metrics also play an essential role. The Coefficient of Determination (R^2) is commonly used to indicate the proportion of variance in the dependent variable explained by the model. The general definition of the coefficient of determination is given in Eq. 5.6, where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the mean of the true data.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}. \tag{5.6}$$

While widely reported, R^2 has limitations, especially in non-linear contexts or when comparing models with different complexities [213]. Adjusted R^2 Eq. 5.7 corrects the number of predictors, reducing the risk of artificially inflated values in high-dimensional settings, considering the total number of variables in the model p.

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}. \tag{5.7}$$

In addition, recent literature emphasizes the importance of considering the distribution of residuals. Metrics such as the Normalized RMSE (NRMSE) or Relative Absolute Error (RAE) allow comparisons across different datasets or target scales. Beyond point estimates, prediction interval coverage probability (PICP) and prediction interval normalized average width (PINAW) are increasingly used to evaluate models that provide uncertainty estimates, especially in probabilistic regression and Bayesian approaches [214].

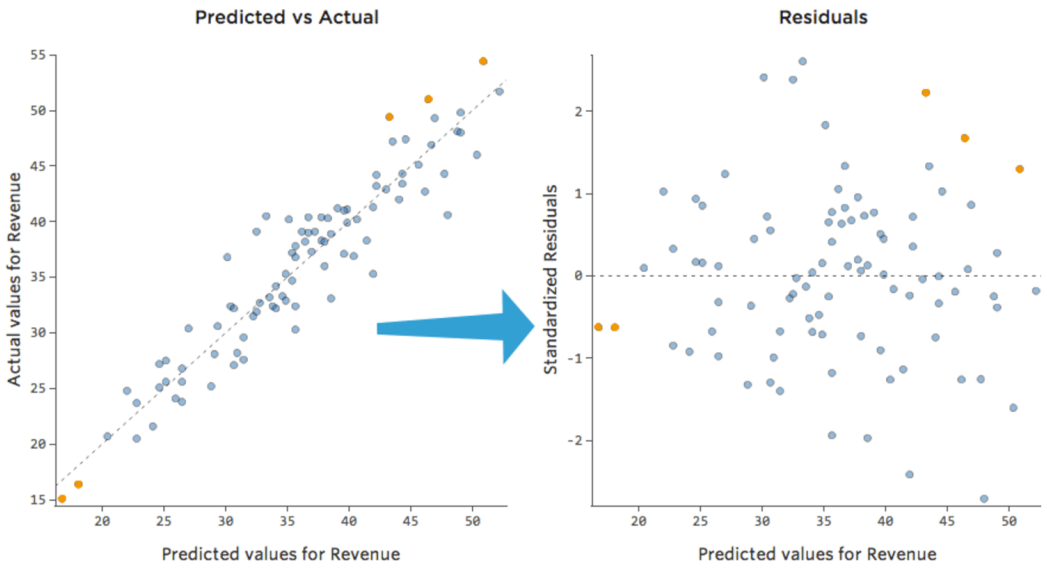


Figure 5.2 Example of Predicted vs. Actual plot (on the left) and Residuals plot (on the right) [Qualtrics].

Visualization tools complement numerical metrics. Residual plots help identify heteroscedasticity, autocorrelation, or systematic biases in predictions. In probabilistic settings, calibration curves are employed to assess whether predicted confidence intervals adequately capture the true variability of the data.

In summary, regression performance evaluation relies on a combination of error magnitude metrics (MAE, RMSE, MAPE), goodness-of-fit measures (R^2, adjusted R^2), and, increasingly, uncertainty-aware metrics for probabilistic modeling. The choice of metrics depends on the application context, the scale of the target variable, and the importance of balancing interpretability with robustness.

5.2.2 Classification

One of the most common measures is accuracy (Eq. 5.8), which quantifies the proportion of correctly classified instances relative to the total number of samples [215].

$$Acc = \frac{TP+TN}{TP+TN+FP+FN}. \tag{5.8}$$

Where TP is true positive, TN is true negative, FP is false positive, and FN is false negative. However, accuracy alone can be misleading, especially in class imbalance, where one class significantly outnumbers the other. For this reason, complementary metrics such as sensitivity (also known as recall), specificity, and precision are frequently reported. Sensitivity

(Recall, True Positive Rate) measures the ability of a model to correctly identify positive cases, thus reflecting its effectiveness in detecting the condition of interest (Eq. 5.9).

$$Se = \frac{TP}{TP+FN}. \tag{5.9}$$

Specificity (True Negative Rate) captures the ability to correctly identify negative cases, ensuring that false alarms are minimized (Eq. 5.10).

$$Sp = \frac{TN}{TN+FP}. \tag{5.10}$$

Precision (Positive Predictive Value) indicates the proportion of optimistic predictions that are indeed correct (Eq. 5.11).

$$Prec = \frac{TP}{TP+FP}. \tag{5.11}$$

Additional composite measures are often used for binary classification. The F1-score, the harmonic mean of precision and recall, provides a balanced assessment of classifier performance when both metrics are important (Eq. 5.12) [216].

$$F_1 = 2\frac{Se \times Prec}{Se+Prec} = \frac{2\,TP}{2\,TP+FP+FN}. \tag{5.12}$$

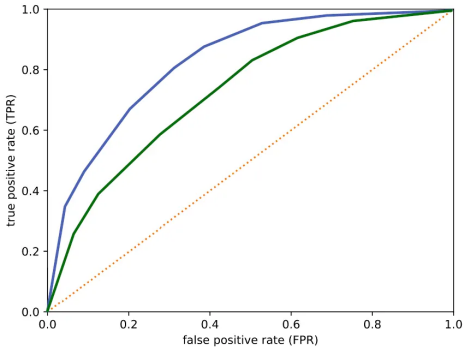
Another metric, Youden's J index, combines sensitivity and specificity into a single statistic that summarizes diagnostic effectiveness (Eq. 5.13) [217].

$$F_1 = Se+Sp-1 = \frac{TP \times TN - FP \times FN}{(TP+FN)(TN+FP)}. \tag{5.13}$$

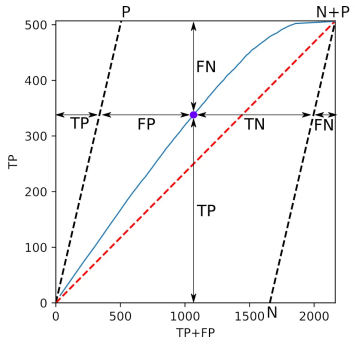
Similarly, the geometric mean (G-mean) evaluates the balance between sensitivity and specificity, offering insight into how well the classifier performs across classes (Eq. 5.14) [218].

$$G = \sqrt{Se \times Sp} = \sqrt{\frac{TP \times TN}{(TP+FN)(TN+FP)}}. \tag{5.14}$$

Beyond scalar metrics, graphical tools are crucial in evaluating classifiers across all possible thresholds. The Receiver Operating Characteristic (ROC) curve is one of the most widely adopted methods (see Figure 5.3a), plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) across threshold values. The Area Under the ROC Curve (AUC) is commonly interpreted as the probability that a randomly chosen positive instance is ranked higher than a randomly chosen negative instance, providing a threshold-independent measure



(a) The ROC curves of two classifiers (blue and green lines) and the curve representing uninformed (random) classifiers (orange dots).



(b) The TOC curve of a classifier (blue line) and the curve representing random classifiers (red dots). All information about the confusion matrix and the ROC graph is available.

Figure 5.3 Receiver Operating Characteristic curve (ROC) (Left) and Total Operating Characteristic Curve (TOC) (Right). Author: Michael Köpf.

of model performance [219]. Classifiers performing closer to the top-left corner of the ROC space demonstrate better discriminative ability.

Despite its popularity, ROC analysis has some limitations, particularly its inability to visualize the distribution of test samples and the whole structure of the confusion matrix. To address this, [220] introduced the Total Operating Characteristic (TOC) curve. Unlike the ROC, which relies on the TPR-FPR coordinate system, the TOC plots true positives (TP) against the sum of true positives and false positives (TP + FP), thereby incorporating information on all four confusion matrix entries (see Figure 5.3b). This provides a more comprehensive representation of classification performance, including dataset composition and the balance between positive predictions and false alarms. As a result, TOC has gained attention as a complementary tool for model evaluation, especially in applications such as spatial modeling, remote sensing, and ecological forecasting, where class distributions are often imbalanced [221].

These metrics and visualization techniques provide a robust framework for evaluating the performance of classifiers. The choice of evaluation method depends mainly on the application context, the cost of misclassifications, and the balance between sensitivity and specificity required for the task.

5.3 Learning Methods

Machine learning methods provide the computational foundation for all the parts of the framework developed later in this thesis. Each method offers specific advantages but also carries limitations that influence its suitability for modeling complex, time-dependent, and highly variable human movement and physiological patterns. This section provides a critical evaluation of the main learning approaches used in the literature and applied throughout this work, highlighting not only their strengths but also the reasons why multiple methodologies were explored in subsequent chapters.

5.3.1 Linear Regression

Linear regression is one of the oldest and most interpretable predictive modeling techniques. It assumes a linear relationship between the input features and the target variable, estimating coefficients that minimize the residual sum of squares between observed and predicted values [222]. Although simple, it remains one of the most straightforward and most interpretable modeling techniques. Its strengths lie in its low computational cost, transparency, and suitability for systems where linear relationships dominate. These characteristics provide a strong baseline for STF-related regression tasks, such as time-to-impact estimation (8.7).

Regularized variants such as Lasso (L1) and Ridge (L2) regression improve generalization by penalizing large coefficients [223]. However, its core limitation is the assumption of linearity. STFs are inherently non-linear dynamic events shaped by abrupt motion changes, multi-segment biomechanics, and noisy wearable data. Linear regression consistently underestimates the rapid curvature of the fall trajectory, failing to capture the fast dynamics leading to ground impact. This limitation motivates the transition toward more flexible non-linear models.

5.3.2 Decision and Regression Tree

Decision trees partition the feature space into hierarchical regions by recursively splitting the data according to feature thresholds that minimize impurity measures such as the Gini index or entropy [224]. Tree-based methods excel in interpretability and can model non-linear decision boundaries without requiring feature normalization. They are particularly useful when working with low-dimensional datasets or when interpretability is essential for deployment in safety-critical systems.

Nonetheless, single trees are prone to overfitting and can be unstable to slight variations in data. Ensemble variants, such as Random Forests [225] and Gradient Boosting Machines [226], overcome this limitation by aggregating multiple weak learners. They also struggle with high-frequency temporal signals, such as inertial time series, since they do not inherently model temporal dependencies. Our experiments confirm this: in the Fall-KAN evaluation (8.7), regression trees show high variance and fail to track smooth temporal trends in angle evolution.

5.3.3 Discriminant Analysis

Discriminant analysis methods, including Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA), offer computational simplicity and assume that data from each class follow a Gaussian distribution and aim to find a linear (or quadratic) combination of features that best separates the classes [227]. These techniques are particularly useful when the covariance structure of each class can be estimated reliably. For fall detection, this yields strong baseline performance when classes are well separated, especially under controlled conditions.

However, real-world inertial and physiological data rarely satisfy Gaussian assumptions. The presence of noise, high inter-subject variability, and overlapping motion patterns (such as fast ADLs vs. soft falls) significantly degrades performance. This limitation becomes evident in Chapter 9, where discriminant models fail to generalize across subjects under LOSO validation.

5.3.4 Logistic Regression

Logistic regression is a probabilistic classifier that models the likelihood of a binary outcome using the logistic (sigmoid) function [228]. Despite its lightweight and simplicity, it offers robust and interpretable decision boundaries, remaining a benchmark for classification tasks. Extensions such as multinomial logistic regression and regularized logistic models enable the handling of multiclass problems and mitigate overfitting. Its limitation, however, is similar to Linear Regression: it cannot learn non-linear temporal relations unless manually engineered features are introduced.

5.3.5 Naïve Bayes

Naïve Bayes models rely on Bayes' theorem under the simplifying assumption of feature independence [229]. They are computationally efficient and perform surprisingly well in practice, especially for small datasets and high-dimensional data, such as text or sensor readings, where conditional independence is approximately held.

However, its performance strongly depends on the independence assumption, which is unrealistic for correlated multimodal signals (such as IMU axes, ECG–EDA interactions). Consequently, Naïve Bayes tends to oversimplify fall-related data structures. In fall-risk classification experiments (Chapter 9), Naïve Bayes systematically underperformed compared to more expressive models.

5.3.6 Support Vector Machine (SVM)

SVMs are powerful discriminative models that seek the optimal hyperplane maximizing the margin between classes [230]. Kernel functions (such as radial basis and polynomial) enable mapping input data into higher-dimensional spaces, allowing for non-linear classification. SVMs are effective for small- to medium-sized datasets with complex decision boundaries and are commonly used in biomedical signal analysis and pattern recognition. They have historically performed well in fall detection tasks using handcrafted features.

However, SVMs scale poorly with large datasets, offer no native temporal modeling, and require substantial feature engineering to perform competitively on inertial time series. As seen in Chapter 8, SVMs lag behind deep learning models in predicting continuous fall trajectories or rapidly changing body angles.

5.3.7 k-Nearest Neighbor

The k-Nearest Neighbor (k-NN) algorithm is a non-parametric method that classifies a sample based on the majority label among its k closest training instances, typically measured using Euclidean distance.

Although simple, its performance depends heavily on distance metrics and feature scaling, and it becomes computationally expensive for large datasets [231]. Besides, predictions depend purely on stored samples, which generates sensitivity to noisy and high-dimensional data and no generalization. In the experiments on fall-risk estimation (Chapter 9), kNN suffered from substantial performance drops under LOSO evaluation due to significant inter-subject variability.

5.3.8 Kernel-based method

Kernel-based methods transform data into high-dimensional feature spaces where linear separability can be achieved. However, exact kernel computations are costly for large datasets. Kernel approximation techniques, such as Random Fourier Features [232], provide efficient approximations that maintain predictive performance while reducing computational complexity, facilitating their use in large-scale machine learning pipelines.

Their limitations include high computational cost, sensitivity to kernel and hyperparameter selection, and a lack of scalability for large streaming datasets. For fall impact prediction (Chapter 8), kernel methods captured some non-linearities but failed to learn short-timescale dynamics as effectively as KANs or recurrent networks.

5.3.9 Ensemble model

Ensemble models combine multiple weak learners to improve predictive accuracy and robustness. Bagging (Bootstrap Aggregation), boosting, and stacking are among the most widely used ensemble strategies. Random Forests and AdaBoost exemplify bagging and boosting, respectively [225]. More recent ensemble frameworks, such as XGBoost [233] and LightGBM [234], have achieved state-of-the-art results across various structured data applications. In general, they handle noisy data well and require minimal preprocessing.

However, they still struggle with temporal dependencies unless time-series features are manually engineered. While ensembles performed respectably in the Fall-KAN comparison (Chapter 8), they were unable to match the temporal accuracy required for millisecond-scale impact estimation. This limitation directly influenced the shift toward specialized new learning models.

5.3.10 Artificial Neural Network (ANN)

Artificial Neural Network (ANN) or Multilayer Perceptron (MLP) is a computational model inspired by the human brain, consisting of interconnected layers of artificial neurons. Each neuron applies a weighted sum of inputs followed by a non-linear activation function. ANNs can approximate complex non-linear mappings and have been widely adopted in diverse domains, from biomedical signal processing to natural language understanding [199]. They outperformed traditional methods in several reported experiments.

However, feedforward networks cannot inherently model temporal dependencies without the use of windowing techniques, and they require large amounts of data to general-

ize effectively. In time-of-impact estimation (Chapter 8), the ANN was surpassed by the Kolmogorov-Arnold Network.

5.3.11 Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN) is a specialized neural architecture designed for spatial data, particularly effective for image and sensor-based pattern recognition. By employing convolutional filters and pooling operations, CNN captures hierarchical local features and spatial dependencies [235]. In wearable sensing applications, CNN can extract movement-related features directly from raw accelerometer or IMU signals without explicit feature engineering.

Its limitation lies in its locality. The CNN excels at short-term dependencies but is less effective for long-term temporal modeling unless combined with recurrent layers. It motivated the LSTM/GRU architectures explored in Chapters 7 and 8.

5.3.12 Recurrent Neural Network (RNN)

A Recurrent Neural Network (RNN) is designed to model sequential data by maintaining an internal state that captures temporal dependencies. They are particularly suited for time-series data, such as ECG, EMG, or gait signals. Variants such as Long Short-Term Memory (LSTM) [236] and Gated Recurrent Units (GRU) [237] effectively mitigate vanishing gradient issues and are widely used for activity recognition and fall prediction tasks. They learn temporal relationships critical for differentiating falls from regular movements and for forecasting motion trajectories.

However, their drawbacks include high computational cost, susceptibility to vanishing/exploding gradients, the need for large datasets, and the potential to overfit dynamic activities (as observed in Chapter 7, where LSTM-based FDNN showed reduced specificity for fast activities such as Jogging Quickly). These limitations justified exploring alternative models such as the Kolmogorov–Arnold Network (KAN) in Chapter 8, which demonstrated superior temporal accuracy for time-to-impact estimation.

5.3.13 Summary

Taken together, the strengths and limitations of these learning methods explain why multiple algorithmic strategies were explored throughout this thesis. The STF problem spans detection, forecasting, and risk estimation, each requiring models with unique capabilities. It motivated

the progression from traditional ML to deep learning and, ultimately, to novel architectures such as KANs and biomechanically informed recurrent networks.

5.4 Imbalance Learning

In many real-world applications, datasets are characterized by a strong class imbalance, where one class (the majority class) vastly outnumbers the other (the minority class). This imbalance is widespread in domains such as medical diagnosis, fraud detection, or fault prediction, where the minority class represents the critical cases of interest (such as the presence of disease, fraudulent transaction, or equipment failure). Standard machine learning algorithms tend to be biased toward the majority class, often achieving deceptively high accuracy while failing to correctly identify instances of the minority class, thereby compromising sensitivity and overall utility [238, 239].

Several strategies have been developed to address class imbalance, which can be broadly categorized into data-level methods and algorithm-level methods. Data-level approaches manipulate the training data distribution to balance class representation:

- Random oversampling duplicates instances of the minority class, while random undersampling removes instances from the majority class. While simple, these methods may lead to overfitting (in oversampling) or information loss (in undersampling).
- Class weighting assigns greater weights to samples from the minority class and lesser weights to the majority class during training. This way, the model pays more attention to the minority class and learns to make more accurate predictions.
- More advanced oversampling techniques, such as the Synthetic Minority Oversampling Technique (SMOTE) [240], generate synthetic minority class samples by interpolating between existing samples and their nearest neighbors. This reduces overfitting compared to naïve oversampling.
- Adaptive Synthetic Sampling (ADASYN) [241] extends SMOTE by adaptively focusing sample generation on minority class regions that are harder to learn, thereby improving classifier performance in challenging decision boundaries.

Algorithm-level approaches adjust learning algorithms to account for imbalance directly:

- Cost-sensitive learning assigns higher misclassification penalties to the minority class, weighting minority class errors more heavily during training [242]. This enables the

model to prioritize the accurate detection of rare events without altering the dataset's distribution.

- Ensemble methods have also been adapted for use in imbalanced learning. For example, RUSBoost [243] integrates random undersampling with AdaBoost, maintaining a balanced training set at each boosting iteration while preserving ensemble diversity and reweighting advantages.

The choice of strategy depends on the application domain, dataset characteristics, and the cost of false positives versus the cost of false negatives. In medical applications, for instance, high sensitivity (recall of the minority class) is often prioritized over overall accuracy, as failing to detect a critical condition may have severe consequences [244]. Increasingly, hybrid methods that combine sampling and cost-sensitive approaches are being developed to enhance classifier robustness in scenarios of severe imbalance.

In summary, imbalanced learning remains a crucial area of machine learning research, as imbalanced datasets are pervasive in practice. Techniques such as SMOTE, ADASYN, and RUSBoost have significantly advanced the field, but ongoing research continues to explore adaptive, ensemble, and deep learning-based solutions tailored to imbalanced domains.

5.5 Conclusions

This chapter presented the machine learning foundations required for the predictive framework developed in this thesis. It introduced the key performance metrics used for both regression and classification, highlighting their importance when evaluating models trained on heterogeneous, temporal, and often imbalanced datasets related to falls. The chapter then critically reviewed a broad spectrum of learning methods, from linear models and traditional machine learning classifiers to deep neural networks, emphasizing not only their strengths but also the practical limitations that influence their suitability for different stages of fall detection, forecasting, and risk estimation.

A central insight emerging from this review is that no single algorithmic family is universally optimal for the fall problem. Linear and kernel methods offer interpretability and low computational cost, but struggle with the non-linear, rapidly changing dynamics inherent to falls. Ensemble models improve robustness but require extensive feature engineering. Deep learning approaches, particularly convolutional and recurrent architectures, offer superior performance for temporal inertial data but demand large datasets and careful regularization to avoid overfitting. These observations provide the methodological justification for exploring

multiple modeling strategies in Chapters 7-9 and for introducing novel architectures, such as the Kolmogorov–Arnold Network (KAN), where classical approaches prove insufficient.

Finally, imbalanced learning techniques were introduced as a necessary element of fall detection and fall-risk classification, given the rarity of true fall events and high-risk physiological states in real-world scenarios. These principles directly inform the design of the experimental pipeline implemented in later chapters.

With these ML foundations established, Chapter 6 introduces the complete experimental framework, showing how the sensing modalities, questionnaires, and ML techniques are integrated into a coherent system. It provides a structural overview of the three main components: reactive fall detection, proactive fall impact prediction, and predictive fall-risk estimation, setting the stage for the detailed methodological and experimental results presented in the subsequent chapters.

Part II

Framework

Chapter 6

Framework overview

Hereinafter, the term “fall” is sometimes used as an alternative to “STF” or to maintain consistency with the terminology used in the literature. The framework developed in this thesis addresses the pervasive challenge of fall safety by establishing a continuum of machine learning methodologies that progress from reactive event classification to proactive mitigation and predictive risk estimation. This holistic approach, depicted in Figure 6.1, is structurally organized into three main components, each of which is covered in a dedicated subsequent chapter. Because each component addresses a different stage of the fall process, a precise conceptual vocabulary is necessary to distinguish between the reactive, proactive, and predictive elements of fall-prevention technology.

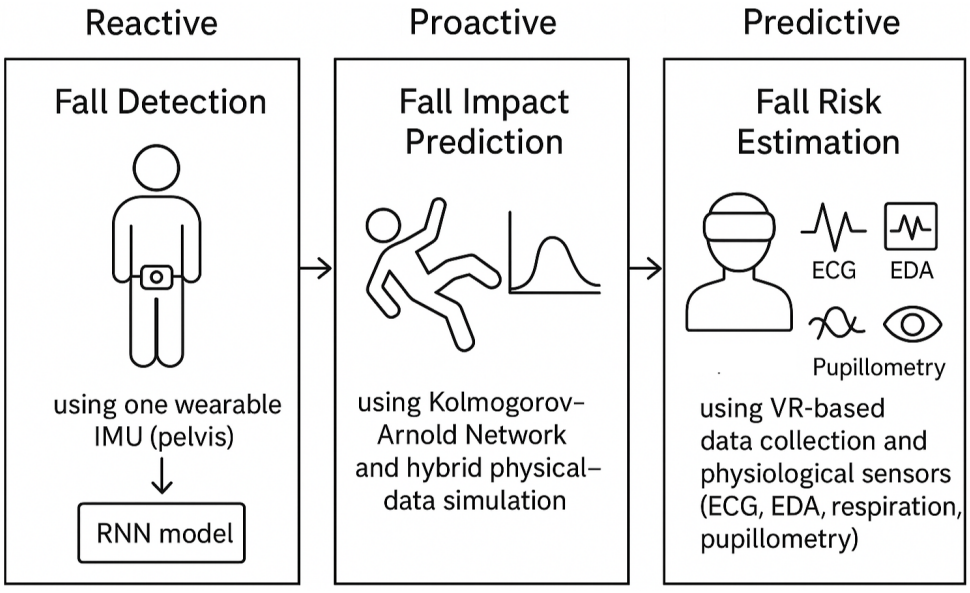


Figure 6.1 Framework overview.

The reactive component refers to methods that operate reactively only after a fall has already begun and aim to recognize the event based on kinematic or physiological evidence. In contrast, the Proactive component identifies methods that respond during the early phase of an ongoing fall, aiming to proactively anticipate its progression, particularly the moment of ground impact, so that mitigation systems such as wearable airbags can be triggered in time. Finally, the predictive component denotes methods that operate before any fall occurs, by estimating an individual's likelihood of falling based on physiological, behavioral, or contextual indicators predictively. For each of these, below the set goal, the methodology used and the key challenges are outlined.

6.1 Reactive: Fall Detection

This component (Chapter 7) focuses on the traditional and fundamental task of accurately identifying an ongoing or recently started fall event using inertial data.

Goal To detect a fall while it happens to enable a timely alert or response.

Methodology Utilizes data from a wearable Inertial Measurement Unit (IMU) positioned at the pelvis. A Deep Neural Network (DNN) that incorporates Recurrent Neural Network (RNN) layers, specifically Long Short-Term Memory (LSTM), is employed for binary classification, distinguishing between Falling and Non-Falling events.

Key Challenge Achieving a high True Negative Rate (TNR) to minimize false alarms during dynamic Activities of Daily Living (ADLs) and a high True Positive Rate (TPR) to maximize protection during falls.

6.2 Proactive: Fall Impact Prediction

In this chapter (Chapter 8), fall detection is not treated as an isolated task, but as a foundational component of a broader framework that also anticipates the evolution of a fall. By coupling real-time fall detection with impact-time forecasting, the proposed approach advances beyond merely identifying a fall to estimating when the impact will occur, an essential requirement for activating wearable mitigation systems.

Goal To predict the trajectory/Time-to-Impact during an ongoing fall event to facilitate life-saving pre-impact mitigation.

Methodology This dual approach employs two novel models. Fall-KAN: Utilizes a Kolmogorov-Arnold Network for highly accurate and interpretable Time-to-Impact estimation. Fall Forecasting Neural Network (FFNN): Employs an RNN model (based on Gated Recurrent Units or GRUs) trained using a hybrid approach involving biomechanical simulations (Inverted Pendulum model) to forecast the subject's angle over time.

Key Challenge Providing accurate predictions within the narrow millisecond window required for actuator deployment.

6.3 Predictive: Fall Risk Estimation

This final component (Chapter 9) moves beyond the event of the fall itself to assess whether physiological and cognitive states increase the overall likelihood of a fall and whether this can be estimated in generalized terms.

Goal To continuously assess and classify the subject's Fall Risk Level (Low, Medium, or High) for prevention purposes.

Methodology Leverages multimodal physiological sensors, including Electrocardiography (ECG), Electrodermal Activity (EDA), respiration, and pupillometry, integrated with kinesthetic data. Data is mainly collected in ecologically valid, high-stress Virtual Reality (VR)-based fall scenarios (XoFallVR dataset). The core finding is the need for a machine learning approach to overcome high inter-subject variability in physiological responses.

Key Challenge Developing models that reliably generalize across different human subjects and contexts, ultimately solved by adopting personalized calibration.

Chapter 7

Fall Detection

Research on fall detection has evolved significantly over the past three decades, progressing from early threshold-based systems that relied on handcrafted features to today's intelligent, multimodal, and adaptive architectures that integrate advanced machine learning (ML) and deep learning (DL) techniques. However, despite this progress, many challenges remain regarding generalizability, real-time adaptability, energy efficiency, and privacy preservation.

7.1 Background

The first wave of research in the 1990s focused on hardware-based systems for sensing impact and motion. Early solutions employed accelerometers and tilt or shock sensors embedded in wearable badges to detect abrupt changes in body orientation. In [30], a miniaturized accelerometer and a microcomputer chip were embedded in a badge to detect falls. At the same time, in [245], a piezoelectric shock sensor and a mercury tilt switch are applied to monitor body orientation and detect falls. These systems were inexpensive and computationally simple but inherently limited by their fixed thresholds, highly sensitive to sensor placement, subject variability, and activity intensity. They also lacked contextual awareness and were unable to distinguish between falls and similar daily activities, such as quickly sitting or lying down.

A significant improvement was achieved by integrating multiple inertial sensors, including gyroscopes and accelerometers, and implementing data fusion techniques. A cutting-edge paper [246] proposed exploring the problem using angular accelerations, angular velocities, and changes in trunk angle, calculated by analyzing data provided by a gyroscope. If the values of the three variables are above their respective thresholds, the activity is reported as a fall. In [247], the fusion of gyroscope and accelerometer data was studied to classify

falls and non-falls more reliably. These allowed for more robust motion characterization through the use of angular velocities and accelerations. However, even fusion-based methods remained largely heuristic, relying on manually tuned thresholds rather than adaptive models. The inability to generalize across individuals or environmental contexts significantly limited their deployment outside controlled laboratory settings.

From the early 2010s onward, research shifted toward vision-based and ambient sensing, driven by advances in computer vision and the availability of low-cost RGB-D (depth) cameras. Depth sensors, such as the Microsoft Kinect, offered a privacy-preserving alternative to RGB cameras, allowing for detailed facial imagery while enabling three-dimensional motion capture. These systems provided richer contextual information than wearables, improving fall discrimination. Nevertheless, their sensitivity to lighting conditions, occlusions, and camera positioning constrained their real-world usability, particularly in domestic environments with variable furniture layouts and lighting [248–250].

Concurrently, radiofrequency (RF) based sensing emerged as a non-contact alternative using Wi-Fi or radar signals to infer body motion patterns [251, 252]. Such systems eliminated the need for body-worn devices, offering unobtrusive monitoring. However, RF sensing requires extensive calibration and is susceptible to interference from multipath reflections and environmental clutter.

Overall, while environmental and vision-based systems enhance contextual perception, they introduce trade-offs between accuracy, privacy, and scalability. In contrast, wearable sensors, especially accelerometers integrated into smartphones and smartwatches, remain the most practical choice for real-world deployment due to their affordability and ubiquity [253, 25]. A comprehensive review of work conducted on fall detection using different sensors is available in [24].

7.2 Methods

Algorithmically, fall detection approaches can be categorized into threshold-based, traditional ML, and deep learning methods.

- Threshold-based algorithms remain attractive for low-power embedded devices due to their interpretability and computational efficiency. However, they exhibit poor scalability and low adaptability to inter-person variability, leading to high false alarm rates when applied to diverse user populations.

- Traditional ML approaches (such as Support Vector Machines, Decision Trees, Hidden Markov Models) addressed some of these limitations by learning discriminative features from data [251–253, 25, 22]. Their reliance on handcrafted feature extraction, however, constrained performance to the designer's expertise. Moreover, these models often require careful feature normalization and are sensitive to sensor noise and orientation. Despite these drawbacks, classical ML remains relevant for low-data or embedded contexts, where model interpretability and energy efficiency are prioritized. Additionally, the research in [27] employed both traditional machine learning (kNN, QSVM, Ensemble Bagged Tree-EBT) and deep learning (ANN), and its experiments demonstrated that traditional machine learning approaches, in certain cases, outperform deep learning.

- The rise of Deep Learning marked a paradigm shift by enabling end-to-end feature learning directly from raw signals. Recurrent models, such as LSTMs, have demonstrated superior performance in handling sequential inertial data, achieving recall rates above 95% in multi-sensor fusion systems [26]. CNN-based architectures dominate vision-based approaches, effectively capturing spatial and temporal features from video data [248–250, 254]. However, deep models come with significant computational costs and require large, diverse, and well-annotated datasets, a notable limitation given that most available datasets involve limited subjects and controlled conditions [255]. This lack of diversity often leads to overfitting and limited generalization to unseen subjects or environments.

Recent studies have explored transformer-based architectures [256, 257] and graph neural networks (GNNs) [258, 259] to model human skeletal dynamics extracted from pose estimation, thereby achieving more robust spatiotemporal representations. These methods, however, remain data-hungry and computationally expensive, making their real-time deployment on wearable platforms challenging.

A promising new direction involves Reinforcement Learning (RL) and Deep Reinforcement Learning (DRL), which learn adaptive decision policies based on environmental feedback. While still in its infancy for fall detection, DRL has shown potential in fall prevention robotics and adaptive threshold optimization [28, 29, 23]. Nevertheless, RL-based approaches are limited by the need for extensive training data and the difficulty of simulating rare but safety-critical fall events. The trend toward multimodal sensor fusion, which combines inertial, visual, and ambient data in parallel, aims to enhance robustness across

various contexts. Hybrid approaches mitigate the limitations of individual sensors but raise challenges in synchronization, feature alignment, and energy consumption.

Finally, Edge-AI and federated learning paradigms have emerged as promising frameworks for privacy-preserving, distributed intelligence in healthcare monitoring. By processing data locally on devices and sharing only model parameters, these methods reduce data transmission overhead and safeguard user privacy [260–263]. However, the trade-off between model complexity and on-device computational capacity remains a crucial constraint.

7.3 Public datasets

Publicly available datasets are a key resource for benchmarking Fall Detection (FD) systems and comparing the performance of different sensor configurations, feature extraction methods, and ML/DL architectures. These datasets provide rich multimodal time-series data acquired from wearable inertial sensors, typically accelerometers and gyroscopes, under controlled conditions. They have been instrumental in enabling reproducibility and cross-comparison of fall detection algorithms, providing standard baselines for the field. Below is an overview of the primary and recent datasets [264].

- SisFall [31]: One of the largest and most cited FD datasets, including signals from accelerometers and gyroscopes recorded from 38 subjects performing 19 types of daily activities and 15 types of falls. It provides high-resolution inertial data but still involves simulated falls and lacks physiological measures.

- UP-Fall [32]: A multimodal dataset combining wearable inertial sensors, environmental sensors, and visual data. It allows testing of multimodal fusion algorithms and ML models, but is still limited by simulated conditions and a homogeneous participant pool.

- UMAFall [33]: Focused on controlled fall scenarios, integrating inertial and ambient sensors. While useful for method benchmarking, its controlled environment limits ecological validity.

- MobiFall / MobiAct [34]: Provides smartphone-based accelerometer and gyroscope data for falls and ADLs (Activities of Daily Living) from 66 subjects. It is one of the first large datasets to utilize mobile devices, demonstrating the feasibility of real-world deployment. However, it shares the same limitation of simulated falls by young subjects and lacks continuous, unconstrained activities.

- SmartFall [35]: Contains data from smartphones and smartwatch sensors across multiple activities and fall types, collected from 51 participants (30 young people and 21 elderly people). It provides useful benchmarks for cross-device robustness but lacks detailed physiological data and real-world noise conditions.

- UniMiB SHAR [36]: One of the most widely used smartphone-based human activity recognition datasets, containing 11,771 daily and fall activity samples performed by 30 subjects of ages ranging from 18 to 60 years. It includes data from elderly subjects, a rarity in public datasets, yet laboratory simulations still constrain it.

- KFall [37]: A comprehensive motion dataset using wearable inertial sensors attached to the lower back. It was initially acquired by 32 young subjects simulating 15 types of falls and 21 ADLs, and later supplemented with ADLs from 10 elderly subjects. It provides detailed, synchronized fall labels (onset and impact) but is mainly limited by simulated falls performed by young subjects.

- IPqM-Fall [38]: A dataset focused on fall detection in combat soldiers using smartwatches and smartphones. It includes simulated falls and daily activities from 15 soldiers, with the intention of developing a casualty detection system.

- FARSEEING (FAll Repository for the design of Smart and sElf-adaptive Environments prolonging Independent livinG) [39]: This is a meta-database and a project that established the most extensive collection of verified real-world fall data recorded with inertial sensors from older people. Its main strength is the ecological validity of real-world fall events (over 200 verified falls), overcoming the standard limitation of simulated falls. However, the data were collected in several studies with varying methods and populations, which can lead to heterogeneity.

Table 7.1 Summary and Comparison of Major Public Fall Detection Datasets

Dataset Name	Year	Sensor Type(s)	Key Participant Details	No. of Activities/Falls	Key Strength	Key Limitation
SisFall [31]	2017	Accelerometer, Gyroscope	38 subjects (23 young adults, 15 elderly, age 19-75)	19 ADLs / 15 Falls	Large-scale, including both young and elderly subjects.	Falls are simulated, and there are no physiological measures.
UP-Fall [32]	2019	Wearable IMUs, Ambient Sensors, Visual/Camera	17 healthy young individuals	11 Activities/Falls	Multimodal (inertial + ambient + visual), supporting fusion algorithms.	Falls are simulated, a homogeneous participant pool (young).
UMAFall [33]	2013	Inertial, Ambient Sensors	(Details vary, focused on controlled scenarios)	Falls/Near-falls/ADLs	Useful for method benchmarking and analysis of fall event stages.	Controlled environment limits ecological validity.
MobiFall / MobiAct [34]	2013	Smartphone Accelerometer, Gyroscope	66 subjects (age, height, gender vary, but generally young)	11 ADLs / 4 Falls	One of the first large datasets using mobile devices (real-world deployment feasibility).	Falls are simulated by young subjects, lacks continuous activities.
SmartFall [35]	2018	Smartphone and Smartwatch	51 participants (30 young and 21 elderly people)	Multiple ADLs / Fall Types	Benchmark for cross-device robustness and multi-sensor data from common devices.	Falls are simulated, lack detailed physiological data, and are contaminated with real-world noise.
UniMiB SHAR [36]	2016	Smartphone Accelerometer	30 subjects (age 18-60, includes some elderly subjects)	9 ADLs / 8 Falls	Widely used benchmark	Falls are laboratory simulations.
KFall [37]	2021	Accelerometer, Gyroscope, Magnetometer	32 young subjects initially, later plus 10 elderly with ADLs	21 ADLs / 15 Falls	Provides detailed, synchronized fall labels (onset and impact moments).	Falls are simulated (mainly by young subjects).
IPqM-Fall [38]	2023	Smartwatch and Smartphone	15 combat soldiers	ADLs / Military Op. / Falls	Focuses on casualty detection in a specific, high-risk user group.	A homogeneous group of young, fit subjects simulates falls.
FARSEEING [39]	2016	Inertial Sensors (various locations, mainly on lower back)	Older people (various cohorts)	>200 real-world falls	Highest ecological validity due to the collection of real-world falls.	Heterogeneity in data collection methods and populations across different studies.

Despite progress, existing public datasets exhibit limited generalization capacity due to several persistent shortcomings:

- Population bias: Most datasets involve young, healthy volunteers rather than workers, the elderly, or pathological subjects, making models poorly representative of at-risk populations.
- Scenario bias: Laboratory simulations omit environmental variability (uneven floors, obstacles, and lighting).
- Temporal fragmentation: Many datasets capture only short, isolated events, rather than continuous daily monitoring, which hinders modeling sequential pre-fall conditions.
- Accessibility and standardization: Several recent datasets remain under restrictive licenses, impeding open comparison and reproducibility.
- Modal limitations: Few datasets integrate physiological and biological measurements such as heart rate variability, respiration, or muscle activity, despite their role in detecting pre-fall fatigue or instability.

Consequently, while such datasets help understand the kinematics of falling, they fail to capture its etiology and the physical, physiological, and environmental causes that precede the event. The absence of naturalistic, diverse data constrains both model generalization and robustness.

7.4 SisFall

The SisFall dataset [31] is utilized to train and evaluate the proposed fall detection approach, as it is one of the largest and most diverse publicly available datasets for fall detection research. It includes recordings from 38 volunteers, consisting of 23 younger adults and 15 older adults (aged 60–75 years), as shown in Table 7.2. The older adult group performed only the 19 Activities of Daily Living (Table 7.3), while the younger subjects primarily simulated 15 falls (Table 7.4). Only one trained older adult (a Judo practitioner) performed fall scenarios. Additionally, a subset of the elderly participants abstained from certain activities due to personal or medical reasons. Each activity was repeated five times per subject.

In total, the dataset contains 2706 ADL sequences and 1798 fall sequences, recorded using three waist-worn inertial sensors sampled at 200 Hz: two accelerometers (Analog Devices ADXL345 and Freescale MMA8451Q) and one gyroscope (Texas Instruments ITG-3200).

Table 7.2 Demographic information of SisFall participants.

Subject	Gender	Age [y]	Height [cm]	Weight [kg]
SA01	F	26	165	53
SA02	M	23	176	58.5
SA03	F	19	156	48
SA04	M	23	170	72
SA05	M	22	172	69.5
SA06	M	21	169	58
SA07	F	21	156	63
SA08	F	21	149	41.5
SA09	M	24	165	64
SA10	M	21	177	67
SA11	M	19	170	80.5
SA12	F	25	153	47
SA13	F	22	157	55
SA14	F	27	160	46
SA15	F	25	160	52
SA16	F	20	169	61
SA17	M	23	182	75
SA18	M	23	181	73
SA19	M	30	170	76
SA20	F	30	150	42
SA21	M	30	183	68
SA22	F	19	158	50.5
SA23	F	24	156	48
SE01	M	71	171	102
SE02	F	75	150	57
SE03	F	62	150	51
SE04	F	63	160	59
SE05	M	63	165	72
SE06	M	60	163	79
SE07	M	65	168	76
SE08	F	68	163	72
SE09	M	66	167	65
SE10	F	64	156	66
SE11	F	66	169	63
SE12	M	69	164	56.5
SE13	M	65	171	72.5
SE14	M	67	163	58
SE15	F	64	150	50

Table 7.3 Types of activities of daily living selected in SisFall.

Code	Activity
D01	Walking slowly
D02	Walking quickly
D03	Jogging slowly
D04	Jogging quickly
D05	Walking upstairs and downstairs slowly
D06	Walking upstairs and downstairs quickly
D07	Slowly sit in a half-height chair, wait a moment, and up slowly
D08	Quickly sit in a half-height chair, wait a moment, and up quickly
D09	Slowly sit in a low-height chair, wait a moment, and up slowly
D10	Quickly sit in a low-height chair, wait a moment, and up quickly
D11	Sitting a moment, trying to get up, and collapsing into a chair
D12	Sitting a moment, lying slowly, wait a moment, and sit again
D13	Sitting a moment, lying quickly, wait a moment, and sit again
D14	Being supine, moving to a lateral position, waiting a moment, and moving back
D15	Standing, slowly bending at knees, and getting up
D16	Standing, slowly bending without bending knees, and getting up
D17	Standing, get into a car, remain seated, and get out of the car
D18	Stumble while walking
D19	Gently jump without falling (trying to reach a high object)

Table 7.4 Types of falls selected in SisFall

Code	Activity
F01	Fall forward while walking caused by a slip
F02	Fall backward while walking caused by a slip
F03	Lateral fall while walking caused by a slip
F04	Fall forward while walking caused by a trip
F05	Fall forward while jogging caused by a trip
F06	Vertical fall while walking caused by fainting
F07	Fall while walking, with the use of hands on a table to dampen a fall caused by fainting
F08	Fall forward when trying to get up
F09	Lateral fall when trying to get up
F10	Fall forward when trying to sit down
F11	Fall backward when trying to sit down
F12	Lateral fall when trying to sit down
F13	Fall forward while sitting, caused by fainting or falling asleep
F14	Fall backward while sitting, caused by fainting or falling asleep
F15	Lateral fall while sitting, caused by fainting or falling asleep

The dataset includes labels identifying whether each recorded sequence corresponds to a fall or a non-fall event. However, the dataset does not provide precise temporal boundaries indicating when the fall begins or ends within the continuous signal.

This lack of frame-level annotation presents a significant challenge for training real-time ML/DL models, as the exact onset of a fall must be accurately identified to supervise temporal models, such as recurrent neural networks. For this reason, a manual annotation effort was conducted by the authors of [265] and their colleagues. Standard annotation criteria were agreed upon, and each recording was labeled independently by multiple annotators, followed by iterative cross-review to reduce inter-annotator variability. Despite these precautions, manual annotation of high-frequency sensor sequences can result in temporal misalignment between the label and the true event, potentially reducing the achievable classification performance.

In [265], SisFall sequences were annotated into three classes: FALL, ALERT, and NO-FALL. The ALERT category corresponds to transitional instability events, such as stumbling with recovery. To partially address the inherent dataset imbalance between fall and non-fall samples, the ALERT class is merged with the FALL class in the present work.

7.5 Fall Detection Neural Network (FDNN)

One of the main contributions of this research is the FDNN (Fall Detection Neural Network) model, which was developed to overcome the limitations of generalization and robustness inherent in traditional fall detection approaches, specifically threshold-based methods and classical Machine Learning (ML) techniques that rely on manual feature engineering.

Deep Learning (DL) models enable automated feature learning directly from raw signals, capturing the complex temporal dependencies that distinguish a fall from a dynamic, yet stable, Activity of Daily Living (ADL). To address the sequential, temporal, and high-variability nature of inertial data, a Recurrent Neural Network (RNN) architecture was adopted. RNNs are recognized for their exceptional ability to overcome the vanishing gradient problem and capture long-term dependencies in sequential data. In the context of a fall, this is crucial for analyzing both the initial phase of instability and the rapid acceleration that precedes impact.

The final architecture (illustrated in Figure 7.1) is the result of an iterative and focused design process. Preliminary analysis using non-temporal models showed that they struggled to distinguish dynamic, high-acceleration events from true falls reliably. This ambiguity

demanded a model with an "internal memory" to capture the signal's evolution over time. Three main variations were examined:

- Long Short-Term Memory(LSTM): Exhibited strong temporal modeling capabilities, but a medium computational cost and training time.
- Bidirectional Long Short-Term Memory (BiLSTM): Improved performance slightly compared to LSTM due to processing sequences bidirectionally, but its double-pass structure made it twice as computationally heavy and impractical for low-latency scenarios.
- Gated Recurrent Unit (GRU): Achieved performance slightly lower than the LSTM while being more lightweight.

Considering the advantages and disadvantages, LSTM layers (Layers 4 and 6) were chosen to offer a compromise between performance and computational costs. It differs from FFNN in Section 8, where greater importance was given to the model's lightweight. Unlike many DL approaches that only operate on raw signals, the FDNN model integrates both subject-level metadata (age, height, weight, gender) and derived inertial signals (orientation and body-to-gravity angle). This fusion enriches the input space (Table 7.5), allowing the network to learn movement patterns that are more specific to the individual and the context. Batch Normalization (Layer 2) and Dropout (Layers 3, 5, and 7 at a 50% rate) were included to stabilize training and prevent overfitting. This is vital given the trade-off between the need for high sensitivity (TPR) and the crucial requirement for a low False Alarm Rate (TNR).

The final nine-layer network, adapted from the recurrent model proposed in [265], processes the multivariate input (19 channels standardized using z-score normalization) for binary classification:

- Layer 1: Fully Connected (FC) layer for initial input projection.
- Layer 2: Batch Normalization layer for input stability.
- Layer 3, 5, 7: Dropout layers (50%) for regularization.
- Layer 4 and 6: Two stacked LSTM layers for capturing complex temporal dependencies, with an internal dimension of 16 units.
- Layer 8: Second FC layer before classification.
- Layer 9: SoftMax output producing the probability of a fall event, $P(falling)$.

Table 7.5 List of FDNN features.

Symbol	Description
y	Age of the subject
h	Height of the subject
w	Weight of the subject
g	Gender of the subject
$\alpha_{x,ADXL345}$	Acceleration data in the X axis measured by the sensor ADXL345
$\alpha_{y,ADXL345}$	Acceleration data in the Y axis measured by the sensor ADXL345
$\alpha_{z,ADXL345}$	Acceleration data in the Z axis measured by the sensor ADXL345
$\alpha_{x,MMA8451Q}$	Acceleration data in the X axis measured by the sensor MMA8451Q
$\alpha_{y,MMA8451Q}$	Acceleration data in the Y axis measured by the sensor MMA8451Q
$\alpha_{z,MMA8451Q}$	Acceleration data in the Z axis measured by the sensor MMA8451Q
$\dot{\alpha}_{x,ITG3200}$	Rotation data in the X axis measured by the sensor ITG3200
$\dot{\alpha}_{y,ITG3200}$	Rotation data in the Y axis measured by the sensor ITG3200
$\dot{\alpha}_{z,ITG3200}$	Rotation data in the Z axis measured by the sensor ITG3200
q_1	Orientation data in Quaternion, the first element
q_2	Orientation data in Quaternion, the second element
q_3	Orientation data in Quaternion, the third element
q_4	Orientation data in Quaternion, the fourth element
θ	Body angle relative to the direction of gravity
$\dot{\theta}$	Body angular acceleration relative to the direction of gravity

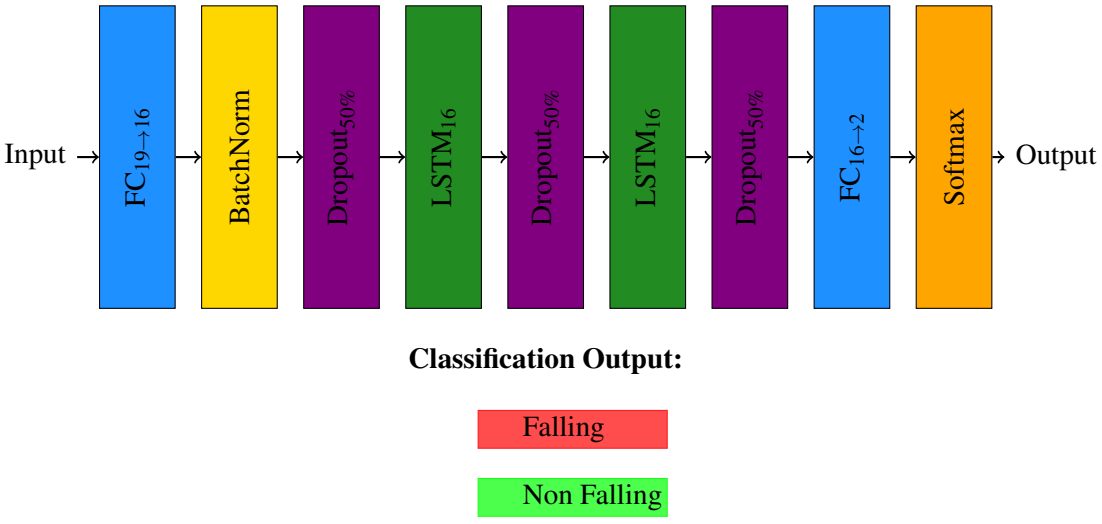


Figure 7.1 Architectural diagram of a nine-layer Deep Neural Network (DNN) for binary classification. The network processes the input through fully connected (FC) layers, batch normalization (BatchNorm), and intermediate hidden layers utilizing LSTM and Dropout regularization. The final layer uses a Softmax activation after a fully connected layer to produce probabilities for the two classes: Falling and Non-Falling.

This configuration proved to be the best compromise between temporal modeling capability and computational efficiency required for potential implementation on smart wearable devices. A fall is detected when $P(falling) > 0.5$.

For training, only sequences containing falls are used, resulting in approximately 1800 sequences. The dataset is randomly partitioned into training (60%), validation (20%), and test (20%) subsets, not splitting for subject. The NN is trained using the Training dataset to adjust its weights, and the validation dataset is used to select the optimal weights, which occur when the NN achieves the highest accuracy on the Validation dataset. The training is performed for 64 epochs using a batch size of 128.

7.6 Results

After training, the model is evaluated on 360 fall sequences from the test subset. As depicted in Figure 7.2, accuracy (ACC) is generally high but is influenced disproportionately by the True Negative Rate (TNR) due to data imbalance. Therefore, Balanced accuracy (bACC = TPR/2 + TNR/2) is reported, which accounts for both TPR and TNR. The accuracy is usually excellent, mainly because the TNR is very high. However, the accuracy is not a good comparison metric because the dataset is not balanced; the TNs weigh more than the TPs. Indeed, the graph of ACC is very similar to the chart of TNR. Instead, the bACC is a more balanced and explanatory metric than accuracy. In this case, the bACC is always more than 80%, except for Sequences 44, 84, 199, 258, and a few others.

Figures 7.3 and 7.4 illustrate the TNR/TPR as functions of the subjects and activities, respectively. The tables display the average value across the repetitions. Blank cells indicate that there are no samples for those subjects and activities in the Test dataset. In particular, the subjects SE01-SE05 do not simulate the Falling events, as described in the SisFall paper [31]. Remember that, being an annotation a manual procedure, the start and end of the actual fall may not precisely match the time defined in the annotated dataset. The TNR value (Figure 7.3) is typically greater than 95% and, on average, 98.4%. The TPR value (Figure 7.4) is, on average, 82.6%. In many cases, the FDNN succeeds in perfectly classifying the fall; however, there are some instances where the TPR is not significantly higher than 50%. Additionally, it is crucial to note that since the fall is a brief event, failure to classify even a few samples significantly lowers the TPR. Analyzing the different activities, activity F10 (Falling forward when trying to sit down) is the most difficult to classify correctly.

The model is then evaluated on the 2701 no-fall sequences (ADLs). Figure 7.5 shows the TNR as a function of subject and activity. The TNR value is typically 100%, but in some

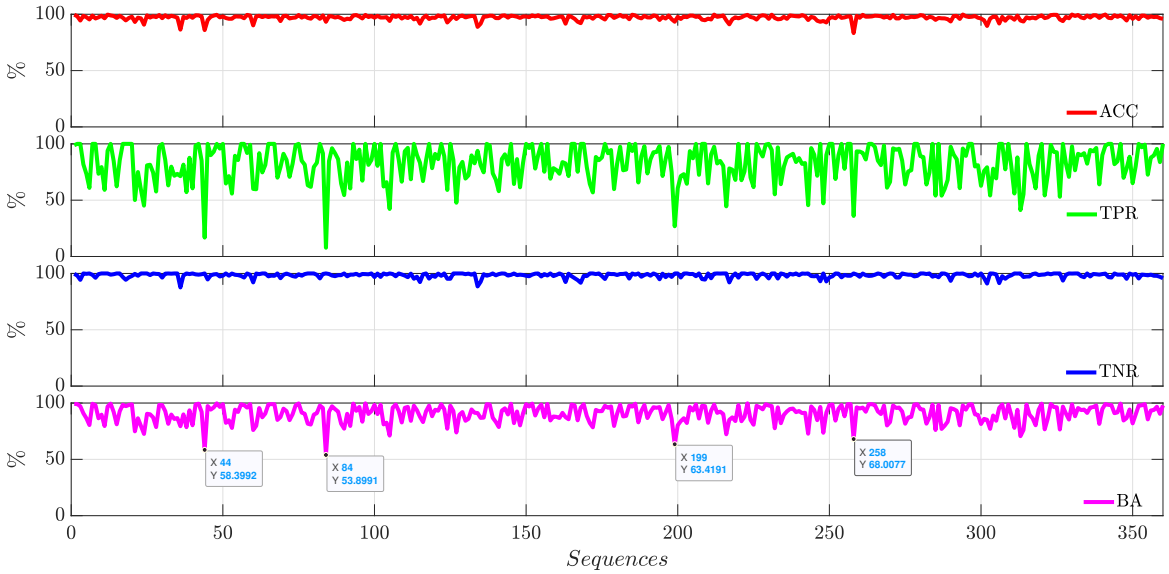
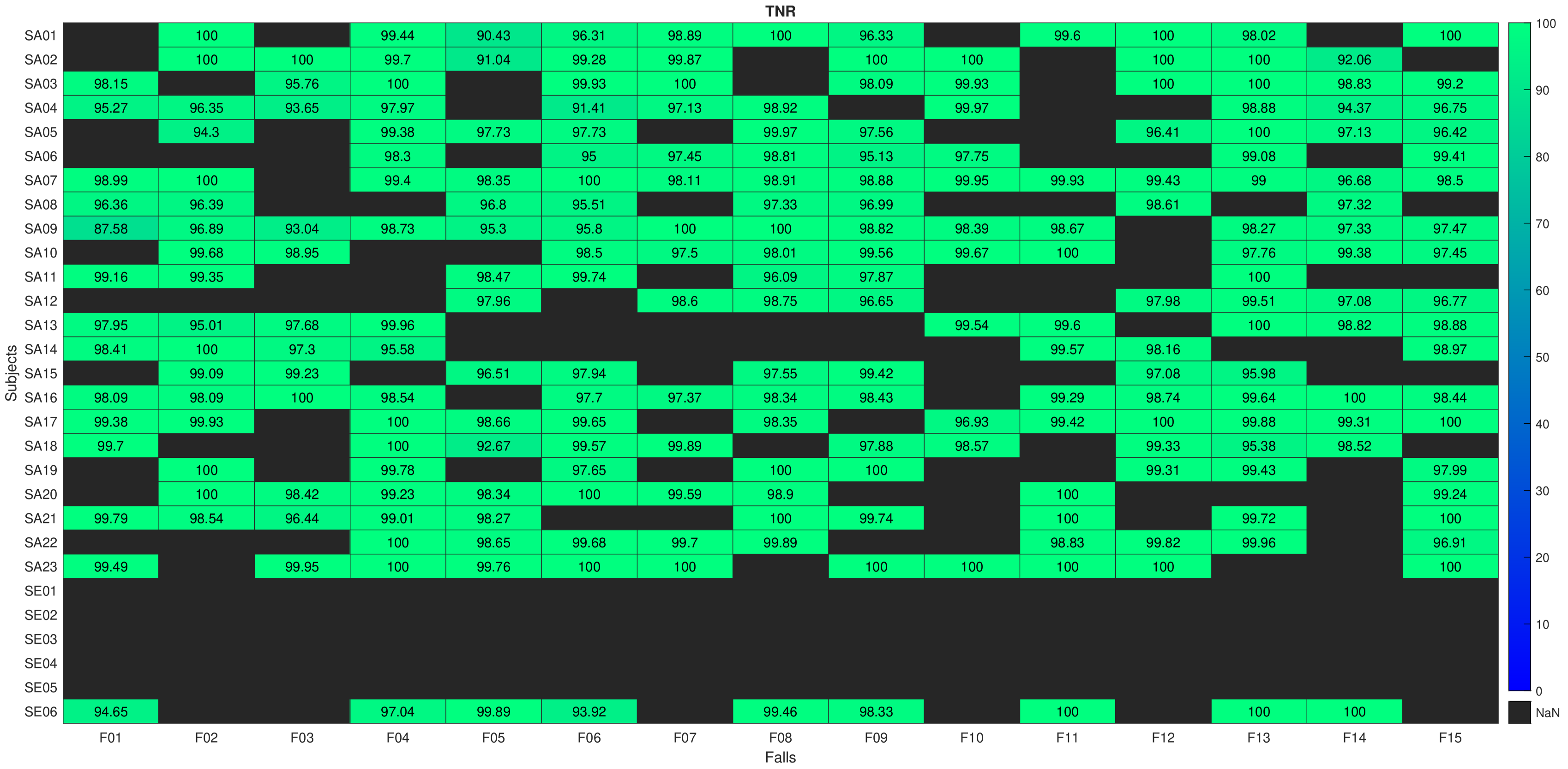


Figure 7.2 Accuracy, True Positive Rate, True Negative Rate, and Balanced Accuracy for Test Dataset.

cases, it is also less than 20%. Particularly for activities with dynamic postural transitions: D03 (Jogging slowly), D04 (Jogging quickly), D12 (Sitting a moment, lying slowly, wait a moment, and sit again), D13 (Sitting a moment, lying quickly, wait a moment, and sit again) and D14 (Being on one's back change to lateral position, wait a moment, and change to one's back). The worst activity for the TNR metric is D04 (Jogging quickly). The overall average TNR across ADLs is 94.4%.



	F01	F02	F03	F04	F05	F06	F07	F08	F09	F10	F11	F12	F13	F14	F15
SA01		100		99.44	90.43	96.31	98.89	100	96.33		99.6	100	98.02		100
SA02		100	100	99.7	91.04	99.28	99.87		100	100		100	100	92.06	
SA03	98.15		95.76	100		99.93	100		98.09	99.93		100	100	98.83	99.2
SA04	95.27	96.35	93.65	97.97		91.41	97.13	98.92		99.97			98.88	94.37	96.75
SA05		94.3		99.38	97.73	97.73		99.97	97.56			96.41	100	97.13	96.42
SA06				98.3		95	97.45	98.81	95.13	97.75			99.08		99.41
SA07	98.99	100		99.4	98.35	100	98.11	98.91	98.88	99.95	99.93	99.43	99	96.68	98.5
SA08	96.36	96.39			96.8	95.51		97.33	96.99			98.61		97.32	
SA09	87.58	96.89	93.04	98.73	95.3	95.8	100	100	98.82	98.39	98.67		98.27	97.33	97.47
SA10		99.68	98.95			98.5	97.5	98.01	99.56	99.67	100		97.76	99.38	97.45
SA11	99.16	99.35			98.47	99.74		96.09	97.87				100		
SA12					97.96		98.6	98.75	96.65			97.98	99.51	97.08	96.77
SA13	97.95	95.01	97.68	99.96						99.54	99.6		100	98.82	98.88
SA14	98.41	100	97.3	95.58							99.57	98.16			98.97
SA15		99.09	99.23		96.51	97.94		97.55	99.42			97.08	95.98		
SA16	98.09	98.09	100	98.54		97.7	97.37	98.34	98.43		99.29	98.74	99.64	100	98.44
SA17	99.38	99.93		100	98.66	99.65		98.35		96.93	99.42	100	99.88	99.31	100
SA18	99.7			100	92.67	99.57	99.89		97.88	98.57		99.33	95.38	98.52	
SA19		100		99.78		97.65		100	100			99.31	99.43		97.99
SA20		100	98.42	99.23	98.34	100	99.59	98.9			100				99.24
SA21	99.79	98.54	96.44	99.01	98.27			100	99.74		100		99.72		100
SA22				100	98.65	99.68	99.7	99.89			98.83	99.82	99.96		96.91
SA23	99.49		99.95	100	99.76	100	100		100	100	100	100			100
SE01															
SE02															
SE03															
SE04															
SE05															
SE06	94.65			97.04	99.89	93.92		99.46	98.33		100		100	100	

Figure 7.3 TNR as a function of subject and fall activity.

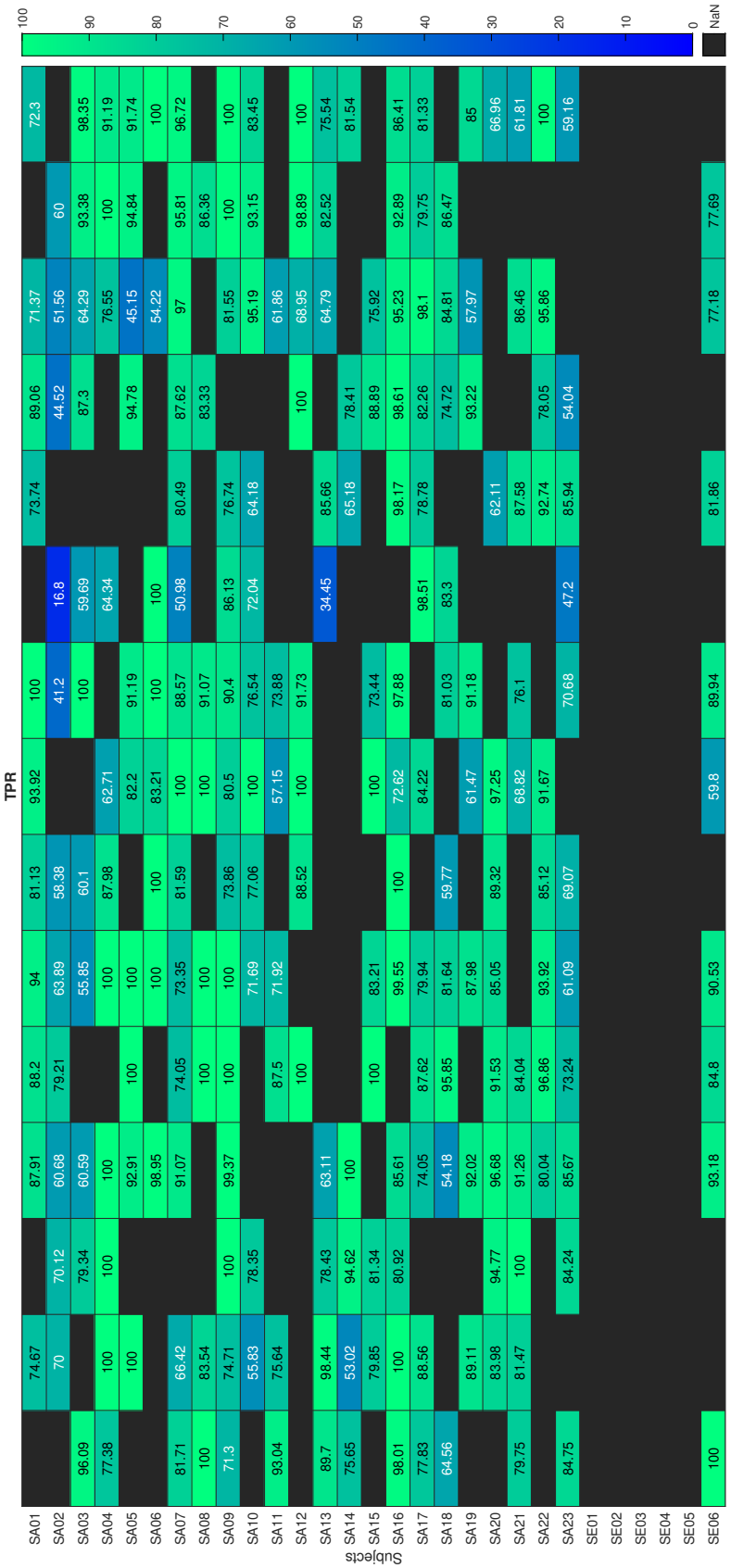


Figure 7.4 TPR as a function of subject and fall activity.

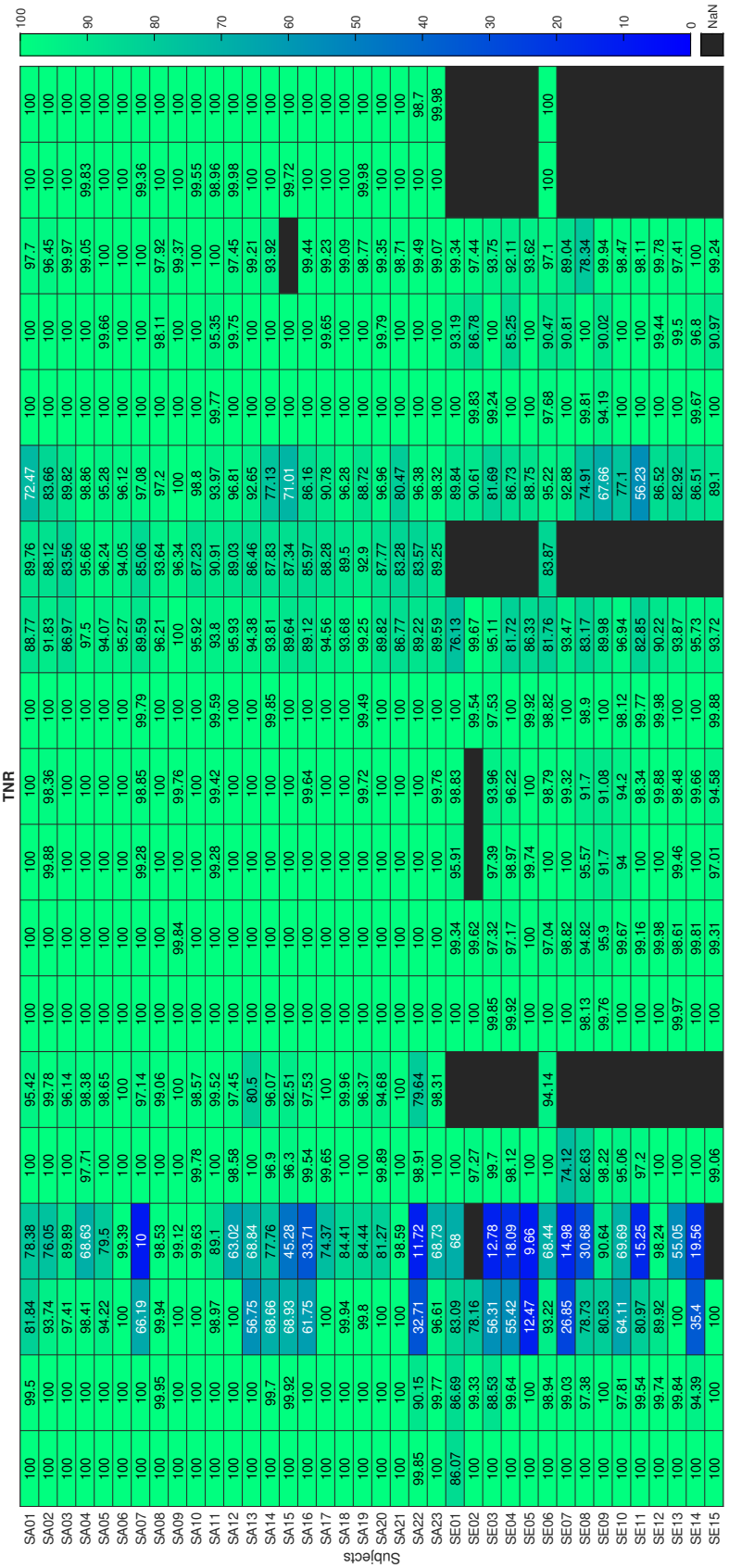


Figure 7.5 TNR as a function of subject and ADL.

7.7 Discussion

The Fall Detection Neural Network (FDNN) developed in this chapter provides a robust model for the binary classification of falls versus ADLs using multimodal inertial sensor data from the SisFall dataset. The results demonstrate a significant strength in minimizing false alarms, which is arguably the most critical operational requirement for a widely deployed wearable system.

7.7.1 Performance Interpretation and Comparison

The average True Negative Rate (TNR) of 94.4% across diverse ADLs (and 98.4% on the non-fall part of the fall dataset) indicates high specificity. This is a crucial achievement, as high TNR ensures user acceptability and prevents alert anxiety, which often plagues threshold-based systems. This suggests that the Deep Learning architecture, particularly the LSTM layers, has successfully learned the complex temporal signatures that distinguish dynamic but stable movements from unstable, potentially injurious ones.

With an average True Positive Rate (TPR) of 82.6%, the model demonstrates strong sensitivity. While the overall balanced accuracy (bACC) remained consistently above 80%, this moderate TPR indicates that a portion of true fall events were missed (False Negatives). This trade-off suggests that the model parameters favored specificity over sensitivity during optimization, a justifiable choice in consumer wearables but a potential concern in high-risk clinical settings.

7.7.2 Limitations

The performance analysis highlighted two main areas of difficulty. The model exhibited a notable drop in TNR during highly dynamic activities, such as Jogging Quickly (D04), where the TNR dropped significantly below the average. Other dynamic postural transitions (D12, D13, D14, Sitting/Lying) also resulted in lower TNR values. This difficulty confirms a well-known challenge: movements involving rapid changes in velocity or large, sudden accelerations often mimic fall kinematics, leading to false alarms. The most challenging fall type to classify correctly was falling forward when trying to sit down (F10), which resulted in a low TPR. This suggests that falls that are "cushioned" or incomplete, where the body's momentum is partially absorbed by compensation or contact with an object (such as a chair), lack the distinct high-impact signature that the network relies on for immediate classification. These limitations are partly inherent to the use of simulated fall data (Scenario Bias) and the

reliance on a single body-worn Inertial Measurement Unit (IMU). Simulated falls may not fully capture the complexity and compensatory movements of real-world falls, and the lack of continuous physiological context (which will be explored in Chapter 9) limits the ability to predict pre-fall states like syncope or fatigue.

7.8 Conclusions

The Fall Detection Neural Network (FDNN) developed in this chapter successfully addresses the challenge of distinguishing between genuine fall events and high-impact Activities of Daily Living (ADLs) using multimodal inertial data from the SisFall dataset.

The LSTM-based deep learning architecture demonstrated a robust capability in the binary classification task. Specifically, it achieved a high average True Negative Rate (TNR) of 94.4% across ADLs, fulfilling the critical requirement for minimizing false alarms in a deployable system. The model also delivered a competitive average True Positive Rate (TPR) of 82.6%. However, the performance analysis confirmed persistent challenges in accurately classifying highly dynamic ADLs and subtle, cushioned falls, underscoring the limitations of purely kinematic models trained on simulated data.

Ultimately, while the FDNN provides a reliable foundation for reactive fall notification, it does not fulfill the crucial need for proactive fall mitigation. To bridge the gap between detection and injury prevention, the research focus must also include the accurate prediction of fall trajectory and Time-to-Impact. This more complex, temporal challenge is the subject of the subsequent chapter.

Chapter 8

Fall Impact Prediction

This chapter adds the missing element to obtain a comprehensive system for addressing the problem of STFs, focusing on transforming fall safety from a reactive to a proactive approach, a necessary step for enabling life-saving pre-impact interventions, such as wearable airbags. The methodologies presented here are designed to accurately estimate the remaining time before a fall impact (Time-to-Impact) and forecast the subject's fall trajectory (Fall Forecasting).

8.1 Background

Research on fall safety has historically focused on detecting a fall after it has occurred, primarily relying on threshold-based methods, inertial signatures of impact, or activity classification algorithms applied retrospectively [21, 23, 26, 27]. While these approaches play a critical role in post-fall assistance, they do not prevent injuries; instead, they only enable notification or response after ground contact. In contrast, the present work shifts the focus from reactive fall detection to proactive fall mitigation, aiming to provide sufficient lead time for a wearable airbag or other protective mechanism to deploy before impact.

Fall impact prediction is particularly challenging due to substantial variability in human movement, environmental factors, and subject-specific responses. To be practically deployable, a fall impact prediction framework must adapt to the diverse characteristics of users of different ages, body types, and fitness levels; operate reliably across heterogeneous environments; and minimize false positives, which may otherwise discourage users from wearing the protective device. Furthermore, unlike systems designed for older adults [24, 25, 29, 30], mitigation technologies targeting workers at height must accommodate

faster reaction times, larger movement variability, and the dynamic, unpredictable conditions typical of construction, maintenance, and industrial settings.

Workplace falls remain among the leading causes of severe injury and fatality worldwide. Workers perform repetitive, high-load, and cognitively demanding tasks, often at elevated heights, under time pressure, or in constrained environments. Recent reports from Eurostat and OSHA [1] confirm that falls from height continue to represent over one-third of fatal workplace accidents, with younger and more agile workers often attempting reflexive recovery strategies that differ significantly from those observed in older adults. This behavioral divergence creates substantial challenges for pre-impact prediction: agile subjects frequently exhibit compensatory movements, such as grabbing railings, braking with limbs, or altering their center of mass, which generate transient kinematic patterns that can be easily misinterpreted as falls, thereby increasing the risk of premature or incorrect airbag deployment.

A significant limitation in existing pre-impact fall research is the insufficient modeling of subject reaction. Classical approaches assume passive, ballistic-like motion once instability begins. However, empirical studies conducted by our group and not yet published show that healthy individuals often attempt to arrest or redirect their falls, generating nonlinear trajectories that complicate prediction. Consequently, accurate estimation of time-to-impact requires models that can capture real-time dynamics, human reflexes, and biomechanical variability while remaining computationally lightweight enough for embedded deployment.

8.1.1 Kolmogorov-Arnold Network (KAN)

A key innovation in this research is the application of the Kolmogorov–Arnold Network (KAN) to estimating the time of fall impact. Grounded in the Kolmogorov–Arnold representation theorem, KANs express multivariate functions as superpositions of univariate functions, offering a more interpretable functional decomposition than conventional multilayer perceptrons. Poluektov and Polar [266] introduced a practical construction of KAN using basis functions and continuous spline-parameterized mappings, along with an efficient Newton–Kaczmarz solver for parameter estimation. More recently, Liu et al. [267] demonstrated that KANs outperform MLPs on numerous regression tasks while requiring significantly fewer learnable parameters, owing to their learnable activation functions.

The rapid rise of KAN research since 2024 has led to domain-specific extensions highly relevant to fall prediction. Vaca-Rubio et al. [268] introduce a new application of KANs to time series forecasting, using their adaptive activation functions to improve predictive model-

ing. They show that KANs outperform conventional MLPs in a real-world satellite traffic prediction task, providing more accurate results with significantly fewer learnable parameters. Bozorgasl and Chen [269] present Wav-KAN, a novel neural network architecture that incorporates wavelet functions into the KAN structure, enabling the network to efficiently capture both high-frequency and low-frequency components of the input data. The wavelet-based approximations utilize an orthogonal or semi-orthogonal basis, striking a balance between accurately representing the underlying data structure and minimizing noise overfitting. Abueidda et al. [270] introduced a new version of neural operators called DeepOKAN, which uses KANs rather than conventional neural network architectures. DeepOKAN is used to develop surrogates for various mechanics problems, noting that it requires fewer learnable parameters than current MLP-based DeepONets to achieve comparable accuracy. Genet and Inzirillo [271] proposed a new neural network architecture, Temporal Kolmogorov-Arnold Networks (TKAN), which combines the strengths of KAN and LSTM. This innovation allows us to predict multi-stage time series more accurately and efficiently. Finally, Kiamari et al. [272] presented Graph Kolmogorov-Arnold Networks (GKANs), a new neural network architecture that extends the principles of recently proposed KANs to graph-structured data. Unlike traditional Graph Convolutional Networks (GCNs), which are based on a fixed convolutional architecture, GKANs implement learnable functions based on splines between layers, transforming the way information is processed across the graph structure and achieving higher accuracy in semisupervised learning tasks on graphs than the traditional GCN model.

This growing body of evidence suggests that KANs offer an ideal architecture for real-time fall impact prediction: lightweight, interpretable, functionally expressive, and capable of modeling nonlinear multivariate relationships with minimal overfitting. This thesis leverages these advantages to address the long-standing challenge of accurately estimating time-to-impact during an ongoing fall.

8.2 Methodology

The concept of fall impact prediction focuses on determining when a falling subject hits the ground or another surface (impact) quantitatively. This is a crucial, proactive step beyond simple fall detection and is a prerequisite for activating pre-impact mitigation systems, such as wearable airbags, which require milliseconds for full deployment.

The methodologies developed in this chapter address this task using two alternative or complementary approaches: a novel, interpretable approach based on the Kolmogorov-Arnold Network (Fall-KAN) for time-to-impact estimation and a Recurrent Neural Network

(RNN) model leveraging Inverted Pendulum (IP) biomechanical simulation for future state forecasting (Fall Forecasting Neural Network, FFNN)

8.2.1 Fall-KAN

The Fall-KAN (Fall impact time estimation Kolmogorov-Arnold Network) approach [43] is designed to provide an interpretable and highly accurate estimate of the time remaining until impact. It is grounded in the mathematical framework of the Kolmogorov–Arnold representation theorem.

Kolmogorov-Arnold Network (KAN)

This section provides an overview of the theoretical framework and implementation details of the Kolmogorov–Arnold Network (KAN) used in this study to estimate the Time-to-Impact, following the approach proposed by Poluektov and Polar [266]. The model is grounded in the Kolmogorov–Arnold representation theorem, a foundational result in functional analysis demonstrating that any continuous multivariate function can be expressed as a finite superposition of continuous univariate functions [273, 274].

More formally, any continuous function $F : \mathbb{R}^m \to \mathbb{R} \in C[0,1]^m$ can be represented as:

$$F(X_1, X_2, \ldots, X_m) = \sum_{k=1}^{2m+1} \Phi^k \left(\sum_{j=1}^{m} f^{kj}(X_j) \right), \tag{8.1}$$

where $f^{kj} : \mathbb{R} \to \mathbb{R} \in C[0,1]$ and $\Phi^k : \mathbb{R} \to \mathbb{R} \in C(\mathbb{R})$.

This theorem provides the mathematical foundation for constructing neural network architectures that can learn general nonlinear relationships without relying on traditional layered perceptron structures. In particular, KANs can approximate any continuous function using additive and compositional transformations of univariate mappings, resulting in interpretable architectures with strong generalization capabilities and reduced potential for overfitting.

In the context of data-driven modeling, this decomposition can be viewed as a hierarchical tree of discrete Urysohn operators, where a single root operator (representing the global function) is composed of $2m+1$ branch operators, each combining m univariate transformations. The internal functions of the model can thus be defined as:

$$f^{kj}(x) = \sum_{p=1}^{n} H_{kjp} \phi^p(x), \tag{8.2}$$

$$\Phi^k(t) = \sum_{l=1}^{q} G_{kl}\,\psi^l(t), \tag{8.3}$$

where H_{kjp} and G_{kl} are the model parameters, $\phi^p : [x_{min}, x_{max}] \to \mathbb{R}$ and $\psi^l : [t_{min}, t_{max}] \to \mathbb{R}$ are compact-support basis functions and $t_{min}, t_{max} \in \mathbb{R}$. The functions ϕ^p are required to be continuous, while ψ^l must be continuous and differentiable almost everywhere. The parameters n and q denote the number of basis functions used for the internal and external mappings, respectively.

The identification problem in KAN consists of estimating both sets of parameters (H_{kjp} and G_{kl}) and the basis functions themselves. The Newton–Kaczmarz method [275, 276] is applied to solve the nonlinear system iteratively, providing efficient convergence toward an approximate solution. In practice, piecewise linear basis functions are used for both $\phi^p(x)$ and $\psi^l(t)$, offering a balance between computational simplicity and expressive power.

Data processing

For this study, the SisFall dataset [31] was used as a benchmark. From each 15-second-long trial, only the portion corresponding to the actual fall phase was extracted, excluding preliminary and final stationary segments (such as walking, sitting, or resting after impact). Following Musci et al. [265], this segment, denoted as FALL, extends from the onset of instability to the point where the standard deviation of the acceleration signal drops below a predefined threshold, marking the moment of impact. The time to impact is computed as the time difference (a multiple of the sampling period, 5 ms) between the current observation and the moment of ground contact (see Figure 8.7).

To limit feature dimensionality and improve interpretability, a fully data-driven feature selection pipeline was implemented using two complementary techniques:

- Correlation analysis to eliminate weakly correlated variables,
- Minimum Redundancy Maximum Relevance (mRMR) [277, 278] to maximize the mutual information between selected features and the response variable while minimizing redundancy among them.

Correlation analysis is the simplest and, very often, one of the most effective methods for feature reduction. It can be used in various ways, such as when two features have a high correlation, allowing one of them to be eliminated. Alternatively, if one feature is poorly correlated with the value to estimate, it can also be eliminated. In this study, the second approach was adopted, and features with correlation scores below a defined threshold were

excluded. The mRMR algorithm identifies a set of mutually dissimilar and maximally distinct features, allowing them to represent the response variable effectively. The algorithm seeks to maximize the relevance of the features to the response variable while minimizing redundancy within the feature set. The method uses mutual information measures, including the mutual information between a feature and the response variable, as well as between variables and features. For each feature, a score is calculated to assess its importance. Only features with both scores (correlation and mRMR) above the thresholds were considered.

The KAN identification process employed piecewise linear basis functions and the Newton–Kaczmarz iterative solver. Key model hyperparameters included the number of internal basis functions n in (8.2), external basis functions q in (8.3), the regularization parameter (μ) of the Newton–Kaczmarz method [266], and the time window (w) used to forecast the time to impact. All hyperparameters were optimized via cross-validation, leveraging the five-repetition structure of SisFall: four repetitions were used for training and validation, and the fifth repetition was reserved for testing to ensure subject-level independence.

8.2.2 Fall Forecasting Neural Network (FFNN)

The concept of Fall Forecasting refers to the real-time estimation of a subject's future state during the process of falling, assuming that the fall will continue unless external intervention occurs. Unlike traditional fall detection systems, which classify an event as a fall only after it occurs, fall forecasting aims to anticipate the moment of ground impact and enable preventive actions, such as deploying protective mechanisms (like wearable airbags) or issuing early alerts to assistive systems.

In this study, a deep learning approach was developed to extrapolate the subject's orientation over time and identify when the risk of impact exceeds a predefined safety threshold. Unlike the models trained directly on wearable sensor data, the training dataset for this forecasting neural network was synthetically generated using a biomechanical simulation based on the Inverted Pendulum (IP) model, a simplified but well-established abstraction of human postural dynamics [279, 280]. While more sophisticated full-body musculoskeletal models are available, such as multi-segment, muscle-driven formulations implemented in platforms like OpenSim [281], these approaches require extensive subject-specific calibration, high-fidelity motion capture, and significant computational resources, which limit their suitability for real-time wearable applications and large-scale data generation.

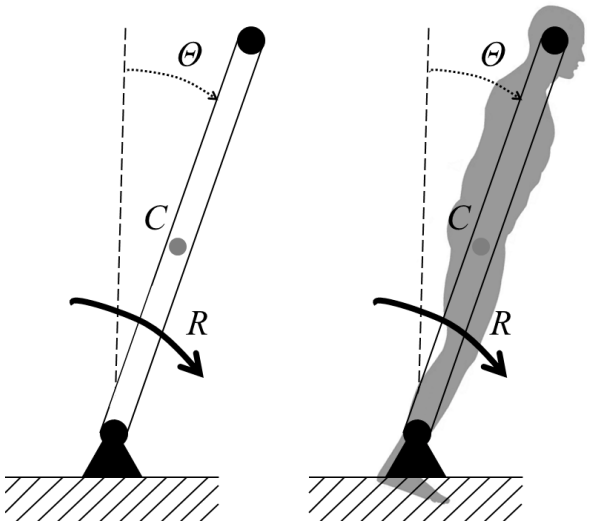


Figure 8.1 model of human as an Inverted Pendulum.

Inverted pendulum (IP) modeling

Figure 8.2 illustrates multiple simulations of the IP model, representing subjects falling from various initial postures. This model offers a computationally efficient method for generating large, diverse datasets for training, with controlled variations in physical and environmental parameters. Two primary types of falls are modeled:

- Tripping: Occurs when a subject's foot collides with an obstacle, leading to a forward rotation around a fixed pivot point. In this scenario, the IP model assumes a fixed base at the point of contact (Figure 8.1). Although tripping and collapsing are kinematically distinct, both involve a rotational descent about a pivot.

- Slipping: Arises from a loss of traction between the foot and the ground, causing the base of support to move backward. This is modeled as an IP with a floating base and a dynamic ground-contact model, where the coefficient of friction determines the onset of instability.

To ensure variability and realism, the simulation framework includes:

- Generation of IP instances with different dynamic parameters (such as inertia, height, and center of mass (CoM) position to represent inter-subject variability.

- Parameterization of ground contact properties by varying the friction coefficient to simulate different flooring materials or environmental conditions.

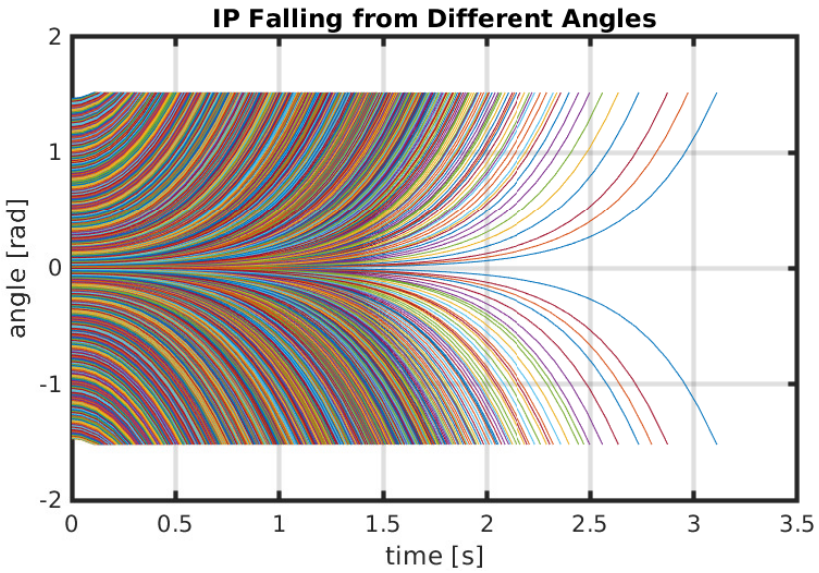


Figure 8.2 Example of an IP falling from different initial angles and velocities.

- Initialization of simulations with diverse CoM angles and angular velocities, corresponding to different pre-fall states.

This approach enables the generation of tens of thousands of synthetic sequences, capturing a wide range of plausible human falling behaviors. Consequently, the FFNN model benefits from a rich training dataset, which is free from the ethical and practical constraints associated with human subject experimentation.

Architecture and Training

For the forecasting task, a Gated Recurrent Unit (GRU)-based recurrent neural network (RNN) was adopted. GRUs were chosen over LSTM architectures because they achieve comparable accuracy with fewer parameters, offering faster inference and lower memory requirements, a crucial feature for embedded real-time systems [282].

The input vector to the network includes:

- Two subject-specific constants (height and weight),
- Two initial dynamic states (fall angle and angular velocity), and
- A time series of subsequent body angles relative to gravity.

All inputs are normalized using Z-score standardization to ensure uniform feature scaling. The architecture, summarized in Figure 8.3, comprises:

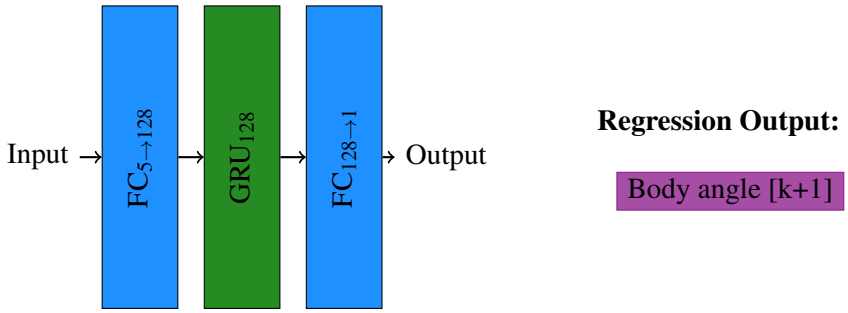


Figure 8.3 Architectural diagram of the three-layer Recurrent Neural Network (RNN) for regression. The network processes the input through fully connected (FC) layers and an intermediate hidden layer utilizing GRU. The final layer uses a fully connected layer to produce the estimation of the angle at the next sample. It is possible to forecast the body angle by feeding the input with the output in a closed circuit.

- A fully connected layer (Layer 1) that processes the inputs,
- A GRU layer (Layer 2) with an internal dimension of 128 units,
- A fully connected layer (Layer 3) that outputs the forecasted orientation angle.

The dataset generated from IP simulations comprises approximately 79,000 sequences, split into training (70%), validation (20%), and testing (10%) subsets. Training follows the same optimization procedure used for the fall detection model, with a batch size of 128 and 128 epochs. Given that wearable airbags typically require 80–100 ms to reach full deployment, the forecasting horizon was set to 200 ms, ensuring sufficient lead time for activation of a fall mitigation mechanism.

8.3 Results

This section presents the empirical validation and performance evaluation of the algorithms developed for real-time proactive fall response: the Fall-KAN for time-to-impact estimation and the Fall Forecasting Neural Network (FFNN) for trajectory prediction. The evaluation focuses on core metrics, such as Root Mean Squared Error (RMSE), for regression tasks, and comparative performance against baseline machine learning models, demonstrating the efficacy of both data-driven feature selection and biomechanically informed deep learning architectures.

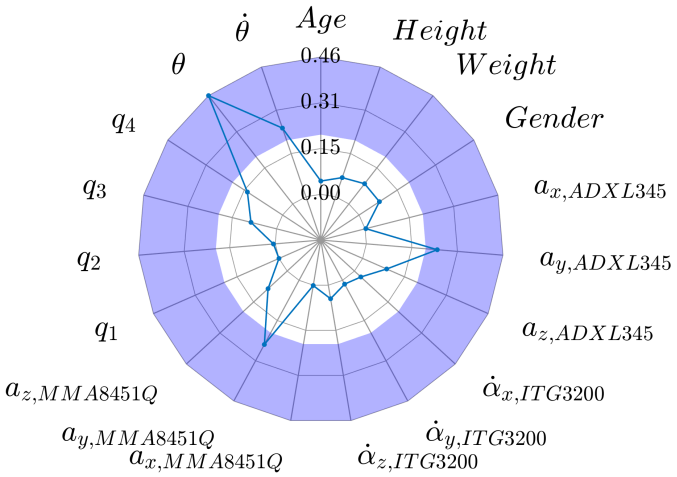


Figure 8.4 Feature correlation scores.

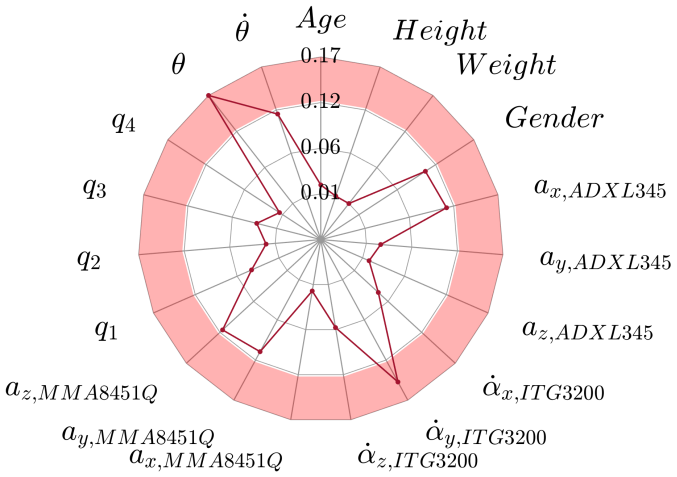


Figure 8.5 mRMR correlation scores.

8.3.1 Fall-KAN

This section reports the implementation outcomes and compares the performance of the proposed Fall-KAN model with several established regression techniques, including Linear Regression, Regression Trees, Support Vector Machines (SVMs), and Multilayer Perceptrons (MLPs).

As anticipated in Section 8.2.1, features with correlation coefficients below a threshold of 0.2 were excluded in the correlation-based step. This procedure retained four highly relevant features: $a_{y,ADXL345}$, $a_{y,MMA8451Q}$, θ and $\dot{\theta}$ (see Figure 8.4). Similarly, the mRMR algorithm was applied with a relevance threshold of 0.12, selecting two additional features, $\dot{\alpha}_{y,ITG3200}$ and θ, as shown in Figure 8.5.

The data were standardized by subtracting the mean and dividing by the standard deviation, ensuring feature comparability. The model coefficients were identified using a 4-fold cross-validation scheme, where three folds were used for training and one for validation, iterating across combinations of hyperparameters:

- $n = 2^j,\ j = 1,\ldots,6$;
- $q = 2^i,\ i = 1,\ldots,6$;
- $\mu = 2^{-k},\ k = 1,\ldots,8$;
- $w = 10,\ldots,50\ ms$.

The best configuration was obtained for $n = 4$, $q = 64$, $\mu = 0.0625$, and $w = 50\ ms$. The training procedure required only one epoch, facilitating direct comparison with traditional ML-based regression models.

Figure 8.6 presents the RMSE distribution across subjects and fall types for the optimal model. On average, the RMSE across all 1,798 falls is approximately $150ms$, indicating high temporal accuracy in estimating the time to impact. Performance variations across subjects (such as SA10) and specific fall types (such as F13) are consistent with prior findings in the literature, where inter-subject variability and activity complexity influence prediction accuracy. Some outlier cases, such as combinations SA05–F13 and SA21–F11, exhibit RMSE values up to four times higher than the mean; however, such deviations are also observed with conventional regression methods. The results of which will not be shown in full in the next section, as space is devoted to other considerations.

Conversely, falls such as F05 and F15 show particularly robust performance, confirming that Fall-KAN effectively generalizes across diverse fall dynamics. In subjects SA03, SA06, and SA12, the model achieves excellent consistency, accurately predicting impact time across almost all scenarios.

To enable fair benchmarking, all comparative models were trained and cross-validated under identical conditions. In the linear regression, no hyperparameter was set. The Regression Tree used a minimum leaf size of 36, thereby balancing the bias–variance trade-off. A very leafy tree tends to overfit, and its validation accuracy is often much lower than its training accuracy. In contrast, a coarse tree with fewer large leaves does not achieve high training accuracy. However, a coarse tree can be more robust, as its training accuracy can be close to that of a representative test set.

For the SVM, several kernels (linear, polynomial, Gaussian) were tested, with the linear kernel selected due to its superior computational efficiency. The MLP was configured with

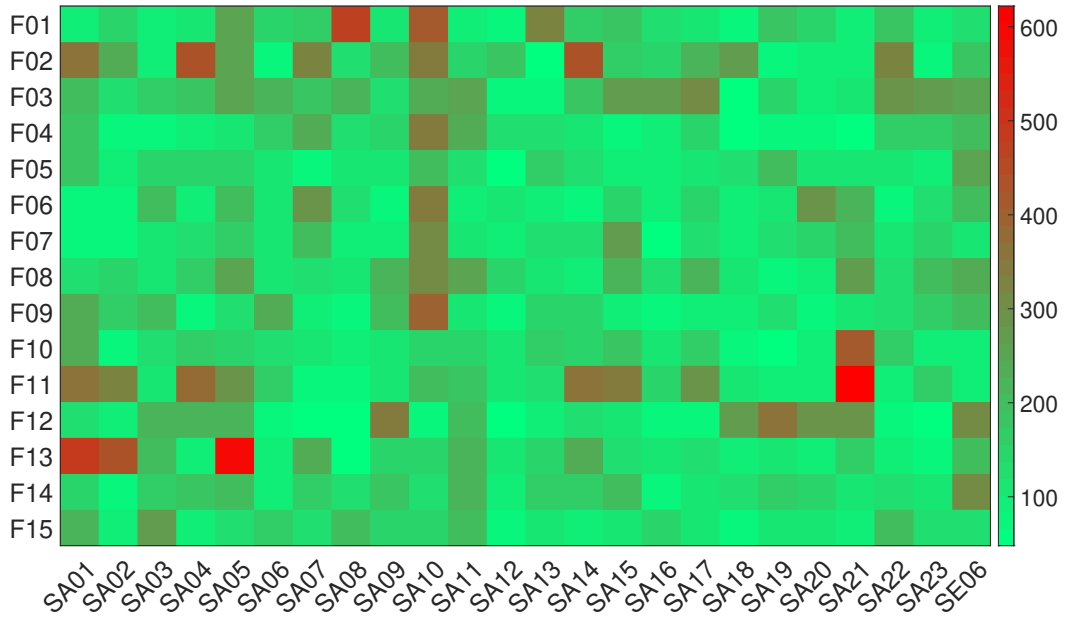


Figure 8.6 Heatmap about RMSE for different subjects and falls [*ms*].

Table 8.1 Comparison between Fall-KAN, Linear Regression, Regression Tree, SVM, and Neural Network.

	Fall-KAN	Linear Regression	Regression Tree	SVM	MLP
$RMSE_{mean}[ms]$	***159***	176	201	164	190
$RMSE_{median}[ms]$	***133***	151	175	*133*	167
$R^2_{median}[ms]$	***0.35***	0.19	-0.01	0.29	-0.01

two hidden layers, mirroring the KAN structure ($2m+1 = 101$ nodes in the first layer and $m = 50$ nodes in the second layer), and employed the ReLU activation function.

All models were trained on a workstation equipped with an Intel Xeon Gold 5220R CPU, 128 GB RAM, and an NVIDIA RTX A6000 GPU. The results, summarized in Table 8.1, show that Fall-KAN achieves the lowest mean RMSE among all tested methods and exhibits one of the lowest median RMSE values, comparable only to the SVM. The median R^2 score is also higher for Fall-KAN, underscoring its stronger explanatory power and generalization capability. In the Fall-KAN case, the mean RMSE is the average of the values shown in Figure 8.6.

Finally, the performance obtained by Fall-KAN and the other methods in the time domain is compared, illustrating the trend of the actual time of impact (Ground Truth) and the projected time of impact. Figure 8.7 compares the temporal predictions of all models for subject 8 performing fall F13 (Forward fall while sitting). The ground-truth time to impact decreases linearly over approximately 600 ms. Most models exhibit an initial transient

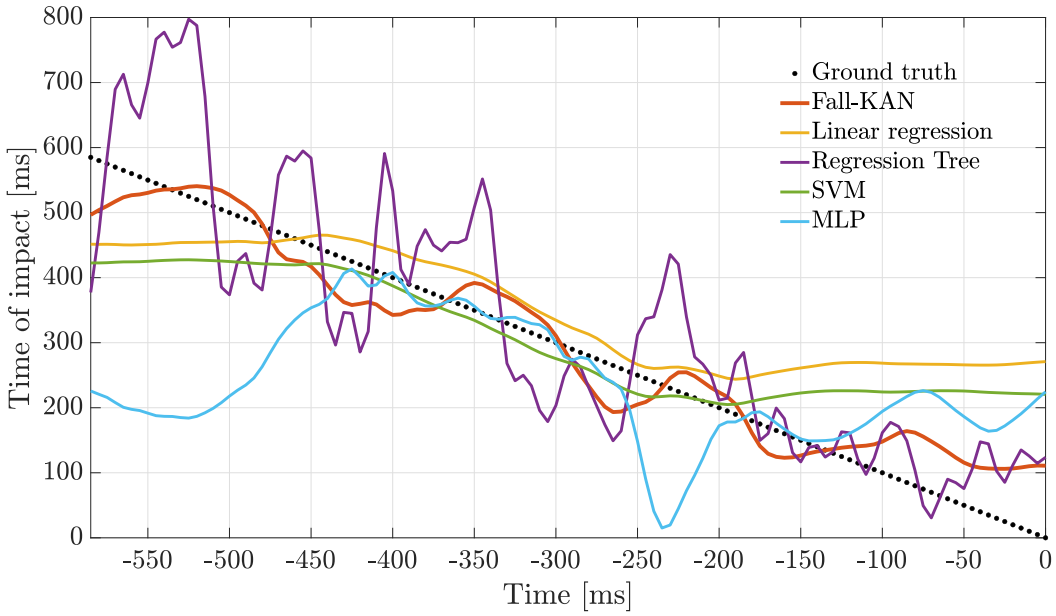


Figure 8.7 Time of impact for the 8th subject during the 13th fall.

followed by a final plateau when the time of impact is too low, with linear regression serving as a baseline reference. The regression tree yields very noisy results that deviate significantly from a straight-line trend, making it one of the least effective methods for this application. Additionally, it is observed that the MLP suffers from a longer settling time than the other methods, resulting in unsatisfactory performance in the case studied, where timeliness is crucial. The SVM performs better but still demonstrates slight final biases. A low error at the beginning of the fall and a final settling value that deviates less than that obtained by the other compared methods.

In contrast, Fall-KAN provides the most consistent and smooth estimation, with minimal initial delay and minor deviations at the end of the fall. Although small oscillations around the reference line persist, the trend is markedly more stable and aligned with the true trajectory, confirming the robustness and suitability of Fall-KAN for real-time fall impact estimation.

8.3.2 Fall Forecasting Neural Network (FFNN)

This section reports the implementation outcomes and compares the performance of the proposed Fall Forecasting Neural Network (FFNN). Figure 8.8 presents an example of the forecasting output for a simulated fall, where the GRU-based network successfully extrapolates the pendulum's future orientation. The real-time estimate (red line) shows the model's ability to predict in the short term (1 sample in advance). In contrast, the forecasted

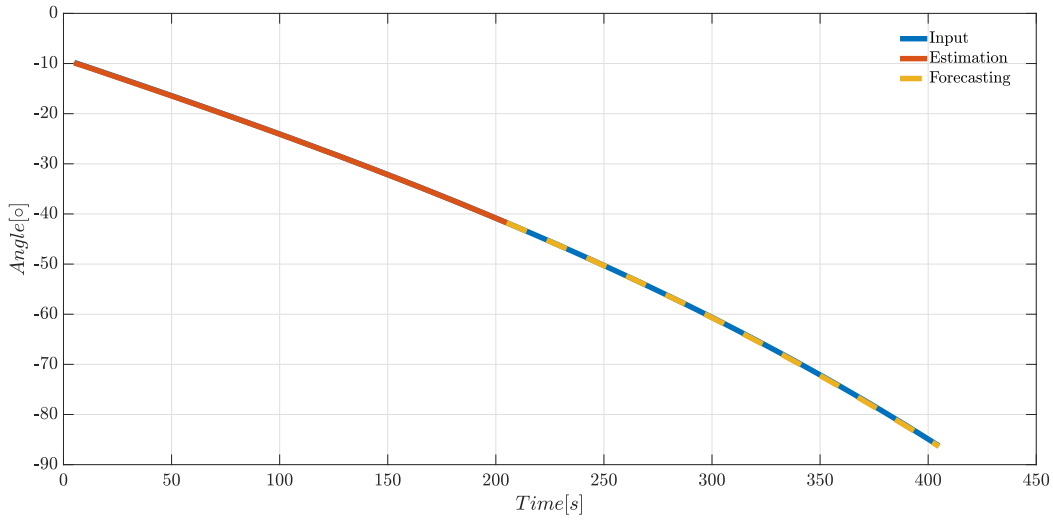


Figure 8.8 Example of fall forecasting of an IP falling with a non-zero initial angle and velocity.

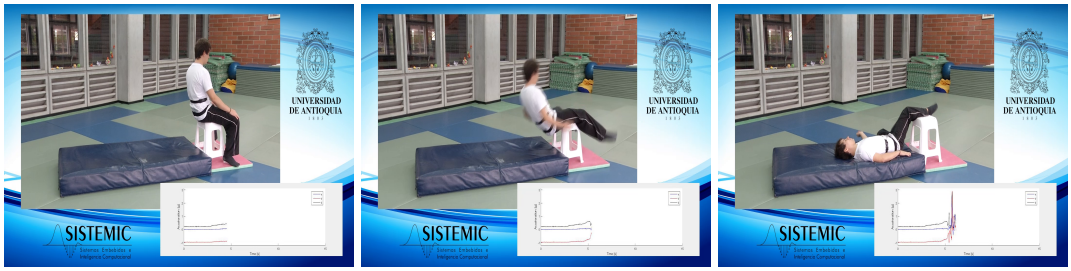


Figure 8.9 Frames from SisFall video - Fall backward while sitting, caused by fainting or falling asleep.

trajectory (yellow dashed line) demonstrates the model's ability to accurately estimate the long-term trajectory. Both closely align with the ground-truth motion (blue line). The system assumes that ground contact occurs when the body reaches an angle of approximately 85° (1.48 rad) relative to the vertical axis.

To assess real-world applicability, the forecasting network was tested in cascade with the Fall Detection Neural Network (FDNN) using sequences from the SisFall dataset. Once a fall event is detected, the forecasting module is triggered to predict the time to impact. This two-stage approach performs best in falls consistent with IP-like dynamics, such as tripping, collapsing, or backward falls (F14). In this scenario, the subject's chair acts as a rotational pivot, and gravity becomes the dominant force after balance loss. Some video frames of the fall are shown in Figure 8.9.

Figures 8.10 and 8.11 show the integrated system's performance on the F14 sequence. The first chart illustrates the falling probability estimated by the FDNN, which rises sharply

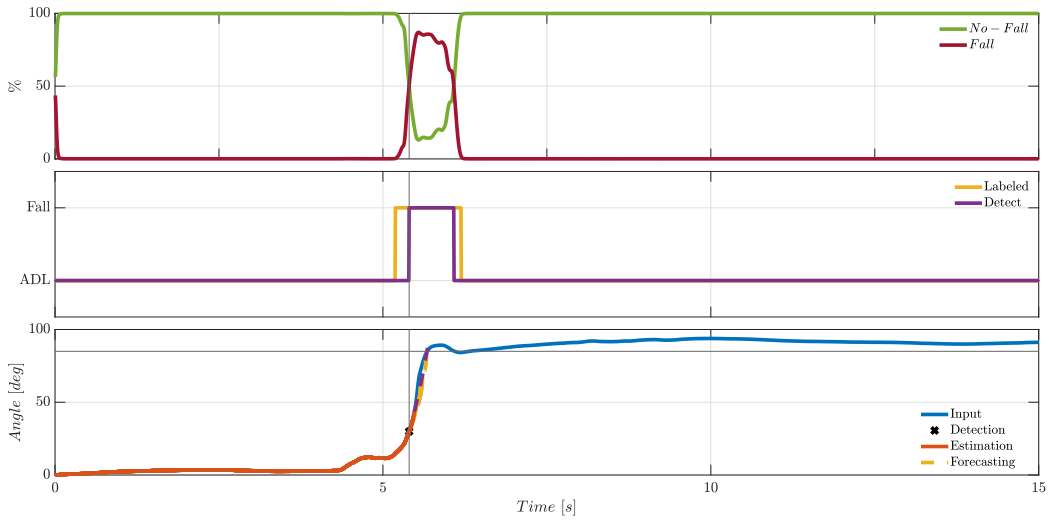


Figure 8.10 Falling probability, Fall Detection and Fall Forecasting.

during the initial phase of instability. The subsequent plots show the evolution of body angle and its forecast over time. Notably, the network underestimates ground contact by approximately 50 ms, which is advantageous for real-time mitigation, providing adequate time for an airbag system to deploy before impact. Overall, this hybrid detection–forecasting pipeline demonstrates the feasibility of combining data-driven deep learning with physics-based modeling to predict and mitigate falls proactively.

8.4 Discussion

The work presented in this chapter successfully extends the fall problem from reactive detection (Chapter 7) to proactive, millisecond-scale prediction, a necessity for deploying life-saving mitigation systems such as wearable airbags. This was achieved through the development and comparison of two distinct, alternative/complementary methodologies: the interpretable Fall-KAN for time-to-impact estimation and the biomechanically informed FFNN for trajectory forecasting.

8.4.1 Fall-KAN

The Fall-KAN model demonstrated superior performance over all tested traditional Machine Learning baselines (Linear Regression, Regression Tree, SVM, and MLP) in estimating the time remaining until impact. The mean RMSE of approximately 159 ms confirms high temporal accuracy, a critical factor for systems requiring rapid response. Notably, the use of

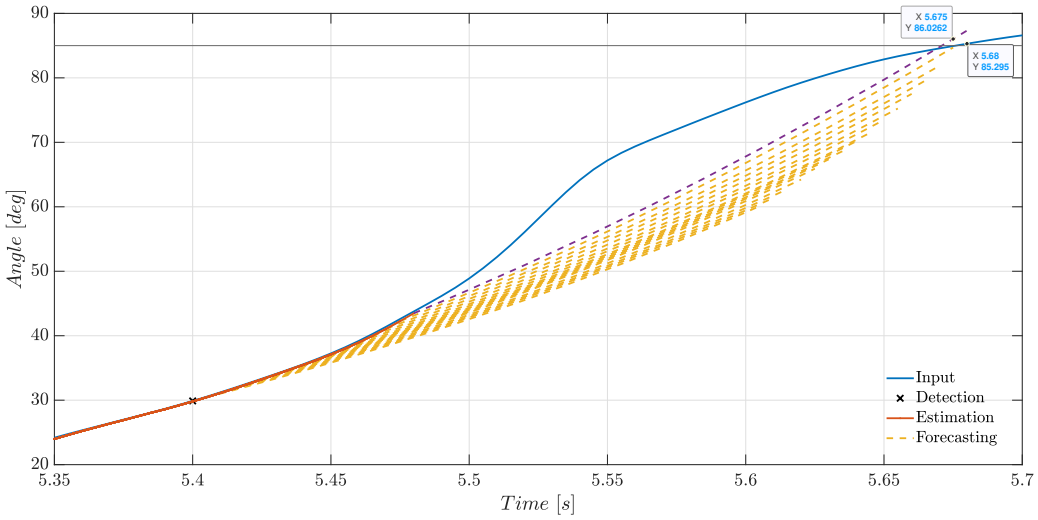


Figure 8.11 Angle time-series of the falling subject's angle relative to the gravity vector.

a Kolmogorov–Arnold Network architecture yielded a more stable and less noisy estimation trend compared to conventional models, particularly during the critical final moments before impact (Figure 8.7). This stability is a direct advantage of KANs, which are theoretically grounded in functional analysis and tend to learn smooth and interpretable relationships.

The data-driven feature selection pipeline (Correlation and mRMR) proved effective in distilling the most informative kinematic variables ($\alpha_{y,ADXL345}$, $\alpha_{y,MMA8451Q}$, $\dot{\alpha}_{y,ITG3200}$, θ, and $\dot{\theta}$), underscoring that a small set of well-chosen features derived from orientation and angular change is sufficient for accurate regression.

8.4.2 Fall Forecasting Neural Network (FFNN)

The FFNN, utilizing a GRU-based Recurrent Neural Network trained on data generated from the Inverted Pendulum (IP) model, showcased the feasibility of real-time trajectory extrapolation. By leveraging physics-based simulation, we overcame the challenge of acquiring extensive real-world fall sequences for a forecasting task. The network's ability to forecast the subject's orientation 200 ms into the future (Figure 8.8) provides a lead time significantly exceeding the required deployment threshold of 80–100 ms for current mitigation systems.

When tested in cascade with the FDNN, the hybrid system successfully anticipated ground contact (underestimating it by ≈ 50 ms in the example of Figure 8.11), which is favorable for early-stage intervention. This integration validates the concept of a complete proactive pipeline, moving from detection to high-speed prediction within the fall event.

8.4.3 Limitations and Future Directions

Despite these advances, the work possesses limitations that define future research scope.

Both the Fall-KAN and FFNN rely on the SisFall dataset (simulated falls) and the highly idealized Inverted Pendulum model. Real-world falls, particularly complex, multi-segmental, or compensatory movements, introduce kinematics that the IP model simplifies. Future work should focus on integrating dynamic parameters derived from real sensor data to personalize the forecasting model.

The Fall-KAN excels in interpretability and localized accuracy near impact, while the FFNN offers a robust, continuous trajectory forecast. A key future direction is the fusion of these approaches to leverage the best of both worlds: interpretable high-accuracy estimation and continuous trajectory prediction.

The demonstrated success of both the Fall-KAN and the FFNN confirms that combining advanced and specialized neural architectures with biomechanical modeling is a powerful strategy for transforming reactive fall detection into proactive, injury-mitigating systems.

8.5 Conclusions

Taken together, the fall detection and fall impact prediction modules form a unified framework that enables both immediate recognition of a fall and a short-term estimation of its evolution. Rather than replacing detection, forecasting complements it, allowing the system to progress from simple event recognition to proactive, time-critical mitigation. Two novel, complementary methodologies were developed and validated using inertial sensor data:

- Fall-KAN: A novel application of the Kolmogorov–Arnold Network achieved the lowest overall RMSE ($\approx$ 159 ms) compared to traditional models for estimating the time remaining until impact. This demonstrated a stable, highly accurate, and interpretable regression trend essential for instantaneous mitigation decisions.

- FFNN: The Fall Forecasting Neural Network utilized a GRU architecture trained on an artificially enriched dataset generated via Inverted Pendulum (IP) simulation. This approach validated the feasibility of extrapolating the falling trajectory, providing a crucial preemptive time window ($\approx$ 200 ms lead time) necessary for activating safety mechanisms, such as wearable airbags.

In conclusion, the work establishes a robust, two-stage proactive pipeline, Fall Detection followed by immediate Fall Impact Prediction, that successfully meets the stringent temporal

demands of pre-impact intervention. While the current findings rely on kinematic data and simplified physics, they lay a vital foundation for the next generation of wearable safety devices, setting the stage for the exploration of physiological risk precursors in the subsequent chapters.

Chapter 9

Fall Risk Estimation

This chapter enlarges the focus from reactive detection (Chapter 7) and proactive forecasting (Chapter 8) of an ongoing fall event to fall risk estimation and prevention. This necessitates a broader, holistic approach that monitors the physiological, biomechanical, and cognitive factors contributing to falls.

9.1 Introduction

While fall detection and forecasting systems primarily rely on inertial signals to determine whether a person is falling or not, fall prevention/fall-risk estimation systems require a broader and more holistic approach. Instead of responding to an ongoing event, these methods aim to anticipate and mitigate the likelihood of a fall by monitoring the physiological, biomechanical, and cognitive factors contributing to balance instability. Fall risk estimation is an inherently multifactorial process, encompassing cardiac, neurological, sensory, muscular, and environmental dimensions.

As reported in Section 2.2, falls can originate from numerous intrinsic and extrinsic causes. Among the most significant are cardiovascular dysfunctions (such as arrhythmias, orthostatic hypotension), neurological disorders (such as peripheral neuropathy, Parkinsonian symptoms), sensory impairments (such as visual or auditory loss), musculoskeletal deficits (such as sarcopenia, gait instability), and pharmacological or metabolic conditions (such as diabetes, osteoporosis, or medication side effects). Psychological factors, such as stress, fatigue, and cognitive decline, also significantly contribute to increased fall risk.

9.1.1 Cardiac and Respiratory monitoring

Cardiovascular anomalies can cause syncope or dizziness, making heart-rate-related monitoring essential for preventing falls. Wearable sensors, such as ECG, SpO2, and respiration monitors, enable real-time tracking of heart rate (HR), heart rate variability (HRV), and respiration rate (RR), which are parameters associated with fatigue and autonomic regulation.

For example, some wearable systems activate alarms or send emergency alerts when heart rate thresholds are exceeded or drop dangerously low, as demonstrated in early work using waist-mounted monitors [283]. Other studies have shown that individuals with frequent premature ventricular contractions (PVCs) exhibit a higher incidence of falls, emphasizing the potential of cardiac monitoring as a predictive biomarker [284]. More recent approaches [285] integrate machine learning algorithms to detect pre-syncope or fainting episodes from physiological data, demonstrating a strong link between early cardiovascular anomalies and fall vulnerability.

9.1.2 Gait, Muscle Activity, and Balance Assessment

Gait irregularities, postural instability, and lower-limb muscle weakness are among the most extensively studied predictors of falls. Modern wearable devices, particularly inertial measurement units (IMUs) and surface electromyography (sEMG) sensors, enable continuous assessment of body dynamics.

Rescio, Leone, and Siciliano pioneered the use of IMU and EMG for evaluating fall risk. In their latest work [286], they prototyped an intelligent sock that detects EMG signals from the gastrocnemius and tibialis anterior muscles of the legs, making it suitable for assessing lower limb-related pathologies, such as changes in gait, sarcopenia, and fall risk. A machine learning scheme was chosen to overcome the known drawbacks of threshold approaches widely used in pre-fall systems. In [287], the authors developed a framework to study the feasibility of fall detection based on sEMG and accelerometers. Their approach demonstrated high accuracy in activity recognition, demonstrating the feasibility of the proposed method in daily activity awareness. Xi et al. [288] developed feature extraction and pattern recognition methods using EMG for monitoring activities of daily living and fall detection. Their study showed that a system with four sEMG sensors can achieve high sensitivity and specificity with low classification errors. Saadeh et al. [289] used the data collected from an IMU sensor to distinguish between ADLs and pre-fall events. Their system accurately predicted fall events, triggering a fall risk alarm before they occurred.

9.1.3 Neurological and Cognitive Monitoring

Beyond motor and cardiovascular factors, falls can result from neurological and cognitive causes, such as impaired attention, mental fatigue, or Freezing of Gait (FoG) in Parkinson's disease. EEG sensors, often used in conjunction with EMG or IMUs, can reveal neurophysiological signatures of stress, fatigue, and reduced alertness.

Annese et al. [290] proposed a cyber-physical system that integrates EEG and EMG signals to compute an online fall-risk coefficient, reflecting the combined influence of neural and muscular dynamics. Similarly, EEG-based methods have been used to monitor workers' stress levels or cognitive overload in high-risk occupations, such as construction [291]. The results demonstrate that the proposed method for field stress recognition can be utilized for the early detection of worker stress using wearable EEG devices, thereby enhancing workers' safety, health, well-being, and productivity. Jeon et al. [292] relate EEG signal patterns to the types of construction hazards and develop an EEG classifier based on experiments conducted in an immersive virtual reality (VR) environment in which hazards of different types (such as falls and slips) were simulated in a VR environment. The research results support the widespread identification of risks, thereby contributing to improved workplace safety in construction. In [293], a similar study was conducted to investigate and understand the effects of working at height on the development of mental fatigue for fall hazard prevention using EEG and VR.

Other research [294] presented a drowsiness detection scheme that fuses behavioral information from user movement, as measured by an IMU sensor, and physiological data from brain activity, as captured by a single EEG electrode. The solution is a wearable device that can detect five different levels of sleepiness. For patients with Parkinson's disease, neural network–based EEG classifiers showed promise for the early detection of FoG episodes, enabling proactive interventions before gait freezing and subsequent falls occur [295].

9.1.4 Overview

There are very few studies that attempt to detect imminent falls from measured biomedical signals. Accurate fall risk assessment demands an integrated, continuous, and personalized approach that fuses diverse physiological and biomechanical signals. As highlighted in recent reviews [296, 297], multimodal data fusion, combining inertial, EMG, ECG, and EEG information, enhances predictive accuracy and reduces false alarms. However, challenges remain, including sensor placement and calibration, long-term comfort, battery life, and privacy constraints, which can limit real-world deployment.

The next generation of preventive systems is expected to feature adaptive, self-learning models embedded in lightweight, ergonomic wearables capable of continuous monitoring and personalized feedback in daily life. These technologies will not only detect fall risk but may actively prevent falls through early interventions, feedback, or actuation (such as exoskeletal support or pre-impact airbag deployment).

Table 9.1 Sensor modalities and their relevance to fall risk factors.

Sensor Modality	**Measured Parameters**	**Associated Fall Risk Factors**
ECG / PPG	Heart rate (HR), heart rate variability (HRV), arrhythmias, blood oxygen (SpO_2)	Cardiovascular dysfunctions (such as hypotension, syncope, arrhythmias); early fainting detection
Respiration Sensor	Respiration rate (RR), breathing irregularities	Cardiorespiratory fatigue, stress, anxiety-related instability
IMU (Accelerometer, Gyroscope)	Linear/angular acceleration, posture, gait parameters	Balance impairment, gait instability, postural sway, freeze of gait
sEMG	Muscle activation, contraction timing, fatigue levels	Lower-limb weakness, sarcopenia, gait disorders, neuromuscular control
EEG	Brain activity, mental workload, drowsiness, cognitive state	Cognitive decline, stress, fatigue, neurological instability, Parkinsonian symptoms
Foot Pressure / Insole Sensors	Plantar pressure distribution, step symmetry, stride analysis	Gait irregularity, imbalance, foot deformities, diabetic neuropathy
SpO_2/Skin Temperature Sensors	Oxygen saturation, thermal stress, peripheral circulation	Fatigue, cardiovascular or metabolic stress
Camera / Depth Sensors	Visual motion capture, body pose estimation, environment perception	Visual imbalance, trip/slip detection, environmental hazard recognition
Environmental Sensors (such as ambient light, floor vibration)	Contextual cues, lighting, humidity, surface vibration	Poor lighting, slippery floors, environmental hazards
Multimodal Fusion Systems	Combination of physiological and motion data	Comprehensive fall risk estimation and personalized prevention modeling

9.2 Data Collection

The complexity of fall risk assessment is widely recognized in recent literature, emphasizing the need for a multifaceted approach and continuous monitoring to accurately predict and

prevent falls in real-life conditions accurately. Traditional fall detection systems primarily rely on kinematic data, such as acceleration or angular velocity; however, these signals alone are insufficient to capture the physiological and psychological precursors of a fall. To move toward accurate fall-risk prediction, it is essential to integrate multimodal physiological data that reflect the subject's overall physical and cognitive state.

In this context, the XoADL dataset [44] was developed to investigate how physiological signals vary across different Activities of Daily Living (ADLs) and to identify latent patterns of fatigue, stress, or instability. The dataset includes 22 ADLs performed by ten subjects, during which multiple biomedical signals, surface electromyography (sEMG), electrocardiography (ECG), respiration bands, and electrodermal activity (EDA), were simultaneously recorded.

Nevertheless, a significant limitation in fall research remains the scarcity of realistic physiological and psychological data recorded immediately before and during a fall. Conducting such experiments in real-world conditions is ethically and practically infeasible, as it is impossible to simulate an unexpected fall while collecting high-fidelity biomedical signals safely. Furthermore, when participants know that a fall will occur, their physiological responses differ substantially from those accompanying spontaneous falls. To overcome these constraints, a novel experimental protocol was conceived to exploit immersive Virtual Reality as a safe yet ecologically valid testing environment.

In collaboration with Dr. Faraci (SUPSI) and the VICARIOS Lab at the Italian Institute of Technology (IIT), a VR-based fall simulation framework was designed to elicit authentic physiological responses under controlled but realistic conditions, resulting in the XoFallVR dataset. The VR environment reproduces three critical construction-site accidents (STFs) identified by the Occupational Safety and Health Administration (OSHA) as the leading causes of injuries and fatalities in the construction sector [298].

The use of virtual reality to collect data on gait, daily activities, and falls is not a new concept. For example, Tian et al. [299] use an Oculus Rift S virtual reality headset, which obscures participants' view of the actual walking path in the laboratory. At a certain point along the path, unbeknownst to the participants, a researcher discreetly pulls the carpet out from under them. This action is performed with varying degrees of intensity, with the intent of causing a fall or stumble, so that the participant either loses their balance and falls or manages to recover and maintain their balance. However, XoFallVR is one of the first attempts to collect biomedical data via a VR platform for fall risk analysis.

Table 9.2 Demographic information of participants in XoADL dataset: age, height, weight, and estimated body mass index (BMI).

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	
Gender	M	M	M	M	M	F	F	F	F	F	M: 5, F: 5
Age [y]	28	30	42	28	25	32	26	26	25	29	29 ± 5
Height [cm]	178	191	178	180	165	168	173	160	165	155	171 ± 10
Weight [kg]	62	67	83	70	64	54	65	68	68	44	65 ± 10
BMI [kg/m^2]	20	18	26	22	24	19	22	27	25	18	22 ± 3

9.3 Datasets

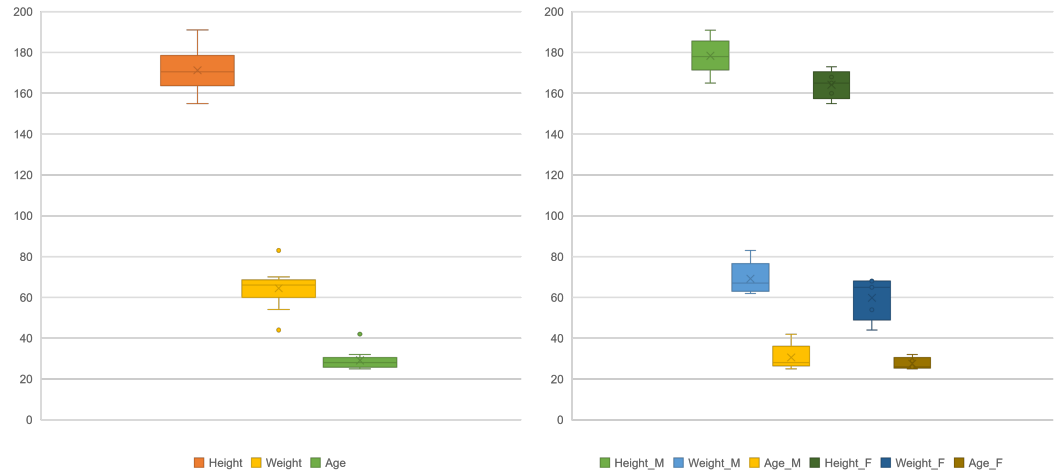
This section provides a detailed description of the participants, sensors, sensor placement, and experimental procedures used to collect the two novel datasets developed for this research: XoADL (Activities of Daily Living) and XoFallVR (Virtual Reality Falls).

9.3.1 XoADL dataset

A series of experiments was conducted with subjects equipped with wearable biosignal sensors to evaluate physiological patterns during Activities of Daily Living (ADLs). The study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of Liguria (protocol reference number: CER Liguria 001/2019) on 28 October 2019. The support of two master's students from the University of Genoa, Natalia Kurvina and Natnael Takele, who were responsible for organizing and managing the experiments during their internship period, was crucial in collecting the data.

Participants

Following the defined study protocol, we specifically analyzed just IIT members. Ten healthy subjects (29 ± 5 years old) with no history of musculoskeletal disorders and who could safely perform the assigned tasks were included. Table 9.2 summarizes the information on the subjects. An effort was made to achieve an equal distribution between males and females. The participants' ages ranged from 25 to 42 years, their heights from 1.55 to 1.91 meters, and their weights from 44 to 83 kilograms, with generally higher values for males than females (see Figure 9.1).



(a) Range of height [cm], weight [kg], and age [y].

(b) Height [cm], weight [kg], and age [y] range depending on gender.

Figure 9.1 Biographical characteristics of subjects in XoADL dataset.

Sensors

Data was collected using the Movella DOT inertial measurement unit (IMU) system (Movella Inc., El Segundo, California, US) and the Biosignalsplux wearable multi-sensor platform (PLUX Wireless Biosignals S.A., Lisbon, Portugal). The Biosignalsplux system included four surface electromyography (sEMG) sensors, one electrocardiography (ECG) sensor, and one piezoelectric respiration (PZT) sensor. This configuration allowed the simultaneous acquisition of kinematic and physiological data, ensuring a comprehensive characterization of movement dynamics and internal physiological responses during the experimental sessions.

The XSens DOT setup comprised five miniature IMU sensors connected via Bluetooth to an Android device for synchronized data acquisition. The sensors were strategically positioned on the lower body to capture lower-limb and pelvic motion, indicated by orange markers in Figure 9.2, as follows:

- one sensor on the pelvic area near the sacrum,
- one on the external lateral side of the left quadriceps,
- one on the external lateral side of the right quadriceps,
- one on the external lateral side of the left ankle, and
- one on the external lateral side of the right ankle.

The Biosignalsplux sensors were arranged to record muscle activity, cardiac activity, and respiratory patterns. Specifically, the four EMG sensors were positioned on the anterior and posterior portions of the quadriceps (blue markers in Figure 9.2), allowing for bilateral measurement of muscle activation and fatigue dynamics. The ECG sensor was placed on the chest near the heart (red marker in Figure 9.2) to acquire single-lead cardiac activity. The PZT respiration sensor was fixed frontally around the thorax, just below the nipple line, using an elastic belt (yellow marker in Figure 9.2), providing real-time detection of respiratory cycles.

Additionally, complementary sensors were included to capture both neurophysiological and psychophysiological signals. An EEG sensor was mounted on the forehead (purple marker in Figure 9.2) to monitor cortical activity. In contrast, an electrodermal activity (EDA) sensor was attached to the index and middle fingers (indicated by the green marker in Figure 9.2) to record skin conductance variations associated with emotional arousal and stress.

Figure 9.2 provides an overview of the complete sensor configuration and body placement. In contrast, Figure 9.3 illustrates the same setup from four different perspectives, highlighting the spatial distribution of the devices along the participant's body.

Experiment procedure

Participants performed activities of daily living (ADLs) and dynamic tasks during the experiments, without experiencing any falls. Table 9.3 summarizes the 22 participants' activities, listed in the order in which they were performed.

The experimental protocol consisted of four calibration phases to measure the maximal voluntary contraction (MVC) of the femoral muscles. For posterior femoral muscles, participants pushed their heels upward against a firm surface (such as the treadmill base) three times. For the anterior femoral muscles, participants sat in a chair and pressed their legs together against resistance provided manually by an experimenter, repeating the exercise three times. Each measurement was performed for both legs. These MVC assessments enabled the tracking of changes in muscle strength throughout the experiment and provided a basis for normalizing electromyographic (EMG) data.

Prior to beginning the trials, participants completed a 5-minute treadmill warm-up at a constant speed of 3.5 km/h to reduce the risk of injury and prevent muscle contractures [300]. The total duration of each participant's session ranged from approximately 1 to 1.5 hours.

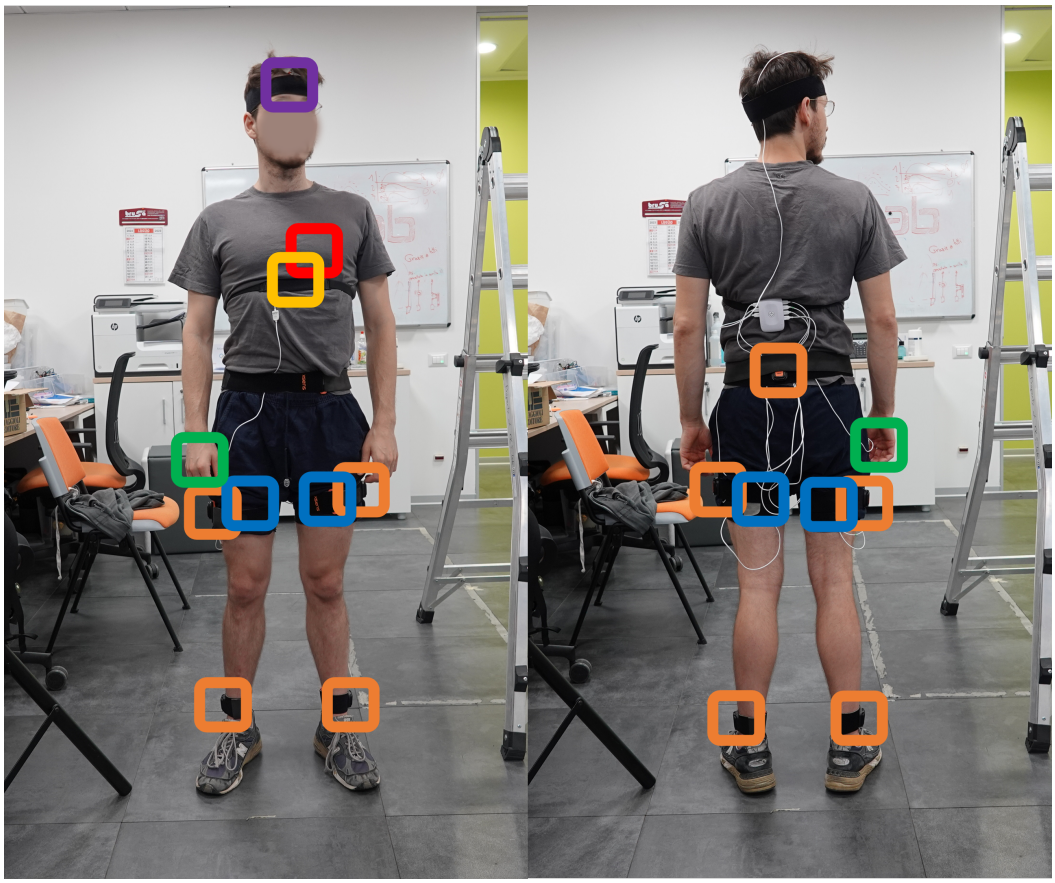


Figure 9.2 Sensor placement in XoADL dataset: five IMUs strategically positioned on the lower body, four EMG sensors on the anterior and posterior part of the quadriceps (blue), the ECG sensor near the heart (red), the PZT sensor was attached frontally with a belt under the nipple line (yellow), the EEG sensor on the forehead (purple), and the EDA sensor on the index and middle finger (green).

Figure 9.3 Sensor placement in XoADL dataset seen from the four points of view: front, back, right side, and left side.

Table 9.3 Activities carried out during the XoADL's data collection.

	Activity	Number of Repetitions
Calibrate	MAX EMG 80%, Standing 10 s	1
Walking	4 km/h for 30s + 10s rest	1
	5 km/h for 30s + 10s rest	1
	Walking around the office (own pace)	1
Jogging	6 km/h 30 sec + 10s rest	1
	7 km/h 30 sec + 10s rest	1
	Jogging around the office (own pace)	1
Running	10 km/h for 30 sec	1
	11 km/h for 30 sec	1
	Running around the office (own pace)	1
Re-calibrate	MAX EMG 80%, Standing 10 s	1
Stairs	$\sim$2 stairs fast + 10s rest	2
	$\sim$2 stairs+ 10s rest	2
	$\sim$2 stairs double steps + 10s rest	2
Re-calibrate	MAX EMG 80%, Standing 10 s	1
Ladder	Ladder climbing (up and down) 5 steps 3s stop	3
	Ladder climbing (up and down) 5 steps no stop + after the last repetition a jump off	3
Jumping	Vertical hop	4
	Leap (2 tiles)	4
Workouts	Squats	5
	Jump squats	5
	Push-ups	5
	Burpees	5
Re-calibrate	MAX EMG 80%, Standing 10 s	1
Down up	Sitting/getting up from a chair	3
	Sitting/getting up from the floor	3

Table 9.4 Demographic information of participants in XoFallVR dataset: gender, age, height, weight, body mass index, wear glasses, virtual reality experience, mother tongue.

	Gender	Age [y]	Height [cm]	Weight [kg]	BMI [kg/m^2]	Glasses	VR experience	Mother tongue
P1	M	28	190	62	17	Y	7M	Persian
P2	M	30	173	78	26	N	2H	Persian
P3	F	26	175	75	24	Y	0H	Italian
P4	F	31	170	71	25	N	4H	Turkish
P5	M	26	175	62	20	N	1H	Amharic
P6	M	23	180	66	20	Y	0H	French
P7	F	23	172	55	19	Y	3M	Arabic
P8	M	29	165	70	26	N	1H	Konkeni
P9	M	29	187	75	21	N	3Y	Italian
P10	M	27	171	80	27	N	1H	Arabic
P11	M	26	160	59	23	Y	2H	Italian
P12	M	25	178	72	23	Y	20H	Italian
P13	M	29	176	65	21	N	4H	Italian
P14	M	44	167	88	32	Y	1H	Portuguese
	M: 11 F: 3	28 ± 5	174 ± 8	70 ± 9	23 ± 4	Y: 7 N: 7	<24H: 10 >24H: 4	Italian: 5 Others: 9

9.3.2 XoFallVR dataset

This section details the participants, employed sensors, sensor placement, and the experimental procedure of the XoFallVR dataset. All participants provided written informed consent prior to the commencement of the study, in accordance with the principles outlined in the Declaration of Helsinki. The experimental procedure adhered to the IIT protocol ADVR TELE01, which received approval from the Ethics Committee of the Liguria Region in Genoa.

Participants

Due to the defined study protocol, we specifically targeted affiliated members of the IIT who were not fully aware of the study's purpose. Colleagues from the same research group who were aware of the study's purpose were explicitly excluded to prevent bias due to preconceived expectations. Similarly, participants were from multiple groups within the institution and were instructed not to disclose the experiment's content during the week of data collection. Fourteen healthy subjects (28.29 ± 5.14 years old) with no history of musculoskeletal disorders were included and were able to perform the assigned tasks safely. Among the participants, half wore glasses when looking at the screen; none had hearing problems. Only 28.57% of the participants had previous experience with VR systems (more than 24 hours). Eleven subjects were male, and the other three were female. Table 9.4 summarizes information about the subjects.

Sensors

For data collection, we used a wrist-worn device (Empatica E4) and a chest-worn device (Polar H10), respectively, along with the Vive Pro 2 visor for virtual reality. Empatica E4 is a wristband worn on the left hand with photoplethysmography (PPT), electrodermal activity (EDA), accelerometer, and skin temperature sensors. The software automatically generates the following signals: blood volume pulse (BVP) at 64 Hz, interbeat interval (IBI), heart rate (HR) at 1 Hz, electrodermal activity at 4 Hz, raw XYZ acceleration at 32 Hz, and skin temperature at 4 Hz. Polar H10 is a high-precision heart rate sensor with Bluetooth and ANT+ connectivity designed for reliable performance in both fitness and clinical settings. It features a chest strap with a built-in memory module for data storage, water resistance to 30 meters, and a long-lasting battery. Known for its accuracy and stability, the Polar H10 is also widely used for research. The VIVE Pro 2 is a high-end VR headset designed for immersive experiences with exceptional visual clarity and a wide 120° field of view. It features a 5K resolution display (2448 x 2448 pixels per eye) and an ultra-smooth 120Hz refresh rate. Empatica E4 recorded data were stored locally, while data from Polar H10 were saved using the Android ECG Logger application [301] on the Android device. Then both were transferred to a computer for further processing after the experiment. In particular, the app can also calculate and provide the breathing rate (BR) through the ECG-Derived Respiration (EDR) technique. Pupillometry data from the VR headset were saved using Unreal Engine (UE) at a rate of 50 Hz. Table 9.5 summarizes the raw signals used by each sensor and their frequency, while an example of the sensor placement on the manikin is reported in Figure 9.4.

Experiment procedure

The study aimed to elicit and analyze three different levels of fall risk (low, medium, high) in participants using virtual reality. The different parts of the experiment procedure according to [302] are described in detail below:

- Preparation: Subjects were enrolled by asking them to express their willingness to participate in a VR study to collect physiological parameters during an assembly task and some simulated incidents that may occur in a work environment. The experiments all took place during a work week, considering a maximum of four daily time slots, excluding breakfast and lunch breaks. This prevented subjects from consuming caffeine and tobacco, although no specific prohibitions were in place against these substances. Prior to the study, the participants read and signed a consent form. Upon arrival at the

Table 9.5 Raw signals collected by each sensor and their frequency in the XoFallVR dataset.

Sensor	Signal	Frequency [Hz]
Empatica E4	Acceleration	32
	Heart Rate (HR)	1
	Blood Volume Pulse (BVP)	64
	Electrodermal Activity (EDA)	4
	Temperature	4
Polar H10	Acceleration	200
	Heart Rate (HR)	1
	RR intervals	-
	Electrocardiogram (ECG)	130
	Breathing Rate (BR)	1
Vive Pro 2	Gaze direction	50
	Pupil diameter	50
	Eye Oppenness	50

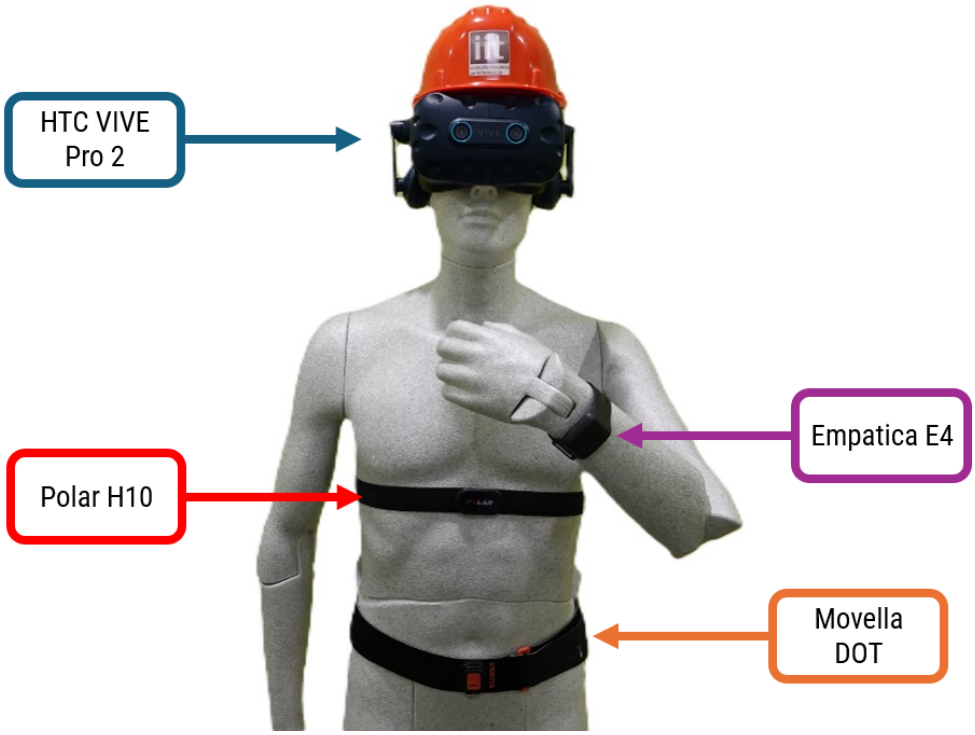


Figure 9.4 Sensor placement on the manikin of the sensors in the XoFallVR dataset.

study location, the participants were equipped with the sensors, and a short sensor test was conducted. Empatica E4 was used to mark the experiment phases via a tap gesture.

- Baseline: After the participants had been equipped with the sensors (except the VR headset), a 1-minute baseline was recorded while they were standing. This condition aimed to collect data during a neutral affective state, without engaging in activities or experiencing VR stress.

- Training: Next, participants were taken to a Cyberith Virtualizer Elite 2 platform and wore VR headsets. They learned to use the motion controllers in an ad-hoc training environment, watched a video displayed on a screen, and practiced using the controls to move objects required for the assembly task. The primary goal is to familiarize the participant with the VR environment and provide them with a general understanding of the task.

- Coffee break: Then, participants enter a randomly chosen virtual environment in which one of three different accident hazards is present: Slip, Trip, and Fall. To collect data on a low fall risk, participants were asked to move around the ground floor of the VR environment, head to the Coffee Break area, and make themselves a coffee. To do this, they must press a button and wait one minute before the coffee is ready and can be drunk. While waiting, the participant can watch a video on a screen. The goal is to make the participant feel comfortable and to encourage them to stay in this condition for at least one minute. Additionally, this sequence is repeated for each repetition and is useful between incidents to help relax biomedical signals.

- Activity: Following, participants were asked to ride an elevator to the top floor of a building, where they had to assemble some scattered railings. They had to take five elements from a container near the assembly station to assemble the railing and drop them near the railing skeleton. The order in which these elements were placed was irrelevant. After placing a random number of elements at the second assembly station, the following incident occurred, depending on the scenario chosen at the beginning, which concluded the scenario.

- Slip: After assembling the first parapet, participants could choose to go and assemble a parapet on one of the two bifurcated platforms. In this way, the participant had no idea that the scenario ended after the next parapet was assembled, but that at least a third guardrail had to be assembled. This was to reduce the expectation of the accident. Once the participant placed the element that triggered the accident animation, the wooden boards on which they walked would tilt, and the participant would slowly slip.

When creating the scenario, care was taken to ensure that the participants had both the perception that the plank was tilting and that they were sliding.

- Trip: After assembling the first parapet, participants had to reach an area where the parapet was to be assembled, and a container with the necessary objects was placed in front of them. In the center was an iron pipe perpendicular to the platforms on which subjects could walk until the animation was triggered, and that object became an obstacle that caused participants to trip. This could happen randomly in either direction, when the participant carried the object to be assembled or when they returned to pick up a new object.

- Fall: After assembling the first parapet, participants see two more parapets to assemble, one closer and one farther away. All subjects decide to continue the activity by assembling the one closer, and when the animation begins, the wooden beam under their feet starts to shatter. After a few seconds, the participants fall in free fall. The time interval between the beam shattering and the moment of the fall was made long enough to allow participants to perceive the incident.

- Surveys: At the end of each scenario, the task completion time (TCT) was recorded, and participants were asked to remove the headset and motion controllers to respond to the following surveys: IMI [160] and SUM [184]. Finally, at the end of the experiment, after participants had played all three scenarios, they were asked to complete the following surveys: IPQ, SUS, UMUX, UFS, and VRSUQ.

The study had a total duration of approximately 1 hour and 15 minutes. Figure 9.5 summarizes the experiment procedure.

9.4 Methodology

The analysis of the fall prediction problem, focused on estimating the risk of falling, required distinct methodological pipelines for the two new datasets: the XoADL (Activities of Daily Living) dataset, analyzed for general physiological models, and the XoFallVR (Virtual Reality Falls) dataset, subjected to a comprehensive machine learning classification framework.

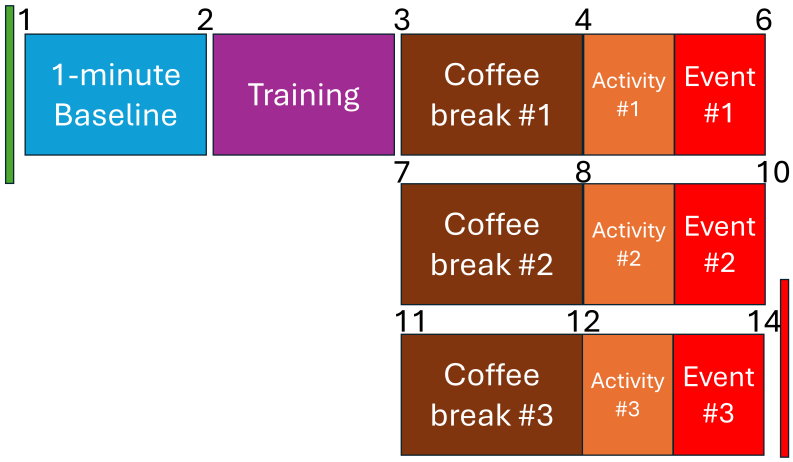


Figure 9.5 Experiment procedure in the XoFallVR dataset.

9.4.1 XoADL dataset

So far, only some of the sensors listed in Section 9.3.1 have been analyzed. Muscle activation (MA), heart rate (HR), and breathing rate (BR) were derived from biomedical signals to assess how these physiological indices varied across different activities.

Electromyographic (EMG) signals were preprocessed using the procedure described in [303], which included rectification and low-pass filtering. The signals were then normalized to the participants' Maximum Voluntary Contraction (MVC) values to allow individual comparability. Muscle activation was further classified relative to the 95th percentile recorded during the preceding calibration phase. After normalization, the median value of each signal was calculated to evaluate overall trends, and these values were subsequently compared for each muscle monitored across the various activities.

Using a frequency-domain approach, heart rate (HR) was estimated from ECG signals. The ECG signals were first filtered using a band-pass filter with cutoff frequencies of 0.9 Hz (corresponding to 55 beats per minute, bpm) and 3.7 Hz (corresponding to 220 bpm). A Fast Fourier Transform (FFT) was then applied to the filtered signals, and the HR during each activity was defined as the sinusoidal frequency with the highest amplitude. The chosen cutoff frequencies were informed by literature values: normal resting HR ranges from 50–90 bpm [304], while the maximum HR is approximately estimated as 220 minus the participant's age [305, 306].

Breathing rate (BR) was derived similarly from signals acquired with a piezoelectric respiratory band. Signals were filtered with a band-pass filter ranging from 0.2 Hz (12 breaths per minute, Bpm) to 1 Hz (60 Bpm), reflecting normal adult resting respiration

(12–20 breaths per minute) [307] and the increased rates expected during exercise (up to 40–60 breaths per minute) [308].

9.4.2 XoFallVR dataset

The analysis and evaluation follow the classic data processing chain, which consists of the following stages: synchronization, feature extraction, feature selection, and classification.

Synchronization

The synchronization of sensor data is a fundamental issue in Internet of Things (IoT) environments. In general, when a measurement environment includes several independent devices, it is essential to ensure the overall consistency of the data with respect to a reference timestamp [309]. In the case study, the problem is even more complex because, in addition to involving multiple sensors, it also consists of acquiring signals at different frequencies, even within the same device.

Instead of training a prediction system based on raw data, as in previous models, given the greater complexity of the study, a feature-based system was developed. This choice simplifies data analysis and the development of the target system. In fact, instead of having all signals at the same frequency, for example, as a result of data interpolation, the features are extracted by considering a time window, as illustrated in the Section 9.4.2 (Feature extraction). To make the signals suitable for feature extraction, it is therefore essential to construct the time axis of each signal so that each data point is placed in the correct time window.

During data collection, the Empatica E4 was used as the primary synchronization device, with the button pressed appropriately and the various stages of the experiment's beginning and end outlined, as shown in Figure 9.5. See the numbers above the blocks. The time is provided in a Unix timestamp relative to UTC. The ECG Logger for Polar H10 provides data based on the time of the Android device used to capture the data. Therefore, the data had to be converted to a Unix timestamp relative to UTC for synchronization. In the case of the Vive Pro 2, synchronization required additional effort because the recording was made using Real Engine, and the timestamp was relative and not absolute with respect to the start of acquisition (sample 1, 2, 3, etc.). Only the absolute time of file saving was known. The data time axis was therefore constructed backwards, taking this information and sampling frequency into account.

Additionally, at the beginning of the animation that preceded the fall, there is an indicator in the data that was useful for synchronizing the accident with other signals. Various problems

arose during the experiments, including a lack of connection between the acquisition devices and PCs, tablets, and smartphones, the need for repeated experiments, and errors in loading the virtual reality environment. To synchronize the data and minimize errors, the video recording of the experience on Real Engine and the notes taken during data collection were also considered.

Feature extraction

All features were calculated within a one-minute window, with the window shifted by 250 ms. These parameters were chosen according to [302, 310]. Some attempts were also made to consider more restrictive time windows (1, 3, 5, 10 seconds) to adapt to the problem under study; however, given the similar results, the suggestion provided in the specific literature on wearables was ultimately chosen. Table 9.6 shows the features extracted from the different modes.

Several features were calculated from the raw ACC signal, including the mean value, standard deviation, median, absolute integral, and peak frequency. These features were calculated separately for each axis and expressed as absolute magnitudes, then summed across all axes (3D). In total, 20 features were extracted from the Empatica E4 and Polar H10 acceleration signals for each sensor.

From the breathing rate collected by the Polar H10, the mean value, standard deviation, median, and slope were obtained. From the heart rates, the mean value, standard deviation, and median were obtained. In total, three features for each sensor.

On the raw BVP/ECG signal, the RR intervals were determined using peak detection algorithms, a crucial metric for assessing heart rate variability (HRV). Besides, Polar H10 already provides an RR (filtered) interval value. Starting from them, multiple features are calculated. In addition to the mean value, standard deviation, and median, other time-domain features are the number/percentage of successive RR intervals that differ by more than 50 ms (NN50, pNN50), root mean square (RMS), root mean square of successive RR intervals (RMSSD), triangular interpolation (TINN), and Baevsky stress index (SI, SIdB). The sum of components and power in different frequency bands was calculated in the frequency domain. The frequency bands used were ultra-low (ULF: 0.01-0.04 Hz), low (LF: 0.04-0.15 Hz), high (HF: 0.15-0.4 Hz), and ultra-high (UHF: 0.4-1.0 Hz). In [302, 311, 312], HR and HRV are described in detail.

Since electrodermal activity (EDA) is controlled by the sympathetic nervous system (SNS), it is susceptible to high levels of arousal. First, statistical characteristics (mean value, standard deviation, median, maximum, minimum, and dynamic range) were calculated.

Furthermore, since the raw EDA signal consists of a tonic component, skin conductance level (SCL), and a phasic component, skin conductance response (SCR), the components were separated using a convex optimization approach called cvxEDA [8]. The SCL represents a slowly varying baseline conductance. At the same time, the SCR is a short-term response to a stimulus [313]. After separating SCL and SCR, additional features were calculated, such as the statistical features mentioned above, the correlation between SCL and time, the number of startles (#SCR), and the sum of duration, prominence, and area of startles in SCR [314].

The standard statistical features mentioned above, as well as the slope, were calculated using the raw TEMP signal. The angular motion between two consecutive samples was calculated from the gaze directions. From the pupil diameters, the time difference in each eye, pupil dilation or constriction (mydriasis), and the difference between eyes, indicators of anisocoria, were calculated. Statistical features, including the mean value, standard deviation, and median, were generated for all raw and derived signals. Finally, the blinks are defined as the number of times the openness is maintained below 10%, and the asymmetry is the difference between the eyes' open positions [315].

Table 9.6 List of extracted features on XoFallVR dataset.

	# Features	Acronym	Description
ACC	4	AA	Mean value for each axis separately, and the magnitude
	4	ASD	Standard deviation for each axis separately, and the magnitude
	4	MA	Median for each axis separately, and the magnitude
	4	AAI	Absolute integral for each axis separately, and the magnitude
	4	APF	Peak frequency for each axis separately, and the magnitude
BR	1	ABR	Mean value of the breathing rate
	1	SBR	Standard deviation of breathing rate
	1	MBR	Median of the breathing rate
	1	SBR	Slope of the breathing rate
HR	1	AHR	Mean value of the heart rate
	1	SHR	Standard deviation of heart rate
	1	MHR	Median of the heart rate
RR	1	ARR	Mean value of RR intervals
	1	SDNN	Standard deviation of RR intervals
	1	MRR	Median of RR intervals
	2	NN50, pNN50	# and the percentage differing by more than 50 ms
	1	RMS	Root mean square of RR interval differences
	1	RMSSD	Root mean square of successive RR interval differences
	1	TINN	Triangular interpolation of the RR intervals
	2	SI, SIdB	Baevsky stress index in linear and dB
	4	PULF, PLF, PHF, PUHF	Power in ultra-low, low, high, and ultra-high bands
	1	LFHF	Ratio of LF and HF band powers of the RR intervals
	2	NPLF, NPHF	Normalized LF and HF band powers of the RR intervals
	4	SULF, SLF, SHF, SUHF	Sum of ULF-UHF frequency components of RR intervals
EDA	3	AEDA, ASCL, ASCR	Mean value of EDA, SCL, SCR
	3	SEDA, SSCL, SSCR	Standard deviation of EDA, SCL, SCR
	3	MEDA, MSCL, MSCR	Median of EDA, SCL, SCR

Table 9.6 List of extracted features on XoFallVR dataset.

	# Features	Acronym	Description
	3	MxEDA, MxSCL, MxSCR	Maximum of EDA, SCL, SCR
	3	MnEDA, MnSCL, MnSCR	Minimum of EDA, SCL, SCR
	3	REDA, RSCL, RSCR	Dynamic range of EDA, SCL, SCR
	1	δEDA	Slope of EDA
	1	CSCL	Time correlation of SCL
	1	#SCR	Number of SCR startles
	1	DSCR	Sum of SCR startle duration
	1	PSCR	Sum of SCR startle prominence
	1	ASCR	Sum of SCR startle area
Temp	1	AT	Mean value of temperature
	1	ST	Standard deviation of temperature
	1	MT	Median of temperature
	1	MxT	Maximum of temperature
	1	MnT	Minimum of temperature
	1	RT	Dynamic range of temperature
	1	δT	Slope of temperature
Eyes	3	AGD	Mean value of gaze direction for each axis separately
	3	SGD	Standard deviation of gaze direction for each axis separately
	3	MGD	Median of gaze direction for each axis separately
	1	AAM	Mean value of angular gaze movement
	1	SAM	Standard deviation of angular gaze movement
	1	MAM	Median of angular gaze movement
	2	ALD, ARD	Mean value of left/right pupil diameter
	2	SLD, SRD	Standard deviation of left/right pupil diameter
	2	MLD, MRD	Median of left/right pupil diameter
	2	ALM, ARM	Mean value of left/right mydriasis (pupil dilation)
	2	SLM, SRM	Standard deviation of left/right mydriasis
	2	MLM, MRM	Median of left/right mydriasis
	1	AIA	Mean value of anisocoria indicator
	1	AIS	Standard deviation of anisocoria indicator
	1	AIM	Median of anisocoria indicator
	2	ALO, ARO	Mean value of left/right eye openness
	2	SLO, SRO	Standard deviation of left/right eye openness
	2	MLO, MRO	Median of left/right eye openness
	2	LB, RB	Number of blinks in the left or right eye
	1	EA	Eye asymmetry
Tot.	115		

Feature selection

To restrict the number of features, a fully data-driven pipeline is adopted. First, constant and quasi-zero variance features were removed. Next, a feature reduction was made using two well-known filter-type feature selection algorithms. Specifically, the χ^2 test and the Minimum Redundancy Maximum Relevance (mRMR) algorithm [277]. The scores were calculated for each feature selection method, and only features above the 75th percentile in both methods were taken.

In the next stage, highly correlated features with an absolute value greater than 0.8 were removed, except the feature with the highest variance. Finally, if at least three features remained, the correlation of the remaining features was calculated to remove possible outliers. Empirically, features with too low median correlation with the other features (< 0.01) were removed. Since only a few features were extracted for breathing rate and temperature, filter-type feature selection was not applied. Pseudo-code 1 summarizes the pipeline.

Algorithm 1 Feature Selection

{Remove features with quasi-zero variance}
for $feat$ **in** $Features$ **do**
  if var($feat$) = 0 **then**
    remove $feat$ from $Features$
  end if
end for
{Filter-type feature selection}
for $feat$ **in** $Features$ **do**
  $chi2\_score[feat] \leftarrow$ chi2test($feat$, $label$)
end for
$mRMR\_score \leftarrow$ mRMR($Features$, $label$)
for $feat$ **in** $Features$ **do**
  if $chi2\_score[feat] > Th_{\chi^2}$ **and** $mRMR\_score[feat] > Th_{mRMR}$ **then**
    $newFeatures \leftarrow feat$
  end if
end for
{Correlation analysis}
$corrF \leftarrow$ abs(corr($Features$))
$varF \leftarrow$ var($Features$)
for $feat$ **in** $Features$ **do**
  $corrFeatures \leftarrow Features : corrF[feat, Features] > Th_{corr}$
  $best\_feat \leftarrow corrFeatures : varF[corrFeatures] = \max(varF[corrFeatures])$
  remove $corrFeatures$ from $Features$ except $best\_feat$
end for
{Remove possible outliers}
if length($Features$) > 3 **then**
  for $feat$ **in** $Features$ **do**
    if median($corrF[feat]$) $< Th_{median}$ **then**
      remove $feat$ from $Features$
    end if
  end for
end if

This study employed filter-type feature selection methods instead of iterative feature selection methods because we were interested in comparing models trained with the same set of features. Filter-type feature selection is not related to the training algorithm. It measures the importance of features based on their characteristics, such as the variance of features and their relevance to the response. In contrast, iterative feature selection methods add or remove features based on variations in model performance. The χ^2 test examines whether each predictor variable is independent of the response variable. A small p-value for the test statistic indicates that the corresponding predictor variable is significantly dependent on the response variable and is therefore an important characteristic. The output score is $-log(p)$. Therefore, a high score value indicates that the corresponding predictor is important. The mRMR algorithm identifies a set of mutually dissimilar and maximally distinct features, allowing them to represent the response variable effectively. The algorithm seeks to maximize the relevance of the features to the response variable while minimizing redundancy within the feature set. The method utilizes the mutual information of variables, specifically the pairwise mutual information of features and the mutual information between a feature and the response.

Classification models

The selected features, described above, serve as input for the classification phase. Six machine learning algorithms were applied and compared in our benchmark: decision tree (DT), discriminant analysis (DA), naive Bayes (NB), k-Nearest Neighbor (kNN), classification tree ensemble (Ens), and multilayer perceptron (MLP). All models were evaluated using the leave-one-subject-out cross-validation (LOSO-CV) procedure and a stratified k-fold cross-validation procedure with $k = 10$. The LOSO results indicate how a model would generalize and behave on data from a previously unseen subject. In contrast, k-fold results describe how a model performs, considering previous experience on the subject. Two sequences (considering two scenarios: Trip and Fall, Slip and Fall, or Slip and Trip) are used to train the models, and one scenario is used for testing (Slip, Trip, or Fall).

To reduce inter-subject differences, each subject's data were normalized by subtracting the median of the baseline data. Before training, validation, and testing, the data were also standardized against the training-validation dataset, and the models were optimized using Bayesian optimization for 10 iterations. Results obtained with a random guesser (RG) and a majority guesser (MG) are also reported. The random guesser is defined as one who chooses one of the possible classes at random (with the same probability). In contrast, the majority guesser always chooses the majority class.

Balancing

Three fall risk levels (low, medium, and high) can be distinguished based on the parts of the experiment procedure. Specifically, low risk is present during the coffee break, medium risk during the activity, and high risk when the accident starts. Two classification tasks are identified from this subdivision. First, a three-class problem was defined: low, medium, and high. Second, a binary classification task was defined to classify low and high values, reducing the problem's complexity. In both cases, however, the dataset is highly unbalanced; in fact, the duration of low fall risk is just over a minute, medium fall risk is a few minutes, and high fall risk is a few seconds. This implies that the dataset follows approximately the following proportions: 32.5% low fall risk, 60.6% medium fall risk, and 6.9% high fall risk. Three techniques were tried to balance the dataset: the synthetic minority over-sampling technique (SMOTE) [240], the adaptive synthetic sampling approach for imbalanced learning (ADASYN) [241], and class weighting. The first two techniques are applicable only to binary classification problems, and in the case of our dataset, they yielded lower performance than the inverse-squared frequency class weighting. During model training, each sample is weighted differently according to the class to which it belongs and the class in which it is classified. A generic element $w_{i,j}$ in the cost matrix is equal to $\sqrt{S_j/S_i}$, where S_j defines the number of samples of the predicted class and S_i defines the number of samples of the true class. Instead, the elements on the diagonal are 0.

Evaluation Metric

Balanced accuracy and F1-score were considered as evaluation parameters. Balanced accuracy is defined as the average recall obtained on each class. The F1-score can be interpreted as a harmonic mean of accuracy, indicating the reliability of the results in a specific class, and recall, which measures the completeness of the results. To obtain the final F1-score, precision and recall were calculated separately for each class and then averaged. The application of the F1-score is recommended for unbalanced classification tasks, as in this case, since the various conditions were performed at different lengths during the experiment procedure [302].

9.5 Results

This chapter section presents the empirical results of the fall risk estimation work, divided into analyses of physiological patterns during Activities of Daily Living (ADLs), the assessment

of the Virtual Reality (VR) environment, and the performance of Machine Learning models for fall-risk classification.

9.5.1 Activities of Daily Living

This section synthesizes the findings from analyzing various biosignal indices collected from 10 subjects performing a series of Activities of Daily Living (ADL) and dynamic activities. The study's primary goal was to investigate the use of biological signals as potential predictors of falling, contributing a novel dataset to the literature.

Muscle Activation

The analysis of muscle activation patterns, quantified via Electromyography (EMG) indices, demonstrated a clear relationship between movement speed and muscle exertion. The results, which are visually summarized in Figure 9.6, show that across all subjects, the median muscle activation for the four major muscles analyzed increased with increasing speed during locomotor activities (walking, jogging, and running), following a near-linear profile. Specifically, walking at 5 km/h resulted in higher average activation than 4 km/h. A similar trend was noted for jogging (7 km/h vs. 8 km/h) and running, with overall muscle activation for running being significantly greater than for walking and jogging. Regarding activity-specific patterns, in explosive activities such as vertical jumping and leaping, the quadriceps muscles showed predictably higher average activation than the front leg muscles. Conversely, a lower median and variance in muscle activation were observed among subjects in ladder activities compared to more explosive tasks, such as jumping or jump squats.

Heart and Breathing Rate responses

The physiological responses measured by Heart Rate (HR) and Breathing Rate (BR) confirmed the differentiation between activities of varying intensity.

As illustrated in Figure 9.7, HR demonstrated an increase proportional to the imposed speed during locomotion, mirroring the EMG findings. The highest HR and significant variability were observed during running, jumping/hopping, and the first recorded activity (walking at 4 km/h). The lowest HR was recorded during walking, going up and down stairs, and straightforward stair climbing, possibly reflecting the high frequency and familiarity of these activities in the study cohort.

The calculated BRs for all activities are reported in Figure 9.8, showing that the highest rate was found during highly strenuous or compound activities, including jumps, jump squats,

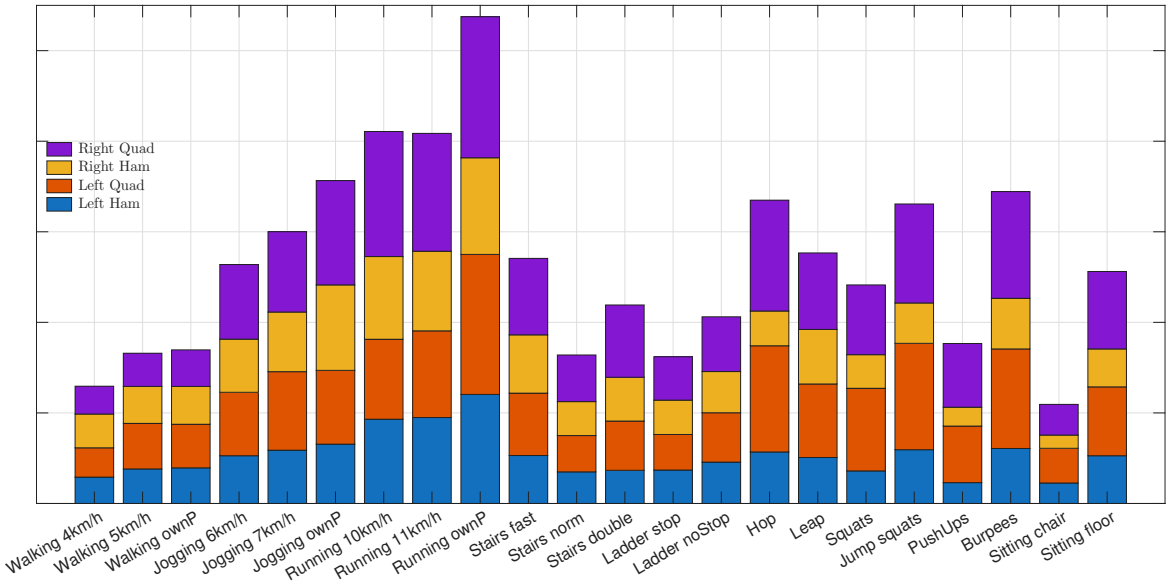


Figure 9.6 EMG indexes for activities.

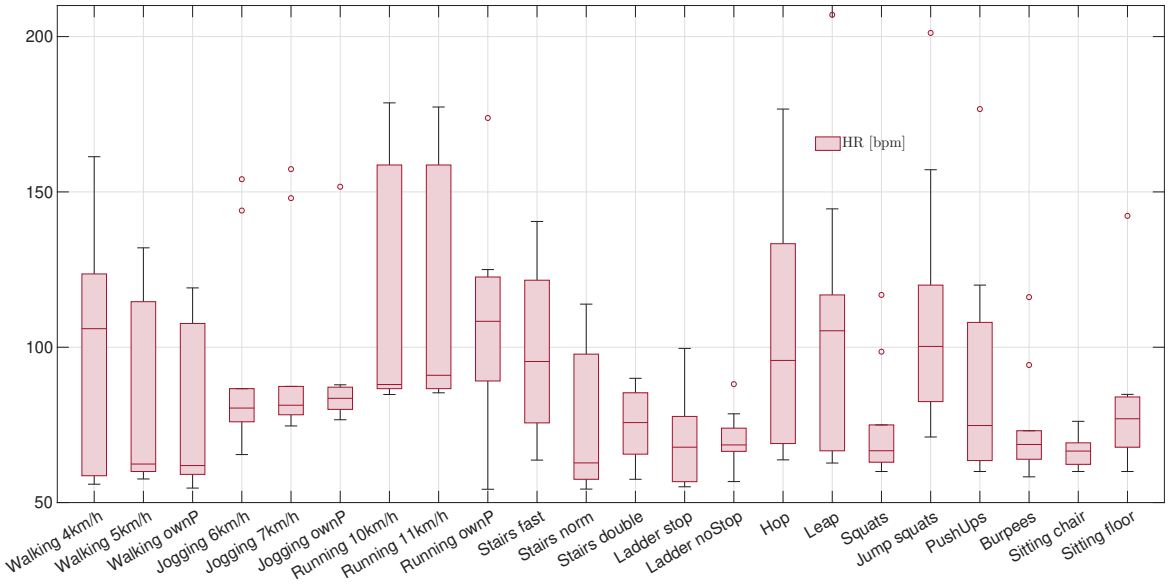


Figure 9.7 Heart Rate (HR) for activities [bpm].

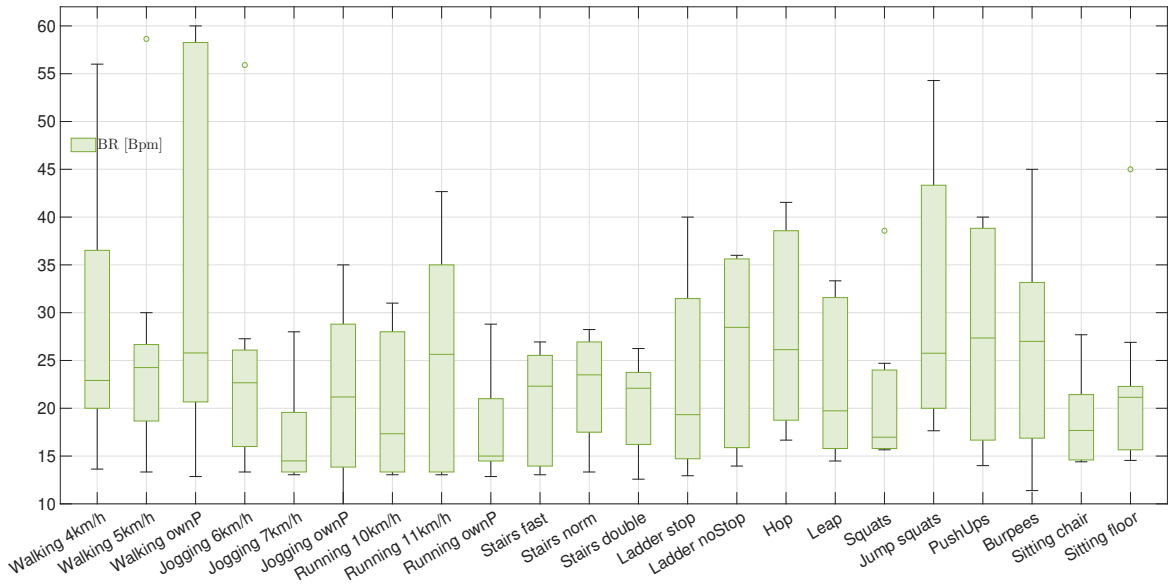
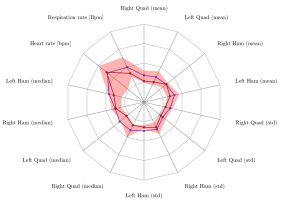


Figure 9.8 Breathing Rate (BR) for activities [Bpm].

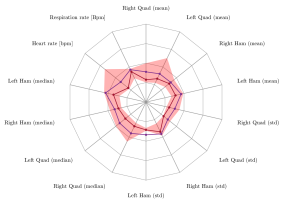
and push-ups. The lowest BR was associated with activities such as jumping, squatting, and sitting. An anomalous result was observed when walking at one's own pace, indicating unexpectedly high variability between subjects. This variability may be attributed to the subjects' natural walking cadence potentially falling within the low-frequency band-pass filter range ($0.2 - 1$ Hz), thereby influencing the BR measurement.

Joint Biosignal Patterns

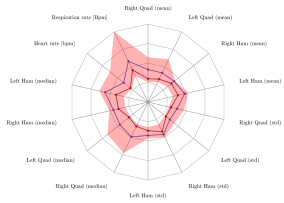
A joint analysis of all measured indices and metrics is represented as a Kiviat diagram (Figure 9.9), which revealed distinctive and characteristic physiological patterns for various activities, enabling the discrimination between heavy/strenuous and less strenuous activities. Different activity types exhibited unique patterns, while similar activities showed similar patterns. For example, walking activities shared similar patterns characterized by a relatively high heart rate. Jogging activities were primarily identified by high muscle activation. Running activities on a treadmill were characterized by high activation of the right leg muscles, a high heart rate, and a low respiratory rate. Additionally, stair activities (going up and down, and straightforward climbing) presented similar patterns, which remained consistent across different repetitions.



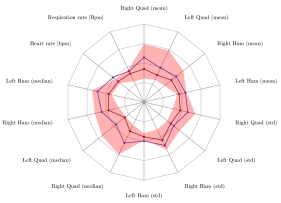
(a) Walking 4km/h



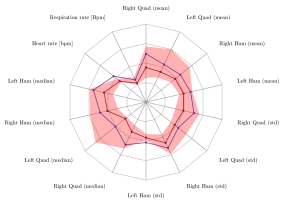
(b) Walking 5km/h



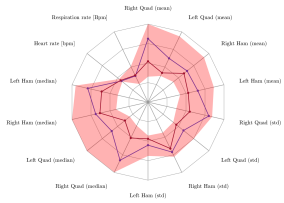
(c) Walking own pace



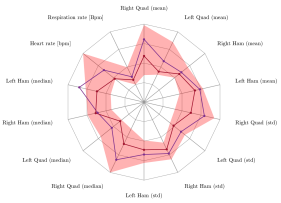
(d) Jogging 6km/h



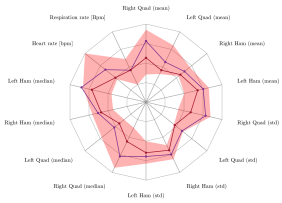
(e) Jogging 7km/h



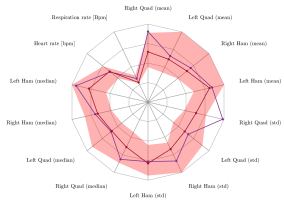
(f) Jogging own pace



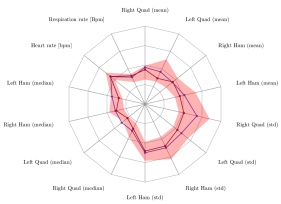
(g) Running 10km/h



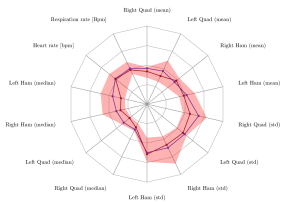
(h) Running 11km/h



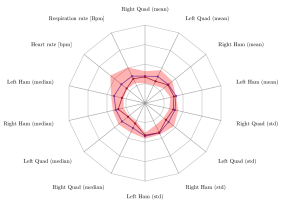
(i) Running own pace



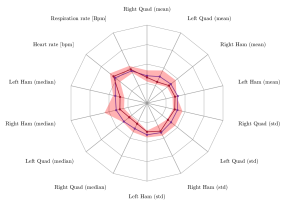
(j) Stairs fast #1



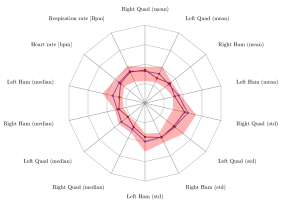
(k) Stairs fast #2



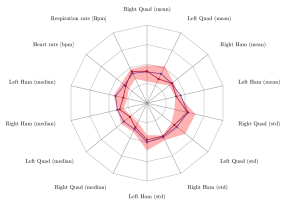
(l) Stairs norm #1



(m) Stairs norm #2

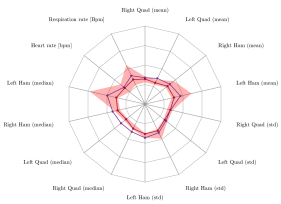


(n) Stairs double #1

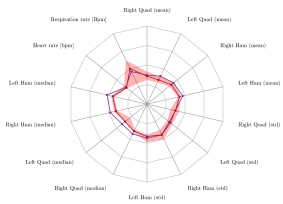


(o) Stairs double #2

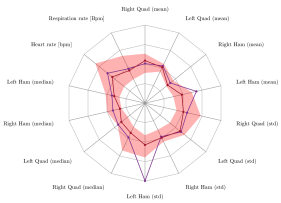
Figure 9.9 Kiviat diagrams of activities. Axes: metrics, Purple line: mean, Red line: median.



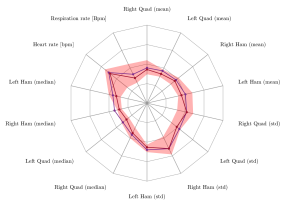
(p) Ladder stop



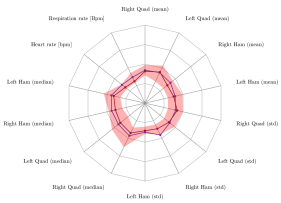
(q) Ladder noStop



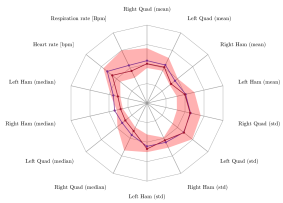
(r) Hop



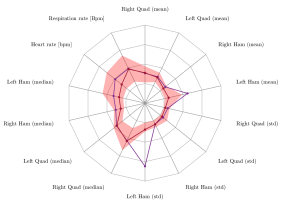
(s) Leap



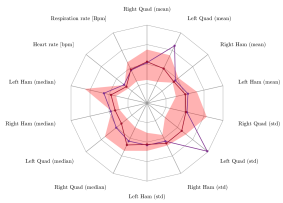
(t) Squats



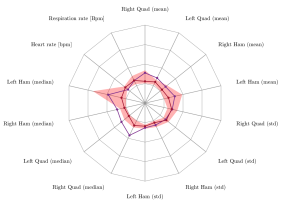
(u) Jump squats



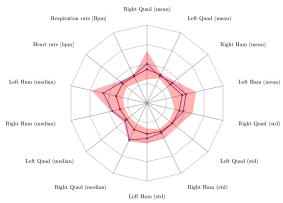
(v) PushUps



(w) Burpees



(x) Sitting chair



(y) Sitting Floor

Figure 9.9 Kiviat diagrams of activities. Axes: metrics, Purple line: mean, Red line: median, Shaded region: Range between the lower quartile (25th percentiles) to the upper quartile (75th percentiles).

9.5.2 VR-based environmental assessment study

Before training the machine learning models for fall-risk estimation, a series of validated questionnaires was administered to assess the virtual reality (VR) environment in terms of presence, immersion, intrinsic motivation, user experience, usability, and motion sickness.

Although previous research has explored VR systems for safety and training in high-risk occupations, most studies have focused primarily on training efficacy or user acceptance, rather than predicting fall risk. For example, studying the impact of VR systems on workers at height (WAH) training employees [316], or investigating the impact of VR-based postural training and balance rehabilitation on postural stability and proprioception [317, 318]. At the same time, others analyzed mental fatigue induced by immersive environments [293] or proposed frameworks for safety training simulations [5], [319, 320]. Moreover, studies such as [321] focused on the usability assessment of VR training modules for construction workers. However, this work represents a novel contribution, as it is the first to evaluate the impact of motivational, perceptual, and usability factors within a VR-based experimental environment designed explicitly for fall-risk estimation in workers at height.

As with the other data collected during the doctoral period, not all data has been analyzed yet. The current outcomes are organized into two main sections: Section 9.5.2 (Post-Activity), referring to data collected immediately after each task, and Section 9.5.2 (Post-Experiment), obtained after completing all sessions. A summary of system performance and error reporting is also presented in Section 9.5.2 (System Performance Analysis).

Post-Activity

Following each virtual assembly activity, participants were invited to complete post-task surveys to assess their intrinsic motivation, perceived usability, and task efficiency. These metrics provided valuable insights into how users interacted with the VR environment in real time, revealing changes in effort, satisfaction, and confidence as they became familiar with the system.

- Intrinsic Motivation Inventory (IMI): Figure 9.10 depicts the IMI results obtained after each assembly task. The scale's internal consistency was measured using Cronbach's alpha, which was $\alpha = 0.777$. Task 3 obtained the highest scores in the effort/importance (E/I) dimension with a mean value of 5.464 ± 1.262, in interest/enjoyment (I/E) with a mean value of 6.142 ± 0.101, and perceived competence (PCo) with a mean value of 5.392 ± 0.151. Task 1 reported the highest perceived choice (P-Ch) score with a mean value of 5.821 ± 0.959. Participants reported the highest pressure/tension (P/T) rating

in Task 2, with a mean value of 4.107 ± 0.353. Finally, the value/usefulness (V/U) dimension obtained the highest score in Task 1, with a mean value of 6.285 ± 0.

According to the effort/importance dimension, participants progressively came to value the importance of assembly activities. Similarly, the enjoyment/interest dimension showed an increasing slope across the three tasks, indicating that participants enjoyed performing the tasks. Perceived choice dimension results define engagement with the task, and participants' scores suggest that they felt slightly forced to complete the activities. The perceived competence dimension results were generally the second lowest, which suggests that participants did not feel confident enough to complete the tasks. In the pressure/tension dimension, the results were the lowest among all dimensions, indicating that the participants did not experience excessive tension during the assembly activities. Finally, the value/usefulness dimension was the highest rated among the three tasks, indicating that participants perceived the activities as meaningful.

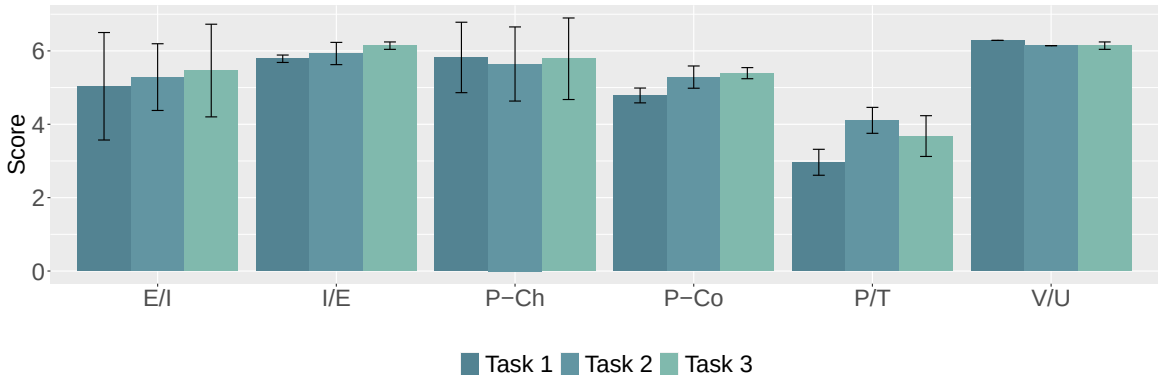


Figure 9.10 Intrinsic Motivation Inventory results. E/I: Effort/Importance, I/E: Interest/Enjoyment, P-Ch: Perceived choice, PC: Perceived competence, P/T: Pressure/Tension, V/U: Value/Usefulness.

- Single Usability Metric (SUM): The computed Cronbach's alpha for the dataset was $\alpha = 0.795$, indicating excellent internal reliability. Figure 9.11 shows the SUM results for the three assembly activities. The highest rated activity in the task ease attribute was Task 1 with a mean value of 3.714 ± 1.204; in the satisfaction attribute, Task 2 and Task 3 were the highest rated with a mean value of 4.214 ± 0.801 and 4.214 ± 0.892, respectively; lastly, in time on task attribute, the top rated activity was Task 1 with a mean value of 3.857 ± 1.027. The Friedman test was conducted, considering a χ^2-distribution with $k = 2$ degrees of freedom for each group of attributes, to determine

whether task ease, satisfaction, and task on-time scores differed between the first, second, and third trials. The results show no significant difference in any of the three groups, Friedman statistic $\chi^2(2) = 1.2$, $p = 0.5488$ (task ease), $\chi^2(2) = 0.5$, $p = 0.778$ (satisfaction), and $\chi^2(2) = 2.545$, $p = 0.280$ (time on task). Therefore, we fail to reject the null hypothesis and conclude that there is no difference in attribute scores (task ease, satisfaction, and time on task) between the first, second, and third assembly activities for each group.

Figure 9.11 Single Usability Metric, results of the three virtual assembly activities.

Participants reported Task 1 with the highest rating ($11.866/15$). Although the scores are close to each other, it is notable that the scores for Task 2 ($11.466/15$) and Task 3 ($11.600/15$) were lower than the first one. This suggests that participants experienced less ease and satisfaction, and spent less time, as the experiment progressed.

- Task Completion Time (TCT): Figure 9.12 presents the recorded times of seven activities during the experimental protocol as described in Sec. 9.3.2 (Experiment procedure). A training phase was implemented to allow the user to adapt to the VR environment, followed by a coffee-making task and an assembly task. The training phase had a duration with a mean value of $241.14 \pm 74.06s$. Coffee-making in Task 3 registered the lowest time with a mean value of $99.57 \pm 17.65s$, a difference of $40.0s$ compared with Task 1, which registered a mean value of $139.5 \pm 23.89s$. Task 3 obtained the lowest time with a mean value of $263.0 \pm 75.45s$, a difference of $46.64s$ compared with Task 1, which registered a mean value of $309.64 \pm 86.6s$. The Friedman test was conducted on each activity time to determine whether time scores differed between the first, second, and third trials. The results show no significant difference among the three activities, with a Friedman statistic $\chi^2(2) = 5.333$ and $p = 0.069$.

Therefore, we did not reject the null hypothesis and concluded that no time difference was found between assembly activities.

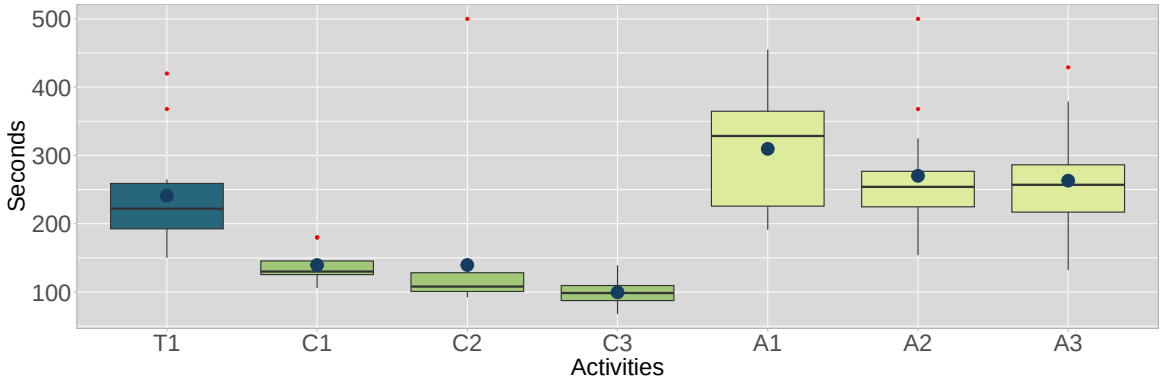


Figure 9.12 Time on Task results, mean value is represented by the dark circle on the plot.

Post-Experiment

While post-activity assessments focused on task-specific reactions, the post-experiment questionnaires provided a global perspective on the user experience. These measures enabled the evaluation of immersion quality, perceived system usability, and physical or cognitive strain, contributing to a comprehensive assessment of the VR platform's suitability for fall-risk monitoring.

iGroup Presence Questionnaire (IPQ) Table 9.7 presents the IPQ results. The scale's internal consistency was measured using Cronbach's alpha, which was $\alpha = 0.519$. The confidence interval was 0.95. The following attributes are ranked in order of decreasing rated score, from highest to lowest. For each attribute, the qualitative grading scale proposed by Melo in [322] is also mentioned. The highest ranked score was "Presence" with a qualitative score of excellent; then, "Spatial presence" with a qualitative score of unacceptable; followed by "Involvement" with a qualitative score of unacceptable; finally, "Experienced realism" with a qualitative score of unacceptable. The results suggest that the VR system effectively created a convincing immersive experience that engages users. However, the spatial presence was slightly lower than the general presence, suggesting that some elements may reminded participants that they were in a simulation. Involvement proposes that participants may experience distractions, cognitive load, or external factors. Finally, the virtual environment did not feel entirely realistic for the participants.

Table 9.7 IGroup Presence Questionnaire.

Attribute	Mean	Standard Deviation	Lower Bond	Upper Bond
Presence	5.285	0.913	4.758	5.813
Spatial presence	3.500	0.951	2.318	4.681
Involvement	3.267	0.604	2.306	4.229
Experienced realism	2.357	0.728	1.198	3.516

System Usability Scale (SUS) Figure 9.13 presents the SUS results obtained from the 14 participants. To assess the reliability of the questionnaire, Cronbach's alpha was computed, yielding a value of $\alpha = 0.229$. The highest individual score was 92.5 (best imaginable), and the lowest was 55 (good). The combined mean score was $74.10/100$ (excellent). Martínez-Gutiérrez presented the interpretation of the SUS score in [323]. The combined mean score indicates that participants found the virtual environment easy to use and rated the performed interaction as not complex, thereby making the participant confident during the session.

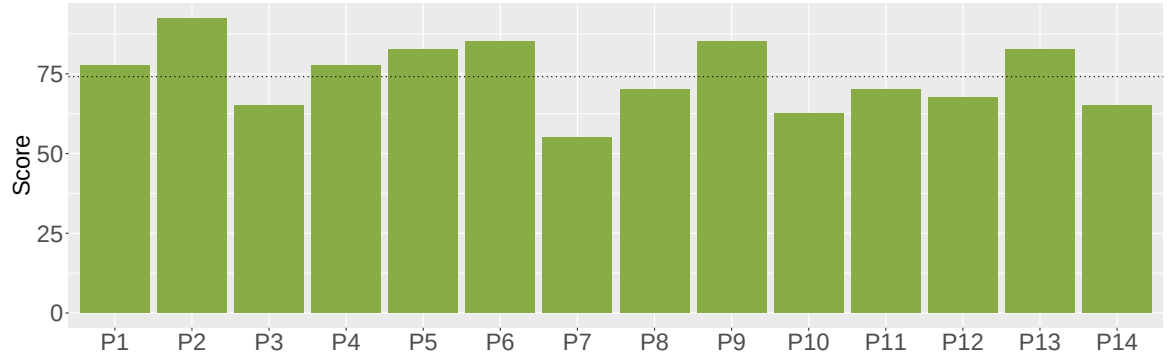


Figure 9.13 System Usability Scale, participants' individual score and mean combined score.

Usability Metric for User Experience (UMUX) Table 9.8 shows the UMUX survey results. The survey was adapted to the context of the virtual environment (e.g., item 1: "The virtual environment capabilities meet my requirements"). The internal reliability of the scale was confirmed with a Cronbach's alpha of $\alpha = 0.144$. The total UMUX score was $(76.190/100)$. The confidence interval was 0.95. The maximum and minimum scores represent a single score for each attribute. The "Overall" attribute received the highest score among participants, while the "Effectiveness" attribute received the lowest score. We can infer from the results that the proposed system may be aesthetically pleasing, engaging, or easy to use, resulting in a high overall perception.

Table 9.8 Usability Metric for User Experience.

Attribute	Max	Min	Mean	Standard Deviation	Lower Bond	Upper Bond
Effectiveness	6	3	4.500	0.759	4.061	4.938
Satisfaction	6	1	4.642	1.150	3.978	5.307
Overall	6	1	4.857	1.350	4.077	5.636
Efficiency	6	2	4.285	1.540	3.396	5.175

Table 9.9 User Fatigue Survey.

Symptom	Max	Min	Mean	Standard Deviation	% Average and Below Average
Tired eyes	3	0	1.285	0.825	64.28%
Clear Vision	3	1	2.214	0.699	64.28%
Eye dryness	2	0	1.000	0.784	71.42%
Tired back	3	0	1.000	1.109	71.42%
Tired neck	3	0	0.857	1.167	57.14%
Severe headache	3	0	0.428	0.851	71.42%
Sleepy feeling	3	0	1.285	1.069	57.14%

However, low effectiveness indicates that participants struggled to complete their activities successfully or efficiently.

User Fatigue Survey (UFS) Table 9.9 presents the UFS results. The maximum and minimum scores represent a single score for each symptom. At least one participant experienced the highest scale of severity (3) in all user fatigue symptoms. Clear vision was the highest-rated symptom, as in "too clear" vision after the experimental protocol. More than half of the participants experienced fatigue equal to or below the average symptom score.

Virtual Reality System Usability Questionnaire (VRSUQ) Table 9.10 presents the results of the VRSUQ questionnaire. The internal consistency of the survey data was evaluated using Cronbach's alpha, yielding a coefficient of $\alpha = 0.664$. The confidence interval was 0.95. The "Satisfaction" attribute received the highest score among participants, while the "Effectiveness" dimension received the lowest score. We can infer that participants enjoyed the experience (satisfaction) despite the potential challenges. The VR system should have engaging visuals (efficiency), good interaction design, or an immersive experience contributing to positive user sentiment. However, with low effectiveness, users may struggle with the controls, navigation, or precision, rendering the work equipment assembly less effective. The VRSUQ single score was $69.841/100$. Nomenclature and abbreviations for Table IV: LB: lower bond, UB: upper bond.

Table 9.10 Virtual Reality System Usability Questionnaire Scores.

Attribute	Mean	Standard Deviation	Lower Bond	Upper Bond
Effectiveness	3.380	0.329	2.561	4.200
Efficiency	3.952	0.459	2.811	5.093
Satisfaction	4.027	0.459	2.906	5.188

Table 9.11 System and Users Errors Reports.

Task	Repetitions	Mistakes	Interactions	Accuracy
Training	1	9	70	87.15 %
Activity 1	1	6	140	95.72 %
Activity 2	1	3	140	97.86 %
Activity 3	1	3	140	97.86 %

System Performance Analysis

Table 9.11 reports the system error statistics. There are two conditions in the current report: a) repetitions, which are the number of times that the activity was reset and repeated because of a system fault (the virtual environment crashed or froze); and b) mistakes, which are considered a mistake when the participant dropped off an object with which it was interacting. The virtual environment system ran 56 times (14 sessions of training and 42 sessions of assembling material) and recorded four repetitions after having all participants' sessions; the virtual environment system's accuracy was 92.86%. During the training session, 14 participants interacted with five pieces of assembling material, resulting in 70 interactions, with nine mistakes (12.85%), and an interaction accuracy of 87.15%. After each assembly task, participants interacted with 10 pieces of assembly material, resulting in a total of 140 interactions. In Task 1, the participants reported six mistakes (4.28%), resulting in a virtual assembly accuracy of 95.72%. Then, in Tasks 2 and 3, participants made three mistakes each, respectively (2.14%), which resulted in a virtual assembly accuracy of 97.86%.

9.5.3 Fall-risk Estimation

This section presents the results of applying machine learning models to the XoFallVR dataset for classifying the three fall-risk levels (Low, Medium, and High) and the simplified binary problem (Low vs. High risk). The findings are divided into the performance of a Generalized Model (using Leave-One-Subject-Out cross-validation) and a Personalized Model (using k-fold cross-validation per subject), contrasting the models' ability to generalize with their ability to achieve maximum individual accuracy.

Table 9.12 Evaluation of the given modalities and classifiers on the three fall risk levels (low vs. medium vs high) classification task - Validation. Abbreviations: DT = decision tree, DA = discriminant analysis, NB = naïve Bayes, kNN = k-nearest neighbor, Ens = ensemble, MLP = multilayer perceptron. Baseline: RG = random guesser, and MG = majority guesser, balanced accuracy: 33.33%, and F1-score: 29.12±0.25% and 25.15±0.18%, respectively.

	DT		*DA*		*NB*		*KNN*		*Ens*		*MLP*	
	bACC	F1	bACC	F1	bACC	F1	bACC	F1	bACC	F1	bACC	F1
Empatica E4:												
ACC	44.44±2.34	23.42±5.4	42.22±2.7	25.32±7.36	43.18±1.23	28.32±1.91	**63.93±2.93**	**50.15±2.89**	52.52±3.25	42.41±4.48	41.17±3.42	20.14±7.35
BVP	39.54±4.62	18.75±9.15	38.61±2.5	15.25±5.07	41.27±1.3	24.93±3.2	**71.72±10.42**	**58.53±11.96**	48.53±7.18	36.59±8.86	37.62±4.16	13.29±8.61
EDA	47.43±1.96	23.8±3.76	47.36±2.54	29.01±6.63	44.78±1.8	35.4±5.01	**73.6±5.75**	**60.24±5.81**	59.46±9.54	51.3±9.94	43.98±4.76	20.33±7.42
TEMP	40.21±0.87	20.21±1.69	38.4±1.98	17.3±4.94	39.46±1.02	21.76±1.49	**72.15±8.38**	**58.12±9.82**	53.69±14.64	43.31±19.87	35.23±2.33	9±5.9
All	52.85±7.56	35.61±14.47	47.57±3.14	31.56±7.16	44.53±3.54	33.83±6.55	**80.85±7.88**	**68.79±10.19**	59.68±5.94	52.06±7.05	44.97±6.62	22.97±11.3
Polar H10:												
ACC	59.7±2.14	51.06±3.29	54.22±1.06	35.79±3.54	55.95±0.88	39.52±1.67	**65.19±5.29**	52.01±5.4	60.8±2.5	**53.38±3.3**	52.19±5.49	29.96±8.22
BR	36.75±7.55	12.69±14.18	34.6±1.38	9.08±5.06	37.24±1.19	16.17±3.17	**64.53±8.47**	**51.08±9.95**	43.71±6.37	33.86±6.42	33.71±0.95	5.75±3.76
ECG	53.71±13.45	40.06±17.29	38.4±2.88	17.71±7.74	39.25±2.42	22.4±4.32	**77.53±9.07**	**65.09±11.91**	52.87±15.32	42.85±16.66	37.42±4.47	14.57±11.21
All	59.52±5.64	47.99±9.76	54.78±2.09	37.28±5.35	52.45±4	37.12±5.1	**79.17±8.26**	**65.87±10.39**	65.83±13.83	58.96±15.48	53.09±5.96	33.68±10.45
Vive Pro 2:												
EYE	55.76±2.68	44.38±3.39	51.98±2.18	42.37±3.41	51.64±1.68	44.42±1.64	**74.34±9.33**	**67.15±11.72**	61.95±11.32	53.92±12.29	44.67±8.05	25.52±15.78
All Sensors:												
ACC	60.52±5.04	50.27±8.13	54.96±1.61	37.71±4.23	57.46±1.08	42.11±2.11	**76.54±6.28**	**63.6±7.18**	62.63±2.81	55.94±3.05	54.93±1.76	37.44±7.36
Heart	44.75±7.22	30.26±11.37	39.52±3.26	20.71±8.72	41.12±2.69	25.83±4.86	**78.86±9.08**	**66.57±10.99**	53.03±15.33	42.81±17.24	40.11±3.95	22.46±9.27
All	62.78±5.35	50.56±11.26	60.32±3.83	49.51±5.57	54.39±3.71	44.94±4.5	**90.29±5.43**	**80.26±7.29**	83.51±16.19	81.32±17.33	56.18±11.08	36.59±18.31

Generalized model

The following results report the tentative training of models that can generalize across all subjects. The data were evaluated using Leave-One-Subject-Out (LOSO) cross-validation, considering the 14 subjects, and the results are reported as the mean and standard deviation. Considering the task with three classes in validation, as shown in Table 9.12, the best model is always KNN, with balanced accuracy ranging from approximately 65% to 90%, and F1-score ranging from 50% to 80%. The best model is the one that considers all sensors. However, the test results are significantly worse because the model fails to generalize well, as shown in Table 9.13. Some considerations are that accelerations and pupilometry are signals that improve classification. Additionally, the data from the Polar H10 yields better results than those obtained from the Empatica E4. The best models are Naive Bayes, which utilizes the acceleration data from the Polar H10, and the decision tree ensemble, considering all available data. In both cases, the models outperform the random or majority classifier, although the results are not particularly strong. They show how difficult it is to generalize across subjects.

In the case of binary classification (low vs. high risk), KNN consistently performs the best algorithm in validation, achieving a balanced accuracy of up to 98% and an F1-score of up to 95% using all sensors, as shown in Table 9.14. The performance is about 50% better than that of the random and majority classifiers. In the test, however, performance drops substantially, reaching a balanced accuracy of 67% in the best case using the multilayer perceptron as a model and a best F1-score of 75% with discriminant analysis using Polar H10 accelerations, Table 9.15. In some cases, however, the performance of the models is

Table 9.13 Evaluation of the given modalities and classifiers on the three fall risk levels (low vs. medium vs high) classification task - Test. Abbreviations: DT = decision tree, DA = discriminant analysis, NB = naïve Bayes, kNN = k-nearest neighbor, Ens = ensemble, MLP = multilayer perceptron. Baseline: RG = random guesser, and MG = majority guesser, balanced accuracy: 33.33%, and F1-score: 28.58±1.78% and 24.97±2.48%, respectively.

	DT		*DA*		*NB*		*KNN*		*Ens*		*MLP*	
	bACC	F1	bACC	F1	bACC	F1	bACC	F1	bACC	F1	bACC	F1
Empatica E4:												
ACC	**39.86±10.56**	18.78±9.05	35.07±4.1	15.01±5.96	35.39±4.43	17.74±9.27	36.65±6.87	23.63±9.18	37.07±8.76	**28.15±9.86**	36.12±5.94	15.49±8.58
BVP	30.98±4.59	13.35±4.51	33±5.54	10.69±6.22	32.55±6.59	16.95±9.69	31.79±7.44	**24.97±6.23**	31.97±11.29	23.44±14.43	**33.71±5.51**	10.75±10.52
EDA	**40.07±9.72**	19.19±10.54	39.59±9.38	20.99±5.86	35.62±5.81	27.71±6.04	40.05±7.89	**32.72±4.75**	38±6.72	28.33±11.06	38.77±5.8	18±7.47
TEMP	34.45±1.64	14.55±6.25	33.21±0.44	7.56±7.48	**34.67±3.95**	13.09±9.59	34.02±6.88	**26.7±9.47**	34.12±3.9	22.41±9.17	33.33±0	5.13±5.31
All	**40.07±12.23**	24.23±10.52	35.73±7.6	19.4±8.4	34.77±4.06	24.07±6.39	35.43±5.97	**28.06±9.31**	37.75±7.84	27.24±10.3	37.53±6.35	18.13±9.68
Polar H10:												
ACC	44.21±9.56	35.23±10.43	50.62±10.14	29.89±10.07	**51.2±10.12**	30.93±11.63	48±8.17	36.49±6.6	46.04±8.67	**39.18±9.51**	48.6±10.94	24.15±12.25
BR	32.28±2.84	8.52±10.84	33.33±0	4.04±3.93	33.04±2.96	7.62±4.94	33.15±9.98	**24.3±8.9**	**35.27±9.04**	18.42±8.85	33.46±0.49	4.29±3.96
ECG	**36.34±6.42**	23.61±8.37	31±6.6	8.15±8.4	32.71±4.87	14.6±10.05	29.8±6.39	**25.5±6.41**	32.37±4.79	21.26±8.58	31.78±6.97	7.48±8.59
All	41.22±10.14	33.61±8.88	**50.88±10.52**	29.91±12.01	45.3±10.44	28.79±16.22	42.04±9.23	32.12±8.81	44.72±8.46	**37.97±9.48**	47.52±11.39	26.73±9.05
Vive Pro 2:												
EYE	40.51±6.12	29.11±10.74	**41.52±10.18**	30.58±15.97	40.35±9.36	31.24±16.75	37.42±6.49	**31.89±4.58**	38.91±6.63	30.12±11.72	35.75±4.6	17.78±10.48
All Sensors:												
ACC	42.73±9.51	33.01±12.03	**49.43±9.14**	30±11.52	48.61±8.93	31.41±7.12	46.12±8.87	37.77±11.36	44.34±10.21	**37.88±9.7**	47.82±8.98	30.56±13.64
Heart	31.97±6.44	15.94±7.67	31.62±5.2	9.51±7.4	31.86±6.02	16.11±10.33	**34.45±7.35**	**28.4±8.75**	33.44±6.51	25.35±6.57	32.45±5.94	10.23±7.62
All	42.83±9.04	27.21±11.35	45.45±14.87	33±17.09	41.17±6.78	34.99±11.23	47.57±9.52	39.71±8.77	46.06±10.82	**41.85±12.46**	**49.31±12.72**	26.29±13.11

Table 9.14 Evaluation of the given modalities and classifiers on the two fall risk levels (low vs high) classification task - Test. Abbreviations: DT = decision tree, DA = discriminant analysis, NB = naïve Bayes, kNN = k-nearest neighbor, Ens = ensemble, MLP = multilayer perceptron. Baseline: RG = random guesser, and MG = majority guesser, balanced accuracy: 50%, and F1-score: 44.11±0.61% and 45.18±0.41%, respectively.

	DT		*DA*		*NB*		*KNN*		*Ens*		*MLP*	
	bACC	F1	bACC	F1	bACC	F1	bACC	F1	bACC	F1	bACC	F1
Empatica E4:												
ACC	65.62±2.63	55.44±4.38	62.65±2.35	53.63±2.21	62.92±1.65	51.8±2.01	**83.44±6.11**	**74.68±7.08**	72.58±6.84	66.48±5.99	62.03±3.69	50.88±11.38
BVP	70.16±5.98	63.06±6.45	59.79±1.03	50.69±1.74	60.49±1.22	47.14±1.6	**83.88±7.65**	**75.58±8.46**	67.21±10.24	60.37±10.5	60.38±3.31	49.18±10.65
EDA	76.01±5.52	67.54±6.55	71.57±2.01	63.67±3.07	59.94±2.51	44.82±4.02	**92.13±4.08**	**85.56±6.35**	75.2±7.24	69.34±8.15	67.07±9.39	54.08±21.25
TEMP	83.98±12.18	77.97±14.19	56.73±4.22	37.61±8.86	59.25±1.11	41.64±1.89	**89.74±5.06**	**80.79±6.87**	81.7±12.83	78.22±14.92	53.85±4.9	26.45±14.26
All	76.49±6.17	68.57±7.63	70.84±2.26	62.99±3.09	66.44±2.13	54.85±3.44	**89.16±6.02**	**81.08±8.71**	83.7±3.69	77.5±4.95	70.73±2.39	62.42±3.58
Polar H10:												
ACC	77.97±2.81	68.65±3.55	78.65±1.13	74.52±1.94	78.75±1.13	74.27±1.66	**81.44±1.47**	**74.61±1.69**	78.81±2.69	71.43±3.7	79.71±1	73.36±1.77
BR	76.57±12.86	68.87±13.2	50.52±1.51	28.84±12.28	54.56±1.46	31.68±4.19	**83.91±6.05**	**76.06±7.76**	75.18±17.23	68.62±17.21	50.98±2.27	19.77±11.38
ECG	77.77±11.4	69.82±12.48	51.37±4.36	35.43±13.29	60.69±1.28	47.32±3.65	**93.11±5.04**	**89.75±8.06**	79.2±16.96	73.15±18.61	51.89±3.87	23.01±16.03
All	79.22±3.91	72.08±6.1	79.73±2.33	75.53±2.79	77.45±3.12	72.17±5.28	**88.96±7.02**	**83.36±8.89**	79.37±3.31	71.07±4.55	78.36±8.46	70.39±16.07
Vive Pro 2:												
EYE	76.94±6.99	68.69±8.59	69.32±2.92	62.19±2.79	67.12±1.57	61.86±3.05	**84±6.09**	**77.67±5.5**	77.34±8.57	71.7±11.15	69.03±5.86	59.32±13.99
All Sensors:												
ACC	77.62±3.25	69.17±4.14	78.78±1.89	73.56±2.38	79.81±1.3	74.29±2.15	**91.4±4.95**	**85.68±7.42**	82.01±4.89	74.95±7.05	79.46±2.33	72.41±2.13
Heart	73.97±8.75	66.96±9.27	56.57±6.33	43.56±14.89	60.9±2.78	48.18±3.95	**87.2±9.72**	**81.26±12.86**	76.12±12.7	69.96±14.57	57.86±6.77	39.06±19.33
All	80.1±5.58	72.33±8.61	84.13±3.08	79.23±3.48	76.32±1.84	69.14±2.67	**98.17±2.19**	**95.96±3.96**	90.06±9.03	85.88±12.64	85.07±11.02	76.84±18.77

comparable to, or even worse than, that of the random or majority classifier. Even in the case of the binary model, generalization is not straightforward.

Personalized model

Following the results from the Leave-One-Subject-Out (LOSO) cross-validation, which demonstrated the challenge of generalizing fall-risk classification models across subjects, the classification strategy was shifted to a personalized approach. This approach utilizes k-fold cross-validation, training and testing the model exclusively on a single subject's data from different repetitions. It uses two sequences for training and validation, and the last one

Table 9.15 Evaluation of the given modalities and classifiers on the two fall risk levels (low vs high) classification task - Test. Abbreviations: DT = decision tree, DA = discriminant analysis, NB = naïve Bayes, kNN = k-nearest neighbor, Ens = ensemble, MLP = multilayer perceptron. Baseline: RG = random guesser, and MG = majority guesser, balanced accuracy: 50%, and F1-score: 42.17±3.2% and 45.62±3.96%, respectively.

	DT		*DA*		*NB*		*KNN*		*Ens*		*MLP*	
	bACC	F1	bACC	F1	bACC	F1	bACC	F1	bACC	F1	bACC	F1
Empatica E4:												
ACC	**57.28±12.07**	42.3±16.38	53.23±7.55	39.2±13.8	52.75±6.25	39.85±18.4	56.03±9.71	**49.18±7.8**	56.61±11.23	42.91±14.64	52.79±7.55	38.37±14.04
BVP	48.25±10.46	41.17±12.52	51.53±9.28	38.27±16.86	50.16±8.95	37.15±11.68	47.41±10.72	**42.03±11.91**	46.94±9.49	38.83±13.55	**51.45±10.38**	34.48±17.51
EDA	57.28±12.44	49.23±14.5	**64.48±11.28**	**55.03±17.1**	57.57±10.79	35.8±18.93	60.36±11.25	54.44±11.51	57.52±13.78	51.83±16.84	59.01±12.11	44.8±22.21
TEMP	53.15±7.89	45.64±11.06	51.57±7.75	25.91±16.65	52.27±4.22	28.35±14.96	49.79±13.45	43.17±13.98	**55.76±12.66**	**45.79±19.69**	50.62±2.05	16.56±11.85
All	56.63±11.78	50.24±16.03	**59.45±14.1**	**53.43±14.4**	56.5±12.41	45.28±17.49	54.82±10.88	50.82±9.64	56.33±11.44	49.33±15.66	58.25±10.31	49.49±14.75
Polar H10:												
ACC	68.33±15.27	54.7±23.98	74.95±15.22	**67.09±19.4**	72.14±17.8	63.95±20.78	72.36±15.63	62.72±22.5	70.84±16.53	60.69±23.5	**75.15±15.84**	65.32±21.46
BR	45.8±10.84	36.24±13.5	50.62±1.08	22.71±17.57	50.9±2.82	22.72±13.05	**51.09±12**	**45.71±7.46**	46.03±6.32	33.15±14.28	50±0	12.61±7.26
ECG	45.59±13.46	40.39±11.72	**52.04±8.58**	30.09±18.01	48.9±10.46	32.54±17.7	49.52±5.88	**46.53±7.27**	44.15±13.2	35.63±15.09	49.52±3.49	16.25±14.18
All	70.42±13.75	57.63±21.96	73.82±14.8	**65.33±20.93**	68.03±14.91	60.4±19.17	68.35±10.72	58.94±15.02	69.74±15.18	57.74±23.06	**74.87±15.3**	65.3±21.32
Vive Pro 2:												
EYE	54.63±13.8	42.7±18.3	**57.56±10.78**	48.96±17.34	53.6±9.53	44.79±19.38	56.83±9.46	**54.24±8.32**	52.61±12.26	45.22±17.19	56.37±8.63	44.47±16.23
All Sensors:												
ACC	69.18±15.29	55.89±23.38	**71.99±15.46**	62.4±23.52	70.8±14.11	**64.96±16.5**	70.81±16.86	63.78±18.7	67.93±11.09	56.79±19.46	71.51±15.4	62.08±22.49
Heart	49.99±13.35	37.98±16.48	46.18±9.34	30.3±14.45	47.47±12.06	34.31±18.11	44.8±10.64	**42.38±9.59**	**50.9±11.61**	38.98±18.65	48.46±9.23	24.76±15.76
All	68.02±16.36	59.45±21.6	69.97±16.36	**64.85±22.75**	63.41±13.91	53.99±15.71	68.14±13.35	63.61±13.82	66.37±13.95	59.5±18.62	**72.05±15.96**	61.79±26.23

for testing. The goal is to establish the upper bound of achievable performance when the classifier is tailored to the individual's unique physiological and kinematic patterns.

Some models, following the feature selection and training process, overfitted the validation data and were removed from the results reported below because they were considered ineffective and uninformative for the problem at hand. For this reason, some models with more initial features may perform worse than models with fewer features, or the models reported in the tables are RG or MG.

The models designed to classify the three fall-risk levels (Low, Medium, and High) demonstrated near-perfect performance during the validation phase (Tables 9.16 and 9.17). The best overall personalized performance, achieved by the optimal classifier/modality combination per subject and repetition (Table 9.18), yielded an average balanced accuracy of $99.87 \pm 0.10\%$ for the Slip scenario and an F1-score of $99.63 \pm 0.22\%$. This indicates that within-subject variability between training and validation folds is minimal, confirming that the extracted features are highly consistent for individual physiological and psychological states across repeated trials.

However, the definitive test performance on unseen repetitions (Table 9.21) revealed a noticeable reduction, although the results still significantly surpass the generalized LOSO test outcomes (Table 9.13). Across the three simulated scenarios (Slip, Trip, and Fall), the average balanced accuracy for the personalized models clustered around 60%, with the highest overall average balanced accuracy being $62.46 \pm 13.96\%$ (F1-score $51.08 \pm 12.15\%$) for the Fall scenario. For the Trip scenario, the performance was also robust, achieving a balanced accuracy of 60.33 ± 11.87 and an F1-score of $54.73 \pm 9.02\%$.

In contrast to the LOSO findings, where performance was close to random guessing for some models, the personalized models consistently outperformed the baseline. The dominant classifiers shifted, with Decision Trees (DT), k-Nearest Neighbors (kNN), and Ensemble (Ens) models often achieving the highest scores across various modalities, especially when utilizing combined data such as Polar H10 ACC, All Sensors, and Empatica E4 ACC (Tables 9.19 and 9.20).

Table 9.16 Balanced accuracy of the provided modalities and personalized classifiers on the three levels of fall risk (low vs. medium vs. high) for each subject - Validation.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14
Empatica E4:														
ACC	84.38±7.23 (DT)	90.68±8.92 (Ens)	79.98±19.91 (Ens)	96.91±1.61 (DT)	93.61±5.23 (DT)	87.4±6.34 (Ens)	84.95±8.43 (Ens)	86.54±11.24 (DT)	84.42±13.99 (Ens)	87.61±15.28 (DT)	88.37±7.38 (Ens)	88±9.08 (Ens)	96.19±3.38 (DT)	96.56±3.52 (DT)
BVP	99.62±0.42 (KNN)	87.1±18.51 (Ens)	96.19±3.68 (DT)	92.46±1.47 (Ens)	97.34±4.34 (DT)	99.34±1.08 (KNN)	90.18±6.6 (Ens)	83.12±3.17 (Ens)	92±5.11 (DT)	96.2±3.72 (DT)	91.83±0.49 (Ens)	92.85±1.97 (Ens)	96.05±3.31 (DT)	98.61±1.24 (DT)
EDA	92.41±7.92 (Ens)	95.25±4.79 (Ens)	96.08±1.93 (Ens)	92.27±5.67 (Ens)	99.4±0.26 (KNN)	94.76±4.3 (Ens)	99.12±0.42 (Ens)	99.14±0.57 (KNN)	94.65±5.14 (DT)	94.28±7.54 (Ens)	98.41±2.52 (DT)	92.37±5.25 (Ens)	99.25±0.97 (KNN)	92.55±9.82 (DT)
TEMP	99.75±0.29 (KNN)	99.24±0.76 (DT)	99.53±0.71 (DT)	98.71±0.85 (DT)	99.34±0.52 (KNN)	99.58±0.3 (KNN)	99.79±0.23 (DT)	99.52±0.33 (KNN)	98.48±1.66 (DT)	99.26±0.75 (DT)	99.2±0.9 (DT)	99.69±0.25 (DT)	86.93±15.73 (DT)	99.21±0.5 (DT)
All	99.81±0.22 (DT)	99.32±0.84 (DT)	99.58±0.51 (DT)	99.73±0.39 (DT)	99.82±0.09 (KNN)	99.58±0.32 (Ens)	99.89±0.1 (DT)	99.95±0.03 (DT)	99.74±0.3 (KNN)	99.38±0.35 (Ens)	99.9±0.08 (DT)	98.96±1.55 (DT)	99.78±0.16 (DT)	99.86±0.11 (DT)
Polar H10:														
ACC	89.63±6.35 (Ens)	92.54±9.21 (Ens)	92.11±5.82 (Ens)	93.38±8.16 (Ens)	98.45±1.31 (DT)	81.01±1.69 (Ens)	97.77±0.79 (DT)	95.2±4.9 (KNN)	88.84±3.39 (Ens)	97.96±0.53 (DT)	99.49±0.18 (DT)	93.2±8.03 (Ens)	99.05±0.28 (DT)	97.94±0.82 (DT)
BR	99.27±0.4 (KNN)	99.69±0.41 (KNN)	99.73±0.14 (KNN)	99.9±0.02 (KNN)	99.85±0.08 (DT)	99.82±0.09 (DT)	99.78±0.17 (KNN)	99.64±0.2 (KNN)	99.75±0.19 (KNN)	99.92±0.02 (KNN)	99.84±0.02 (KNN)	99.72±0.35 (KNN)	99.89±0.02 (DT)	99.79±0.11 (KNN)
ECG	89.09±17.24 (DT)	99.6±0.02 (KNN)	94.08±4.48 (Ens)	92.46±10.39 (Ens)	90.48±6.24 (DT)	90.15±5.54 (Ens)	93.06±1.07 (Ens)	95.58±6.82 (KNN)	98.5±2.08 (KNN)	91.36±12.68 (DT)	97.55±2.67 (DT)	93.33±8.96 (DT)	97.94±0.57 (DT)	98.3±0.1 (DT)
All	99.73±0.22 (DT)	99.62±0.31 (DT)	98.87±1.26 (KNN)	99.77±0.3 (DT)	87.06±3.29 (DT)	98.18±1.37 (DT)	99.38±0.81 (DT)	99.75±0.17 (DT)	98.64±1.32 (DT)	98.01±1.7 (Ens)	98.8±1.8 (DT)	99.55±0.19 (DT)	77.71±4.26 (DA)	99.55±0.36 (DT)
Vive Pro 2:														
EYE	91.66±6.21 (Ens)	99.56±0.36 (KNN)	98.61±0.7 (DT)	99.9±0.08 (DT)	99.64±0.31 (DT)	99.24±1.12 (KNN)	99.41±0.77 (DT)	99.75±0.14 (DT)	88.14±11.72 (DT)	94.69±6.52 (DT)	98.33±1.46 (DT)	98.54±1.25 (KNN)	84.77±18.67 (KNN)	99.34±0.84 (DT)
All Sensors:														
ACC	97.8±1.07 (KNN)	98.48±2.2 (KNN)	97.71±1.37 (KNN)	99.79±0.22 (DT)	99.54±0.25 (DT)	91.67±9.94 (Ens)	99.72±0.23 (DT)	99.91±0.09 (DT)	97.2±3.2 (KNN)	98.86±1.1 (DT)	97.8±3.33 (DT)	98.94±0.67 (DT)	99.45±0.32 (Ens)	99.65±0.09 (DT)
Heart	99.76±0.22 (DT)	98.18±2.2 (DT)	98.72±0.52 (Ens)	98.79±1.41 (Ens)	96.2±6.09 (DT)	98.6±1.27 (Ens)	97.65±0.91 (Ens)	98.25±0.41 (Ens)	99.15±1.18 (DT)	99.15±0.18 (DT)	96.17±2.36 (Ens)	98.28±0.56 (Ens)	98.97±0.79 (DT)	99.46±0.52 (DT)
All	99.79±0.13 (KNN)	99.83±0.2 (Ens)	99.93±0.07 (DT)	99.87±0.06 (KNN)	99.93±0.04 (DT)	99.58±0.35 (KNN)	99.87±0.12 (Ens)	99.81±0.21 (DT)	99.58±0.33 (KNN)	99.67±0.16 (KNN)	99.92±0.11 (DT)	99.85±0.16 (DT)	88.22±8.88 (DA)	99.59±0.39 (DT)

Table 9.17 F1-score of the provided modalities and personalized classifiers on the three levels of fall risk (low vs. medium vs. high) for each subject - Validation.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14
Empatica E4:														
ACC	73.69±5.28 (Ens)	82.83±14.16 (Ens)	68.41±16.59 (Ens)	85.11±5.96 (DT)	84.15±7.22 (Ens)	74.63±10.63 (Ens)	70.39±10.69 (Ens)	72.42±21.06 (Ens)	73.25±18.7 (Ens)	81.58±21.07 (Ens)	73.17±12.91 (Ens)	79.97±16.04 (Ens)	91.67±7.3 (DT)	95.98±3.59 (DT)
BVP	99.7±0.15 (KNN)	81.67±21.48 (Ens)	91.98±6.32 (KNN)	80.43±0.95 (Ens)	92.97±9.35 (Ens)	97.49±3.91 (KNN)	75.43±12 (Ens)	65.36±6.32 (Ens)	77.93±8.27 (Ens)	92.23±4.16 (Ens)	82.03±1.29 (Ens)	83.9±2.25 (Ens)	95.71±3.35 (KNN)	98.49±1.22 (DT)
EDA	85.47±16.05 (Ens)	87.88±14.42 (Ens)	85.56±9.01 (Ens)	79.31±12.82 (Ens)	96.57±1.41 (KNN)	87.89±7.61 (Ens)	97.58±1.79 (Ens)	94.38±4.37 (KNN)	86.08±8.51 (Ens)	88.26±17.56 (Ens)	93.47±9.62 (DT)	83.62±12.27 (Ens)	98.34±1.7 (KNN)	91.57±10.15 (DT)
TEMP	98.81±0.3 (Ens)	95.66±4.36 (DT)	96.85±4.73 (DT)	91.76±4.56 (DT)	95.69±3 (KNN)	97.64±1.7 (KNN)	98.34±1.75 (DT)	96.5±0.12 (Ens)	92.91±5.71 (DT)	96.08±6.22 (DT)	96.21±2.24 (Ens)	98.52±0.4 (KNN)	77.78±21.36 (Ens)	99.17±0.5 (DT)
All	99.05±1 (DT)	96.63±4.16 (DT)	96.53±4.03 (DT)	97.99±2.69 (DT)	98.63±0.7 (KNN)	98.37±1.55 (Ens)	99.28±0.85 (DT)	99.67±0.16 (DT)	98.36±1.85 (KNN)	96.95±4.45 (Ens)	99.34±0.33 (DT)	95.65±5.95 (DT)	99.37±0.31 (DT)	99.76±0.28 (DT)
Polar H10:														
ACC	79.2±12.68 (Ens)	86.66±14.66 (Ens)	81.01±11.94 (Ens)	83.48±14.92 (Ens)	92.25±6 (DT)	67.44±3.13 (Ens)	88.36±3.44 (DT)	86.78±12.59 (Ens)	76.74±8.15 (Ens)	93.91±6.34 (DT)	97.59±1.21 (DT)	85.77±13.48 (Ens)	97.22±0.57 (DT)	97.57±1.14 (DT)
BR	98.79±1.55 (KNN)	99.29±0.55 (KNN)	98.19±1.32 (KNN)	99.41±0.11 (KNN)	99.24±0.52 (DT)	99.07±0.72 (DT)	99.86±0.09 (KNN)	98.71±0.67 (KNN)	99.73±0.23 (KNN)	99.8±0.13 (KNN)	99±0.18 (KNN)	99.71±0.18 (KNN)	99.74±0.11 (DT)	99.79±0.13 (KNN)
ECG	81.22±25.95 (Ens)	98.27±1.3 (KNN)	84.77±10.76 (Ens)	84.1±19.08 (Ens)	77.14±7.91 (Ens)	79.76±8.74 (Ens)	80.89±2.56 (Ens)	89.58±17.56 (KNN)	94.9±8.51 (KNN)	89.16±12.3 (Ens)	90.5±9.24 (DT)	84.83±10.42 (Ens)	95.3±0.97 (DT)	98.01±0.22 (DT)
All	98.54±1.14 (DT)	97.74±1.54 (DT)	93.59±5.25 (Ens)	98.44±1.86 (DT)	71.03±3.87 (Ens)	92.57±5.14 (DT)	95.73±5.61 (DT)	98.23±0.64 (Ens)	93.56±5.6 (DT)	96.18±2.52 (Ens)	95.27±6.11 (Ens)	97.68±0.84 (DT)	70.22±7.83 (DA)	99.53±0.33 (DT)
Vive Pro 2:														
EYE	88.22±3.76 (Ens)	97.53±1.48 (KNN)	91.78±3.36 (DT)	99.34±0.63 (DT)	98.06±1.64 (DT)	97.15±2.96 (Ens)	96.42±4.32 (DT)	98.06±1.63 (DT)	74.71±19.97 (Ens)	87.63±19.15 (DT)	91.99±5.63 (DT)	93.16±5.25 (KNN)	84.9±20.61 (KNN)	99.32±0.81 (DT)
All Sensors:														
ACC	90.33±4.64 (Ens)	94.14±4.68 (Ens)	90.11±3.74 (Ens)	98.35±1.54 (DT)	97.05±1.58 (DT)	82.74±15.6 (Ens)	97.91±1.72 (DT)	99.44±0.71 (DT)	92.52±4.88 (Ens)	97.18±2.21 (DT)	92.57±9.92 (DT)	94.56±1.94 (Ens)	98.64±0.87 (Ens)	99.53±0.09 (DT)
Heart	99.06±1 (DT)	91.3±8.54 (DT)	94.21±1.44 (Ens)	96.05±3.9 (Ens)	91.38±8.67 (Ens)	96.31±4.06 (Ens)	89.49±1.26 (Ens)	94±2.3 (Ens)	95.8±5.49 (DT)	97.87±1.09 (Ens)	88.93±6.87 (Ens)	95.2±2.17 (Ens)	97±2.51 (DT)	99.2±0.96 (DT)
All	98.65±0.64 (KNN)	99.07±1.12 (Ens)	99.34±0.73 (DT)	98.91±0.54 (KNN)	99.59±0.23 (DT)	97.67±1.7 (KNN)	98.95±1.17 (Ens)	98.44±1.46 (DT)	98.29±1.61 (Ens)	98.63±1.5 (DT)	99.64±0.45 (Ens)	99.18±0.47 (Ens)	87.78±12.03 (DA)	99.63±0.35 (DT)

Table 9.18 Balanced accuracy and F1-score of the best personalized classifiers on the three levels of fall risk (low vs. medium vs. high), for each subject and repetition - Validation.

		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	
Slip	bACC	99.91	99.94	99.98	99.93	99.91	99.85	99.91	99.97	99.62	99.91	99.79	99.67	99.88	99.89	99.87±0.10
	F1-score	99.55	99.62	99.97	99.45	99.43	99.11	99.82	99.84	99.49	99.89	99.82	99.61	99.62	99.65	99.63±0.22
Trip	bACC	99.95	99.96	99.86	99.81	99.91	99.89	99.78	99.92	99.66	99.90	99.97	99.98	99.91	99.75	99.87±0.09
	F1-score	99.69	99.60	98.55	99.28	99.47	99.77	99.96	99.52	99.77	99.86	99.98	99.62	99.77	99.83	99.62±0.35
Fall	bACC	99.56	99.60	99.95	99.97	99.98	99.71	99.99	99.97	99.96	99.94	99.99	99.92	99.89	99.96	99.88±0.14
	F1-score	99.86	98.66	99.49	99.49	99.86	98.33	99.80	99.65	99.95	99.64	99.14	99.92	99.83	99.90	99.54±0.48

Table 9.19 Balanced accuracy of the provided modalities and personalized classifiers on the three levels of fall risk (low vs. medium vs. high) for each subject - Test.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14
Empatica E4:														
ACC	**55.39±9.5 (KNN)**	39.46±5.76 (DT)	36.84±13.49 (KNN)	48.82±6.09 (Ens)	48.2±13.57 (Ens)	**55.83±6.83 (Ens)**	38.52±9.2 (KNN)	50.46±1.03 (Ens)	45.81±7.54 (DT)	61.49±5.84 (Ens)	49.68±6.92 (DT)	44.88±11.88 (KNN)	61.1±24.52 (KNN)	46.01±5.45 (KNN)
BVP	46.23±8.02 (DT)	37.94±2.7 (MLP)	54.89±8.56 (Ens)	37.97±11.62 (DT)	46.1±5.69 (Ens)	50.21±8.78 (DA)	52.13±12.59 (MLP)	**75.08±4.41 (Ens)**	33.49±0.59 (RG)	50.04±4.89 (DT)	48.14±10.38 (Ens)	35.17±2.27 (KNN)	53.74±4.2 (DT)	39.76±31.56 (DA)
EDA	33.55±0.37 (RG)	33.45±0.78 (RG)	38.43±8.81 (Ens)	36.15±3.76 (Ens)	33.33±0 (MG)	33.33±0 (DA)	40.83±10.03 (DT)	41.25±14.29 (KNN)	43.05±12.75 (KNN)	67.31±13.9 (Ens)	41.91±24.89 (DT)	35.18±5.16 (Ens)	50.77±14.96 (Ens)	43.44±20.18 (MLP)
TEMP	38.74±5.69 (DA)	39.02±8.94 (DA)	41.56±14.36 (DT)	**55.16±22.46 (DT)**	39.72±11.06 (DA)	35.9±16.98 (DT)	45.24±20.63 (DA)	34.75±11.81 (MLP)	43.51±10.34 (KNN)	44.32±6.3 (DA)	47.87±4.35 (Ens)	40.02±11.59 (DT)	38.08±12.21 (DT)	33.33±0 (MG)
All	51.15±3.26 (DT)	42.48±15.7 (MLP)	**56.55±17.13 (DA)**	48.25±18.67 (DA)	38.97±22.45 (KNN)	53.24±17.92 (DA)	42.56±12.76 (DA)	60.18±11.74 (DA)	40.64±11.67 (DA)	69.95±9.41 (DA)	**53.97±11.5 (KNN)**	47.18±23.66 (DT)	40.24±12.35 (MLP)	38.97±23.1 (DT)
Polar H10:														
ACC	35.7±1.95 (Ens)	47.15±7.59 (MLP)	45.11±8.71 (Ens)	45.08±15.35 (DT)	63.65±17.22 (DT)	45.75±13.74 (MLP)	48.34±26 (MLP)	49.67±3.23 (Ens)	**60.73±24.71 (MLP)**	72.63±5.77 (KNN)	52.49±31.21 (KNN)	52.62±19.64 (DT)	48.9±10.83 (DT)	47.04±5 (MLP)
BR	38.88±19.69 (Ens)	49.8±12.07 (DA)	33.33±0 (MG)	36.86±4.06 (MLP)	41.65±13.69 (DT)	37.75±7.65 (MLP)	42.6±12.16 (DA)	44.41±5.64 (DA)	33.82±0.49 (RG)	44.05±17.5 (KNN)	37.46±6.99 (MLP)	41.13±7.07 (MLP)	34.7±23.7 (DT)	46.76±9.77 (DT)
ECG	40.73±10.79 (Ens)	**50.96±12.22 (DA)**	41.01±8.8 (DT)	39.97±16.96 (DA)	53.01±0.43 (Ens)	39.98±24.61 (MLP)	33.46±0.13 (RG)	52.3±23.49 (Ens)	45.4±12.04 (DA)	49.01±22.54 (DT)	41.25±1.33 (MLP)	56.97±18.42 (MLP)	52.58±8.65 (Ens)	44±3.78 (KNN)
All	37.4±12.43 (DA)	49.16±24.14 (Ens)	48.35±13.03 (KNN)	40.04±3.36 (KNN)	55.07±4.06 (Ens)	49.28±2.7 (DA)	49.32±10.91 (Ens)	56.53±3.8 (DT)	60.57±22.19 (Ens)	74.21±6.2 (DT)	48.71±17.89 (MLP)	45.2±12.75 (DA)	49.49±10.87 (KNN)	49.89±6.72 (DA)
Vive Pro 2:														
EYE	44.64±13.14 (MLP)	38.12±25.26 (KNN)	44.64±8.99 (KNN)	39.56±14.33 (DT)	33.33±0 (MG)	50.81±15.15 (KNN)	43±14.32 (DA)	39.99±20.19 (DA)	34.93±2.76 (DA)	45.69±22.78 (Ens)	36.62±7.45 (KNN)	55.25±10.94 (DA)	47.09±11.89 (DA)	**56.45±12.97 (DT)**
All Sensors:														
ACC	33.33±0 (MG)	43.88±6.62 (DA)	35.57±2.81 (MLP)	48.62±6.31 (DT)	**65.93±17.65 (KNN)**	45.94±10.93 (MLP)	**52.18±2.94 (Ens)**	53.5±12.9 (MLP)	47.49±13.89 (DT)	**74.67±4.95 (KNN)**	47±5.1 (MLP)	50.12±15.07 (DA)	**74.8±9.21 (DA)**	45.37±6.18 (DA)
Heart	34±7.4 (Ens)	48.03±6.66 (Ens)	54.11±15.18 (Ens)	53.34±17.87 (DA)	51.57±3.73 (DT)	44.48±11.83 (KNN)	42.32±1.98 (DA)	66.79±14.31 (DA)	42.84±15.47 (DA)	53.7±11.53 (DA)	44.6±15.89 (MLP)	48.47±25.17 (KNN)	53.82±23.86 (Ens)	49.05±23.67 (DA)
All	47.6±7.56 (DA)	43.55±4.96 (KNN)	42.21±9.97 (DT)	49.93±8.11 (Ens)	49.52±12.16 (DT)	50.94±16.75 (MLP)	49.33±9.76 (KNN)	53.65±13.97 (DA)	44.41±11.72 (DT)	67.7±7.42 (DT)	49.82±14.85 (MLP)	**64.88±4.8 (DA)**	64.24±24.66 (DA)	41.95±10.47 (Ens)

Table 9.20 F1-score of the provided modalities and personalized classifiers on the three levels of fall risk (low vs. medium vs. high) for each subject - Test.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14
Empatica E4:														
ACC	**48.52±6.06 (KNN)**	39.46±0.96 (Ens)	29.63±18.85 (KNN)	**47.99±7.95 (Ens)**	42.54±1.4 (Ens)	**52.88±9.42 (Ens)**	34.16±13.69 (Ens)	40.86±6.43 (Ens)	40.68±4.15 (Ens)	48.59±3.36 (Ens)	37.76±8.83 (Ens)	42.16±14.46 (Ens)	50.83±19.51 (Ens)	37.26±4.7 (DA)
BVP	39.93±9.08 (DT)	31.62±9.3 (Ens)	**53.52±12.31 (Ens)**	30.83±7.31 (Ens)	44.74±4.39 (Ens)	30.69±9.79 (DA)	42.16±7.53 (Ens)	**59.55±8.04 (Ens)**	27.97±0.35 (RG)	40.86±12.62 (DT)	39.16±2.74 (Ens)	31.32±5.56 (DT)	48.49±7.92 (DT)	28.78±3.57 (RG)
EDA	28.03±1.04 (RG)	28.05±0.48 (RG)	31.5±8.52 (Ens)	29.92±9.37 (Ens)	27.96±0.25 (RG)	28.72±14.4 (KNN)	38.47±14.07 (DT)	36.99±9.46 (Ens)	34.18±13.86 (Ens)	61.22±11.65 (Ens)	27.27±0.8 (MG)	30.93±5.39 (Ens)	34.31±2.09 (Ens)	28.78±3.57 (RG)
TEMP	27.95±0.96 (RG)	28.11±0.56 (RG)	32.31±18.49 (DT)	47.19±12.8 (DT)	27.99±0.25 (RG)	31.93±21.56 (KNN)	35.95±8.38 (Ens)	30.85±21.24 (Ens)	28.87±13.56 (KNN)	30.51±7.44 (DA)	38.8±14.38 (Ens)	28.25±0.07 (RG)	34.76±17.53 (KNN)	28.74±3.58 (RG)
All	47.78±3.21 (DT)	28.14±0.39 (RG)	48.07±8.02 (KNN)	44.72±19.04 (Ens)	36.59±23.02 (KNN)	31.76±23.88 (DA)	38.43±4.25 (KNN)	44.22±12.25 (DA)	27.83±0.2 (RG)	56.72±5.13 (DA)	46.02±12.28 (Ens)	31.69±19.47 (KNN)	29.67±0.88 (RG)	28.78±3.57 (RG)
Polar H10:														
ACC	28.56±5.85 (Ens)	33.13±11.35 (KNN)	45.19±8.92 (Ens)	42.66±6.13 (Ens)	57.3±2.66 (KNN)	28.63±0.44 (RG)	50.02±5.34 (DT)	50.04±2.93 (Ens)	**48.43±4.95 (Ens)**	57.86±18.37 (DA)	48.15±5.95 (Ens)	36.23±13.59 (DT)	48.34±10.65 (DT)	41.65±4.37 (MLP)
BR	33.65±13.98 (KNN)	**45.48±6.54 (KNN)**	27.12±0.42 (RG)	31.6±5.02 (DT)	41.8±11.99 (Ens)	28.54±0.53 (RG)	41.2±7.58 (Ens)	35.85±12.21 (KNN)	27.99±0.25 (RG)	32.39±13.62 (KNN)	32.88±5.67 (KNN)	40.46±8.33 (Ens)	33.52±20.8 (DT)	28.8±3.55 (RG)
ECG	41.14±12.3 (Ens)	42.03±13.19 (Ens)	32.86±6.21 (DT)	27.09±0.44 (RG)	49.85±2.31 (DT)	39.38±7.04 (Ens)	30.23±12.35 (DT)	45.84±14.41 (Ens)	31.49±6.9 (DT)	33.57±12.09 (KNN)	37.29±5.04 (DT)	42.11±9.81 (MLP)	50.16±8.54 (Ens)	28.78±3.57 (RG)
All	34.36±6.66 (Ens)	41.36±23.6 (Ens)	47.07±10.17 (KNN)	35.92±12.46 (KNN)	51.39±1.39 (Ens)	44.21±14.71 (MLP)	51.24±9.91 (Ens)	55.42±1.95 (DT)	56.03±11.16 (Ens)	56.84±5.75 (DT)	45.38±13.91 (MLP)	42.95±6.14 (Ens)	48.47±10.98 (KNN)	37.95±8.56 (DA)
Vive Pro 2:														
EYE	33.35±16.14 (KNN)	40.14±24.43 (KNN)	29.95±19.55 (KNN)	35.52±14.63 (DT)	31.54±15.89 (KNN)	45.83±24.75 (KNN)	33.71±7.66 (DT)	27.75±0.46 (RG)	27.87±0.27 (RG)	32.07±10.39 (Ens)	29.31±14.94 (KNN)	52.24±0.89 (Ens)	35.26±27.4 (DA)	**43.77±27.04 (DA)**
All Sensors:														
ACC	28.39±14.48 (KNN)	42.04±6.28 (Ens)	34.29±12.3 (DA)	43.66±7.23 (DT)	**58.51±9.04 (KNN)**	31.19±15.37 (KNN)	**52.83±2.28 (Ens)**	50.69±1.43 (KNN)	47.63±9.58 (DT)	**60.65±17.77 (DA)**	40.82±13.79 (Ens)	37.77±11.64 (Ens)	**69.52±12.59 (DA)**	42.07±2.3 (DA)
Heart	30.87±8.65 (Ens)	41.84±9.53 (Ens)	50.12±12.92 (Ens)	38.28±14.32 (KNN)	51.71±4.26 (Ens)	41.78±14.42 (DT)	41.46±5.75 (DA)	58.08±18.51 (Ens)	37.81±3.56 (DT)	40.32±4.77 (KNN)	41.46±7.16 (Ens)	40.95±14.59 (KNN)	50.65±11.51 (DT)	38.17±10.07 (DA)
All	38.9±11.15 (Ens)	37.86±8.75 (KNN)	37.74±14.28 (Ens)	50.33±7.57 (Ens)	44.47±6.35 (DT)	36.27±21.99 (DA)	49.31±9.8 (KNN)	47.33±10.87 (KNN)	42.65±12.12 (DT)	58.88±8.5 (DT)	**50.09±9.99 (DA)**	**53.42±13.36 (DA)**	51.69±10.66 (Ens)	28.88±3.44 (RG)

Table 9.21 Balanced accuracy and F1-score of the best personalized classifiers on the three levels of fall risk (low vs. medium vs. high), for each subject and repetition - Test.

		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	
Slip	bACC	58.26	65.02	39.77	33.45	64.98	48.41	48.98	79.68	82.14	72.48	44.27	67.96	83.81	49.55	59.91±16.28
	F1-score	55.41	53.02	39.50	42.17	64.15	42.06	50.23	53.36	68.62	69.62	52.91	37.99	83.90	72.69	56.12±13.89
Trip	bACC	44.78	42.83	74.01	53.74	48.77	57.24	54.76	74.66	66.37	80.33	50.97	59.35	65.41	71.41	60.33±11.87
	F1-score	44.01	42.03	58.48	51.68	48.09	57.32	54.53	68.64	52.07	66.12	58.37	61.27	64.16	39.50	54.73±9.02
Fall	bACC	63.13	45.04	55.88	78.30	84.03	61.85	52.81	70.90	33.69	71.19	66.67	67.33	75.18	48.39	62.46±13.96
	F1-score	46.15	41.39	62.57	57.14	63.29	59.27	53.71	56.65	47.38	47.92	38.99	61.01	60.48	19.11	51.08±12.15

The binary classification task (Low vs. High risk) demonstrated the strongest performance for the personalized approach. The task simplification significantly improved robustness against internal noise and subtle variations between repetitions. During validation, performance remained exceptionally high (Tables 9.22 to 9.23), with the personalized models reaching an average balanced accuracy of 99.42 $\pm$ 0.35% (F1-score 98.57 $\pm$ 1.30%) for the Trip scenario (Table 9.24), confirming maximum model fitness to the individual data distribution.

The test set results (Tables 9.25 to 9.26) confirmed the viability of this personalized binary approach. The best models, typically Multilayer Perceptron (MLP) and Discriminant Analysis (DA), achieved highly competitive real-world test results (Table 9.27):

- **Fall Scenario:** Average balanced accuracy of $90.39 \pm 12.81\%$ and F1-score of $85.14 \pm 15.46\%$.
- **Trip Scenario:** Average balanced accuracy of $89.83 \pm 8.96\%$ and F1-score of $82.94 \pm 16.15\%$.
- **Slip Scenario:** Average balanced accuracy of $85.13 \pm 14.57\%$ and F1-score of $76.99 \pm 18.97\%$.

This level of performance represents a substantial leap over both the personalized three-class results and the generalized LOSO binary results (with a best balanced accuracy of 67%), providing strong evidence that fall-risk estimation is not an entirely generalizable problem, but one that requires a foundation of subject-specific modeling. When considering binary, personalized risk classification, the framework demonstrates sufficient robustness and accuracy for potential real-world deployment, provided a subject-specific calibration and training period is implemented upon initial setup. The modalities that showed particular efficacy in this personalized setting include the Empatica E4's BVP and Polar H10 signals (ECG and all the features) and the overall fusion of all available sensors.

Table 9.22 Balanced accuracy of the provided modalities and personalized classifiers on the two levels of fall risk (low vs. high) for each subject - Validation.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14
Empatica E4:														
ACC	96.57±3.93 (KNN)	97.3±3.37 (DT)	92.49±6.45 (DT)	97.89±1.86 (DT)	**97.67±3.62 (KNN)**	50±0 (MG)	92.17±4.3 (DT)	99.03±1.21 (DT)	**99.27±0.51 (DT)**	95.18±4.17 (KNN)	94.37±4.56 (DT)	98.91±0.86 (KNN)	50±0 (Ens)	97.6±2.94 (DT)
BVP	89.41±11.4 (DA)	**99.75±0.08 (KNN)**	50.18±0.14 (RG)	93.14±8.7 (DT)	90.19±6.85 (MLP)	99.41±0.59 (KNN)	94.07±1.52 (DT)	50.02±0.22 (RG)	96.86±3.55 (DT)	94.27±5.18 (DA)	96.48±5.77 (KNN)	91.57±7.2 (DA)	**99.71±0.39 (KNN)**	98.66±0.97 (DT)
EDA	65.6±27.01 (Ens)	90.76±9.58 (MLP)	88.24±8.12 (DA)	62.65±17.71 (DA)	87.05±17.69 (MLP)	93.36±2.06 (DA)	92.94±5.78 (DA)	95.63±2.74 (KNN)	72.03±13.53 (DA)	65.74±27.27 (Ens)	83.27±5.5 (DA)	99.15±0.32 (DT)	86.68±11.68 (DA)	90.39±10.94 (DA)
TEMP	83.53±4.45 (DA)	63.01±0.54 (DA)	62.7±16.75 (MLP)	50±0 (Ens)	84.09±13.05 (MLP)	68.91±8.27 (DA)	86.36±9.01 (DA)	72.05±13.15 (DA)	91.46±4.61 (DA)	90.13±7.96 (DA)	50±0 (MG)	92.88±4.45 (DA)	90.64±1.44 (DA)	84.34±8.38 (DA)
All	50±0 (MG)	94.71±2.97 (DA)	50±0 (Ens)	50.02±0.01 (RG)	98.48±1.5 (DA)	50±0 (MG)	98.05±1.72 (DA)	50.15±0.11 (RG)	90.69±5.97 (DA)	50±0.04 (RG)	92.98±11.24 (DA)	93.92±6.06 (DA)	50±0 (Ens)	50±0 (MG)
Polar H10:														
ACC	96.03±3.35 (DT)	95.5±3.79 (DT)	80.82±17.4 (DA)	80.13±10.95 (DA)	50±0 (Ens)	**99.45±0.65 (DA)**	50.12±0.13 (RG)	50.02±0.22 (RG)	50.15±0.13 (RG)	50±0.04 (RG)	50±0 (MG)	94.25±6.54 (DA)	50±0 (MG)	96.15±2.98 (DT)
BR	71.34±13.59 (DA)	50.02±0.24 (RG)	85.72±6.44 (DA)	76.08±7.5 (MLP)	93.16±2.72 (MLP)	85.27±10.74 (DA)	88.07±3.85 (DA)	**99.66±0.14 (DT)**	76.82±7.53 (DA)	61.6±13.66 (MLP)	85.74±5.09 (DA)	71.18±17.24 (DA)	86.4±10.91 (DA)	77.95±10.49 (DA)
ECG	89.95±8.37 (DA)	97.82±2 (DA)	94.75±5.27 (DA)	99.19±1.26 (KNN)	50±0.02 (RG)	50±0 (Ens)	99.7±0.1 (DT)	92.72±8.57 (DA)	92.54±8.02 (DA)	50±0 (MG)	**99.33±0.8 (DT)**	99.16±0.51 (DT)	88.84±10.42 (DA)	98.56±1.68 (DT)
All	92.55±6.75 (DA)	50±0 (MG)	96.92±2.84 (DA)	50±0 (MG)	50±0.02 (RG)	66.54±28.65 (Ens)	50±0 (MG)	66.42±28.44 (Ens)	50±0 (Ens)	50±0 (MG)	66.65±28.84 (Ens)	99.23±0.5 (DA)	50.01±0.12 (RG)	98.21±1.88 (DA)
Vive Pro 2:														
EYE	50±0 (MG)	93.72±7.93 (DA)	94.68±4.6 (DA)	**99.56±0.47 (DA)**	86.3±12.74 (DA)	94±2.71 (DA)	81.12±13.28 (DA)	91.74±4.72 (DA)	50±0 (MG)	91.25±9.43 (MLP)	77.47±23.05 (DA)	97.49±1.32 (DA)	99.61±0.22 (DT)	97.44±4.08 (DT)
All Sensors:														
ACC	**99.38±0.34 (KNN)**	99.36±0.18 (KNN)	**99.12±1.27 (DT)**	50.02±0.01 (RG)	50.01±0.01 (RG)	50.02±0.02 (RG)	50±0 (MG)	50.14±0.14 (RG)	50.15±0.13 (RG)	**99.6±0.45 (DA)**	50±0 (MG)	**99.72±0.1 (DA)**	50±0 (Ens)	**99.3±0.79 (KNN)**
Heart	50.14±0.12 (RG)	99.16±0.7 (DA)	50±0 (MG)	98.7±0.92 (KNN)	50.01±0.01 (RG)	99.41±0.21 (DA)	**99.87±0.09 (DT)**	66.65±28.84 (Ens)	88.56±13.66 (DA)	50±0 (MG)	98.62±0.6 (DA)	99.37±0.95 (KNN)	98.25±1.52 (DA)	93.52±3.3 (DA)
All	50±0 (MG)	50±0 (MG)	50±0 (Ens)	50±0.02 (RG)	50±0 (MG)	83.29±28.83 (Ens)	50.14±0.11 (RG)	50±0 (Ens)	50±0 (Ens)	50±0.04 (RG)	50.01±0.15 (RG)	50±0.22 (RG)	50.09±0.05 (RG)	96.98±2.59 (DA)

Table 9.23 F1-score of the provided modalities and personalized classifiers on the two levels of fall risk (low vs. high) for each subject - Validation.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14
Empatica E4:														
ACC	90.18±10.46 (KNN)	89.97±11.97 (DT)	89.15±11.7 (Ens)	94.52±2.82 (Ens)	**95.45±7.33 (Ens)**	47.04±0.37 (MG)	83.82±4.06 (Ens)	95.02±5.52 (DT)	**97.98±1.37 (DT)**	94.35±6.84 (Ens)	85.15±11.77 (DT)	95.45±3.69 (KNN)	44.73±1.91 (RG)	97.51±2.88 (DT)
BVP	85.19±3.96 (DA)	**98.7±0.17 (KNN)**	47.3±0.37 (MG)	84.46±14.73 (DT)	78.5±13.16 (MLP)	**98.1±2.83 (KNN)**	91.13±3.53 (Ens)	47.84±0.44 (MG)	90.09±10 (DT)	93.47±3.99 (DA)	92.64±11.19 (Ens)	80.31±17.08 (DA)	**99.21±1.11 (KNN)**	98.56±1.02 (DT)
EDA	63.66±28.99 (Ens)	77.3±12.95 (MLP)	85.31±11.61 (DA)	54.86±19.76 (DA)	73.2±31.32 (MLP)	79.38±3.1 (DA)	77.8±12.47 (DA)	86.91±10 (Ens)	63.91±22.97 (DA)	57.85±34.99 (Ens)	81.29±8.2 (DA)	97.59±1.87 (DT)	76.7±12.59 (DA)	90.27±10.77 (DA)
TEMP	62.96±7.23 (DA)	52.41±1.34 (DA)	54.32±18.77 (DA)	47.46±0.17 (Ens)	66.15±23.6 (MLP)	50.07±5.05 (DA)	66.05±16.62 (DA)	50.06±12.45 (DA)	76.08±10.04 (DA)	82.88±9.94 (DA)	46.53±0.48 (MG)	78.65±12.27 (DA)	80.56±5.42 (DA)	85.18±9.11 (DA)
All	46.95±0.12 (MG)	80.78±5.52 (DA)	47.3±0.37 (Ens)	47.46±0.17 (MG)	93.77±5.47 (DA)	47.04±0.37 (MG)	91.64±7.19 (DA)	47.84±0.44 (MG)	87.05±7.37 (DA)	46.16±5.56 (RG)	93.91±6.99 (DA)	90.92±7.09 (DA)	44.75±1.92 (RG)	49.05±1.2 (RG)
Polar H10:														
ACC	89.82±8.43 (DT)	94.37±6.64 (Ens)	74.77±15.48 (DA)	61.78±5.18 (DA)	47.3±0.06 (Ens)	97.82±2.32 (DA)	47.52±0.04 (MG)	47.84±0.44 (MG)	47.03±0.19 (Ens)	46.16±5.56 (RG)	46.53±0.48 (MG)	83.4±15.34 (DA)	44.7±1.94 (RG)	96.03±3.3 (Ens)
BR	52.83±15.73 (DA)	47.59±0.48 (MG)	68.48±7.71 (DA)	55.22±2.14 (MLP)	79.57±7.8 (MLP)	69.7±9.76 (DA)	68.43±3.97 (DA)	**98±0.81 (DT)**	56.56±16.8 (DA)	57.99±1.84 (DA)	67.59±9.99 (DA)	59.43±12.12 (DA)	77.2±17.79 (DA)	76.63±9.39 (DA)
ECG	77.61±12.63 (DA)	94.3±7.2 (DA)	85.95±7.64 (DA)	96.27±5.8 (KNN)	47.3±0.06 (MG)	47.04±0.37 (Ens)	98.44±0.52 (DT)	80.54±12.35 (DA)	82.19±16.98 (DA)	46.13±5.54 (RG)	**98.55±2.05 (DT)**	96.51±1.87 (DT)	86.46±9.29 (DA)	98.52±1.65 (DT)
All	82.74±14.56 (DA)	47.59±0.48 (MG)	89.82±6.17 (DA)	47.46±0.17 (MG)	47.3±0.06 (Ens)	64.33±29.6 (Ens)	47.52±0.04 (MG)	63.35±27.27 (Ens)	47.03±0.19 (Ens)	46.16±5.55 (RG)	64.29±30.78 (Ens)	96.63±2.31 (DA)	44.76±1.95 (RG)	98.14±2.29 (DA)
Vive Pro 2:														
EYE	46.95±0.12 (MG)	84.78±11.08 (DA)	83.32±11.25 (DA)	**97.92±2.07 (DA)**	80.57±13.78 (DA)	84.22±12.47 (DA)	68.21±9.94 (DA)	83.13±2.81 (DA)	47.03±0.19 (MG)	89.3±7.39 (MLP)	75.76±27.42 (DA)	90.54±3.28 (DA)	99.12±0.46 (DT)	97.34±4.1 (DT)
All Sensors:														
ACC	**97.41±1.44 (KNN)**	98.2±2.15 (KNN)	**96.45±4.8 (DT)**	47.46±0.17 (MG)	47.3±0.06 (Ens)	47.04±0.37 (MG)	47.52±0.04 (MG)	47.84±0.44 (MG)	47.03±0.19 (Ens)	**99.49±0.29 (DA)**	46.53±0.48 (MG)	**98.75±0.43 (DA)**	44.73±1.98 (RG)	**99.27±0.77 (KNN)**
Heart	46.95±0.12 (MG)	98.31±0.69 (DA)	47.3±0.37 (MG)	93.96±4.16 (KNN)	47.3±0.06 (MG)	97.49±0.59 (DA)	**99.31±0.46 (DT)**	65.27±29.87 (Ens)	87.1±10.29 (DA)	46.13±5.54 (RG)	95.79±3.64 (DA)	97.29±4 (KNN)	95.59±3.8 (DA)	93.04±2.89 (DA)
All	46.95±0.12 (MG)	47.59±0.48 (MG)	47.3±0.37 (Ens)	47.46±0.17 (Ens)	47.3±0.06 (MG)	82.19±30.39 (Ens)	47.52±0.04 (MG)	47.84±0.44 (Ens)	47.03±0.19 (Ens)	46.16±5.56 (RG)	46.53±0.48 (MG)	47.11±0.37 (MG)	44.8±1.93 (RG)	96.54±3.02 (DA)

Table 9.24 Balanced accuracy and F1-score of the best personalized classifiers on the two levels of fall risk (low vs. high), for each subject and repetition - Validation.

		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	
Slip	bACC	99.46	99.66	97.65	99.95	96.75	99.95	99.92	99.81	98.72	99.80	99.89	99.68	99.95	98.42	99.26±0.97
	F1-score	97.71	98.52	90.95	99.75	86.99	94.83	99.56	98.92	96.41	99.77	99.63	98.72	99.83	98.45	97.14±3.68
Trip	bACC	99.00	99.82	99.76	99.05	99.33	98.72	99.92	99.53	99.36	99.09	99.69	99.84	99.26	99.52	99.42±0.35
	F1-score	95.84	98.83	98.68	95.67	99.59	99.75	99.59	97.66	98.80	99.19	99.82	99.19	97.93	99.39	98.57±1.30
Fall	bACC	99.67	99.77	99.94	99.69	99.36	99.68	99.77	99.65	99.73	99.91	98.41	99.65	99.93	99.97	99.65±0.38
	F1-score	98.66	98.75	99.73	98.33	99.77	99.73	98.78	97.42	98.74	99.51	96.19	98.35	99.87	99.96	98.84±1.03

Table 9.25 Balanced accuracy of the provided modalities and personalized classifiers on the two levels of fall risk (low vs. high) for each subject - Test.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14
Empatica E4:														
ACC	50±0 (MG)	66.06±15.04 (DA)	50±0 (MLP)	68.49±5.98 (DA)	69.02±32.88 (Ens)	50.01±0.06 (RG)	50±0 (MG)	83.91±2.18 (KNN)	64.77±13.2 (MLP)	85.82±1.98 (MLP)	**84.58±3.87 (KNN)**	78.51±32.81 (KNN)	85.54±17.04 (MLP)	50.01±0.06 (RG)
BVP	**77.86±24.29 (MLP)**	51.25±2.17 (KNN)	**87.59±18.88 (KNN)**	**75.86±22.2 (MLP)**	57.68±7.95 (MLP)	78.96±17.16 (DA)	62.15±16.31 (DA)	66.63±28.81 (Ens)	51.06±33.53 (MLP)	75.18±27.66 (DA)	68.36±26.49 (DA)	83.45±1.31 (KNN)	50.1±0.09 (RG)	57.36±13.59 (DA)
EDA	62.78±22.13 (Ens)	75.19±24.95 (KNN)	64.98±11.15 (KNN)	50±0 (DT)	76.2±22.79 (Ens)	56.15±12.31 (KNN)	55.14±8.61 (DA)	53.93±13.47 (DT)	64.57±12.96 (MLP)	84.99±14.04 (KNN)	69.72±18.69 (KNN)	52.61±11.99 (KNN)	**91.61±9.79 (DT)**	65.67±13.61 (MLP)
TEMP	59.31±16.12 (MLP)	50.15±0.26 (KNN)	50.22±0.34 (RG)	68.42±19.02 (KNN)	50.01±0.02 (RG)	50.04±0.03 (RG)	50.55±0.96 (DA)	50.12±0.26 (RG)	**74.76±10.31 (DA)**	66.54±28.65 (DA)	57.97±33.18 (DA)	76.95±24.51 (DT)	78.45±24.66 (DT)	50.01±0.06 (RG)
All	58.97±15.53 (MLP)	58.82±12.47 (DA)	77.97±23.51 (KNN)	54.47±5.88 (DA)	75.89±9.75 (KNN)	77.51±22.02 (DA)	81.05±26.93 (MLP)	**99.42±0.28 (DA)**	53.98±47.22 (MLP)	50.01±0.02 (RG)	60.85±28.43 (MLP)	53.5±16.62 (KNN)	82.18±27.89 (MLP)	60.74±10.43 (DA)
Polar H10:														
ACC	50.13±0.22 (RG)	50±0 (MLP)	80.48±26.45 (Ens)	58.01±17.4 (Ens)	50±0 (Ens)	50±0 (Ens)	**93.84±3.96 (DA)**	50.12±0.26 (RG)	50.1±0.27 (RG)	91.72±11.74 (MLP)	64.49±25.09 (KNN)	65.21±26.34 (KNN)	67.02±14.74 (KNN)	67.68±15.32 (MLP)
BR	67.77±37.47 (KNN)	60.12±17.53 (Ens)	64.87±21.19 (DA)	50±0 (MG)	55.4±14.4 (DA)	55.67±22.07 (MLP)	80.22±16.5 (DA)	73.53±19.17 (DA)	55.98±16.85 (KNN)	59.79±13.78 (DT)	50.1±0.26 (RG)	50±0 (MG)	50.01±0.16 (RG)	58.33±10.69 (KNN)
ECG	65.34±17.49 (MLP)	**86.48±10.72 (KNN)**	75.31±23.66 (Ens)	54.23±7.93 (KNN)	85.58±11.56 (DA)	**93.64±8.27 (DA)**	80.84±27.47 (DA)	70.69±24.58 (DA)	61.64±35.87 (DA)	79.94±7.12 (DT)	75.46±23.89 (Ens)	60.04±12.18 (MLP)	66.68±17.95 (MLP)	74.68±2.09 (MLP)
All	67.57±22.75 (DT)	78.73±12.21 (MLP)	62.96±32.05 (DA)	55.78±12.55 (KNN)	**94.62±4.66 (MLP)**	50.01±0.06 (RG)	93.56±4.46 (KNN)	80.29±26.42 (DA)	69.18±5.68 (KNN)	**92.72±9.92 (KNN)**	66.11±27.91 (DT)	65.91±27.56 (MLP)	75.67±23.66 (MLP)	72.9±25.72 (KNN)
Vive Pro 2:														
EYE	58.7±22.7 (DT)	54.99±17.6 (Ens)	79.37±25.65 (DT)	57.91±13.71 (Ens)	54.66±30.62 (KNN)	77.14±21.8 (DA)	54.37±7.56 (MLP)	54.46±4.01 (DA)	59.49±20.01 (Ens)	68.63±25.9 (KNN)	50±0 (DA)	67.04±20.46 (MLP)	50±0 (MG)	**86.93±10.22 (DT)**
All Sensors:														
ACC	50±0 (MG)	63.29±16.72 (MLP)	71.47±28.87 (KNN)	52.56±9.8 (MLP)	50.01±0.02 (RG)	50±0 (MG)	90.87±2.45 (DA)	96.95±3.71 (KNN)	50.13±0.25 (RG)	90.11±4.44 (DT)	76.99±24.95 (MLP)	**98.41±0.73 (DA)**	84.31±12.43 (KNN)	67.67±10.12 (DA)
Heart	67.39±22.18 (DA)	68.18±2.29 (KNN)	64.1±31.4 (KNN)	63.21±14.15 (KNN)	85.29±9.36 (DA)	90.45±15.51 (KNN)	81.56±25.99 (DT)	64.55±25.19 (Ens)	58.69±11.35 (MLP)	86.55±11.69 (KNN)	68.33±7.17 (MLP)	78.22±17.66 (DA)	50.01±0.16 (RG)	62.35±17 (DA)
All	67.88±19.69 (DT)	61.89±10.3 (DT)	68.41±27.38 (MLP)	74.66±23.3 (MLP)	89±18.34 (KNN)	50±0 (MG)	90.5±2.44 (DA)	88.92±17.66 (KNN)	50.14±0.24 (RG)	78.21±24.48 (DT)	65.61±14.74 (Ens)	67.95±26.1 (MLP)	60.22±13.03 (KNN)	78.96±7.19 (DA)

Table 9.26 F1-score of the provided modalities and personalized classifiers on the two levels of fall risk (low vs. high) for each subject - Test.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14
Empatica E4:														
ACC	46.95±0.24 (MG)	58.06±17.51 (Ens)	48.92±23.55 (KNN)	48.14±21.27 (DA)	62.13±31.67 (Ens)	46.99±0.65 (MG)	47.53±0.08 (MG)	76.14±15.94 (KNN)	53.2±6.53 (Ens)	77.7±18.22 (MLP)	**78.36±19.19 (KNN)**	74.96±28.3 (KNN)	80.39±12.82 (MLP)	46.78±10.96 (MLP)
BVP	**67.22±18.21 (MLP)**	47.51±0.87 (MG)	**87.41±13.63 (KNN)**	**69.39±30.87 (DA)**	47.41±0.78 (Ens)	67.76±21.85 (Ens)	53.97±3.06 (DA)	65.4±29.62 (Ens)	51.96±21.75 (Ens)	71.2±28.38 (DA)	64.98±27.07 (DA)	69.06±10.39 (KNN)	44.66±4.07 (RG)	46.3±3.6 (RG)
EDA	62.21±26.34 (Ens)	66.78±28.34 (KNN)	52.3±4.49 (DA)	47.43±0.35 (DT)	70.31±21.69 (Ens)	46.99±0.65 (MG)	47.53±0.08 (MG)	47.74±0.98 (MG)	49.23±11.36 (DA)	78.2±22.03 (KNN)	60.6±26.63 (KNN)	53.65±11.25 (KNN)	**88.03±15.83 (DT)**	48.93±19.51 (MLP)
TEMP	47.23±0.46 (DT)	47.51±0.87 (MG)	47.29±0.74 (MG)	51.28±13.85 (KNN)	47.3±0.1 (MG)	46.99±0.65 (MG)	47.53±0.08 (MG)	47.74±0.98 (MG)	**57.92±29 (DA)**	60.88±32.99 (DA)	46.51±0.85 (MG)	67.62±22.58 (DT)	74.82±23.74 (DT)	46.3±3.6 (RG)
All	46.95±0.24 (MG)	52.1±24.99 (MLP)	70.87±26.26 (KNN)	47.43±0.35 (MG)	50.8±13.47 (KNN)	66.48±18.54 (DA)	75.96±26.17 (MLP)	**98.34±1.21 (DA)**	47.02±0.36 (MG)	43.77±6.56 (MG)	46.51±0.85 (MG)	54.1±17.49 (KNN)	80.61±28.68 (MLP)	50.63±16.45 (DA)
Polar H10:														
ACC	47.69±11.63 (DT)	47.51±0.87 (MG)	70.64±21.67 (Ens)	47.79±2.93 (Ens)	47.3±0.1 (Ens)	46.99±0.65 (Ens)	79.21±9.49 (DA)	47.74±0.98 (MG)	47.02±0.36 (Ens)	81.63±23.66 (DT)	56.73±18.45 (KNN)	56.83±17.54 (Ens)	51.62±10.77 (KNN)	60.84±12.55 (MLP)
BR	62.99±33.58 (KNN)	56.46±16.25 (Ens)	51.16±7.02 (Ens)	47.43±0.35 (MG)	48.43±23.79 (Ens)	48±3.04 (DT)	63.41±29.15 (DA)	55.49±16.6 (Ens)	57.29±21.49 (KNN)	45.15±0.53 (DT)	46.51±0.85 (MG)	47.05±0.62 (MG)	44.62±4.04 (RG)	48.22±8.24 (KNN)
ECG	57.1±19.8 (Ens)	**81.75±19.86 (KNN)**	72.26±21.66 (Ens)	48.41±1.68 (KNN)	78.32±20.91 (DA)	**83.95±16.08 (DA)**	75.76±25.19 (Ens)	66.65±19.88 (DA)	47.02±0.36 (MG)	69.57±14.77 (DT)	68.51±26.12 (Ens)	52.44±15.14 (DT)	59.73±15.84 (MLP)	65.68±13.88 (MLP)
All	60.81±15.58 (DT)	71.96±30.95 (DT)	62.31±31.54 (DA)	47.43±0.35 (MG)	**82.72±12.56 (MLP)**	46.99±0.65 (MG)	**80.3±9.45 (DA)**	77.05±24.8 (DA)	51.33±9.28 (KNN)	**89.19±9.14 (KNN)**	62.39±27.52 (DT)	60.85±24.5 (MLP)	63.06±22.74 (DA)	70.13±25.38 (KNN)
Vive Pro 2:														
EYE	51.18±11.42 (DT)	53.82±14.87 (Ens)	69.65±21.11 (DT)	47.43±0.35 (MG)	49.66±17.26 (KNN)	71.95±20.41 (DA)	49.56±3.45 (KNN)	47.74±0.98 (MG)	56.27±24.11 (KNN)	51.91±44.49 (MLP)	46.51±0.85 (MG)	59.51±21.16 (DT)	44.58±4 (RG)	**85±7.17 (DT)**
All Sensors:														
ACC	46.95±0.24 (MG)	53.45±11.99 (KNN)	58.63±21.61 (KNN)	47.43±0.35 (MG)	47.3±0.1 (Ens)	46.99±0.65 (MG)	74.46±9.69 (DA)	90.85±14.65 (KNN)	47.02±0.36 (Ens)	75.46±17.19 (DT)	67.68±28.26 (MLP)	**94.08±5.21 (DA)**	76.47±29.16 (KNN)	61.83±5.05 (DA)
Heart	59.58±21.79 (Ens)	55.35±19.85 (Ens)	63.15±31.75 (KNN)	54.35±6 (KNN)	79.55±23.74 (DA)	81.23±27.65 (KNN)	75.02±23.08 (Ens)	63.73±28.3 (Ens)	48.81±3.62 (DT)	77.26±18.09 (KNN)	64.85±11.5 (MLP)	69.41±19.12 (Ens)	44.62±4.04 (RG)	51.87±19.88 (Ens)
All	59.77±12.2 (DT)	52.75±25.36 (MLP)	67.49±27.82 (MLP)	69.38±23.11 (MLP)	79.34±35.01 (KNN)	46.99±0.65 (MG)	71.13±5.09 (DA)	80.87±30.52 (KNN)	47.02±0.36 (Ens)	71.64±22.56 (DT)	61.6±28.19 (KNN)	64.97±23.83 (MLP)	54.49±6.76 (KNN)	74.62±14.46 (DA)

Table 9.27 Balanced accuracy and F1-score of the best personalized classifiers on the two levels of fall risk (low vs. high), for each subject and repetition - Test.

		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	
Slip	bACC	94.59	79.55	65.85	50.65	89.60	97.86	97.29	99.57	68.69	97.73	80.80	97.73	94.44	77.47	85.13±14.57
	F1-score	82.86	86.47	72.55	37.46	69.78	91.76	85.40	97.10	40.17	94.25	56.51	89.14	96.07	78.31	76.99±18.97
Trip	bACC	88.98	81.08	97.04	84.48	95.44	98.95	94.70	99.09	68.92	99.13	84.40	99.18	80.71	85.56	89.83±8.96
	F1-score	71.59	59.95	90.37	99.07	83.51	94.63	86.10	99.51	42.21	94.68	86.10	99.53	69.79	84.13	82.94±16.15
Fall	bACC	50.00	98.83	99.88	92.46	98.82	84.11	89.52	99.59	86.67	81.30	88.54	98.31	99.67	97.76	90.39±12.81
	F1-score	47.23	98.83	99.32	71.65	94.86	65.46	69.40	98.40	91.38	78.64	92.46	93.59	98.22	92.57	85.14±15.46

Then, SHAP (Shapley additive explanations) was used to explain the best-performing personalized binary approaches. The contribution of each feature and the direction (positive or negative) of its effect were calculated. The SHAP analysis reported is limited to discriminant analysis models trained using signals recorded from participant 7 (P7) with the Polar H10 and accelerations from participant 12 (P12) with both the Empatica E4 and the Polar H10. Specifically, Figure 9.14 refers to the model trained using the signals from P7 during the Trip and Fall sequences; Figure 9.15 refers to the model trained using signals during the Slip and Fall sequences; and Figure 9.16 refers to the model trained using signals during the Slip and Fall sequences. Figures 9.17, 9.18, and 9.19 follow the same reasoning regarding P12.

In models using signals recorded from P7 with the Polar H10, SHAP analysis reveals features consisting of accelerations and cardio-respiratory signals. In Figure 9.14, respiratory and ECG-derived features play a more balanced role. Average breathing rate (*ABR*) shows the largest SHAP spread, with higher values contributing positively to the prediction. In contrast, $pNN50$ derived from ECG demonstrates an inverse relationship, with higher values associated with negative SHAP contributions. The median RR (*MRR*) interval contributes modestly in a positive direction, while acceleration features (*MA*, *ASD*) at Polar H10 retain a secondary but structured influence. While in Figure 9.15, both $TINN$ and average BR (*ABR*) are the dominant predictors, exhibiting predominantly positive SHAP values for higher feature magnitudes. Consistent with Figure 9.14, $pNN50$ maintains a negative association with the predicted outcome. Acceleration-based features, such as AAI_{3D}, contribute only marginally, indicating that model predictions in this setting are primarily driven by combined respiratory and heart rate variability dynamics rather than by movement features. In Figure 9.16, the ultra-low-frequency composite feature ($SULF$) emerges as the most influential feature, with higher values strongly associated with positive SHAP contributions. Additional HRV-related measures, including $TINN$ and median heart rate *MHR*, provide secondary contributions, whereas acceleration-derived features exhibit minimal impact, such as AAI_{3D}, ASD_{3D}, and AA_{3D}.

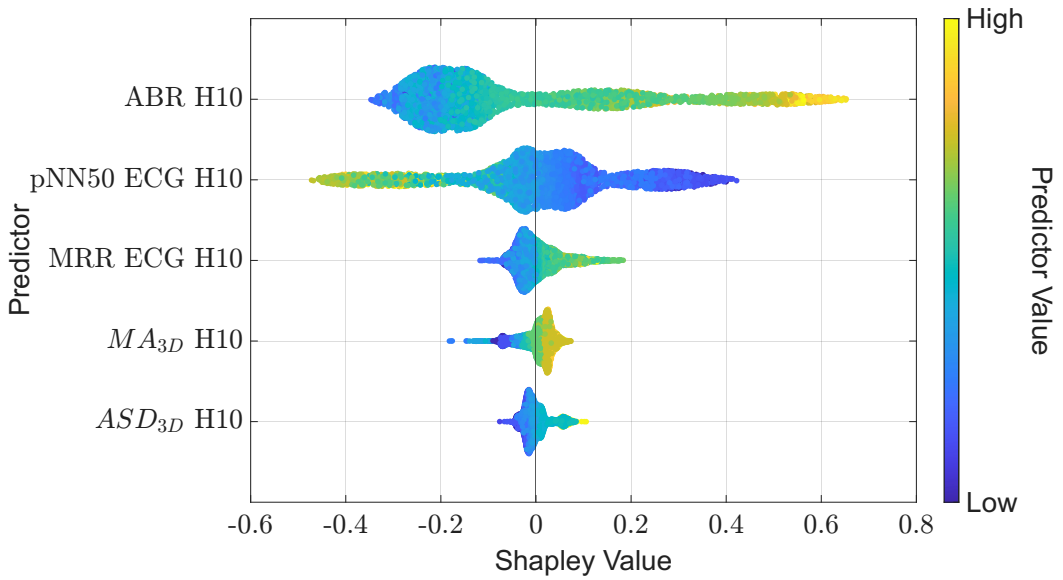


Figure 9.14 SHAP analysis of the Discriminant Analysis (DA) personalized classifier on the two levels of fall risk (low vs. high) trained using P7's signals recorded with Polar H10 during the Trip and Fall sequences, after feature selection. The variables are listed in descending order of importance, with the most important variable at the top. Legend: mean *ABR* = Average breathing rate, *pNN*50 = percentage of RR intervals that differ by more than 50 ms, *MRR*= median of RR intervals, *MA* = median magnitude acceleration, and *ASD* = standard deviation of magnitude acceleration.

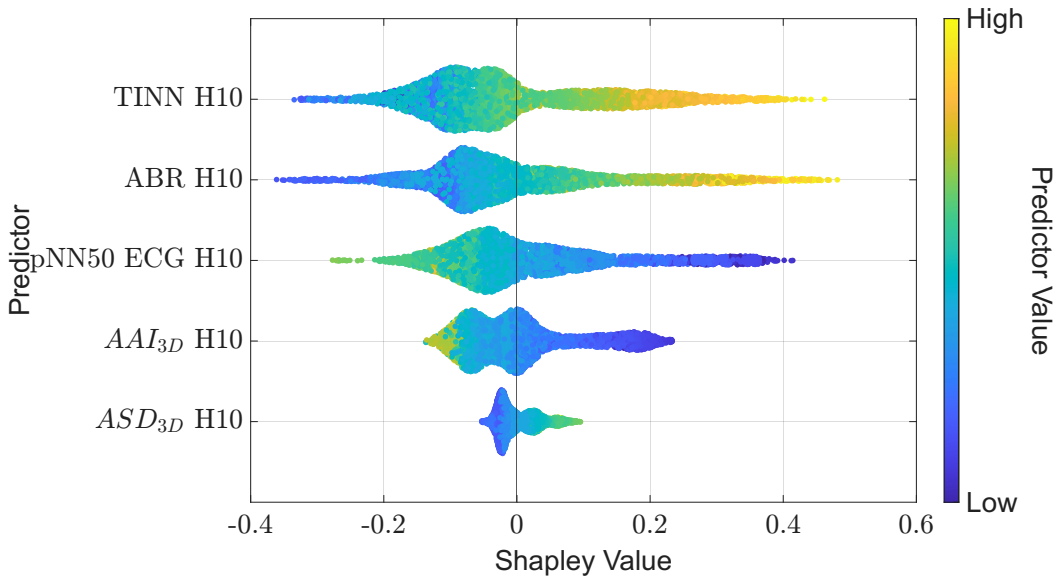


Figure 9.15 SHAP analysis of the Discriminant Analysis (DA) personalized classifier on the two levels of fall risk (low vs. high) trained using P7's signals recorded with Polar H10 during the Slip and Fall sequences, after feature selection. The variables are listed in descending order of importance, with the most important variable at the top. Legend: mean $TINN$ = triangular interpolation of RR intervals, ABR = average breathing rate, $pNN50$ = percentage of RR intervals that differ by more than 50 ms, AAI_{3D} = absolute integral of magnitude acceleration, and ASD = standard deviation of magnitude acceleration.

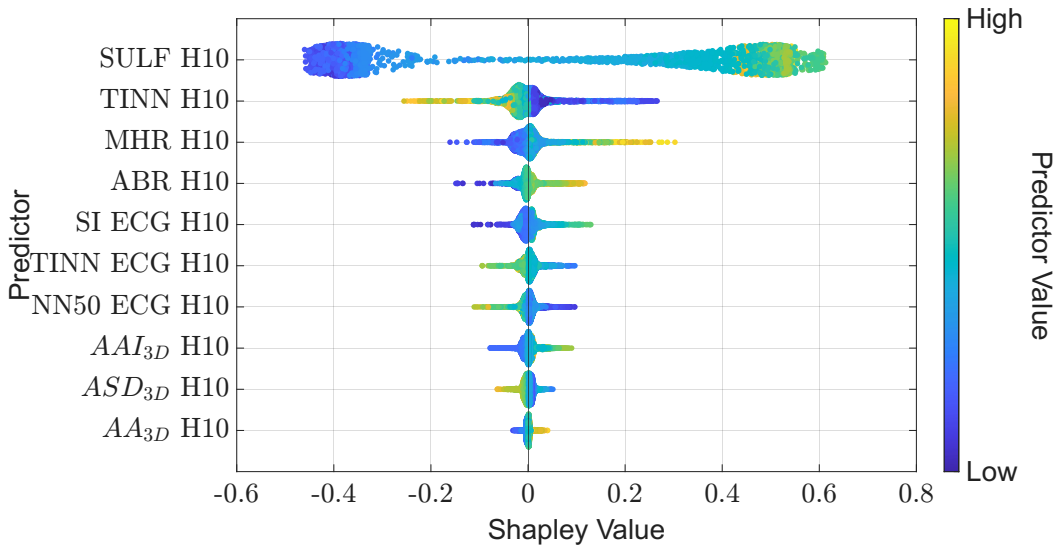


Figure 9.16 SHAP analysis of the Discriminant Analysis (DA) personalized classifier on the two levels of fall risk (low vs. high) trained using P7's signals recorded with Polar H10 during the Slip and Trip sequences, after feature selection. The variables are listed in descending order of importance, with the most important variable at the top. Legend: $SULF$ = sum of ULF frequency components of RR intervals, $TINN$ = triangular interpolation of RR intervals, MHR = median HR, AHR = average HR, SI = Baevsky stress index, AAI_{3D} = absolute integral of magnitude acceleration, ASD = standard deviation of magnitude acceleration, and AA = average magnitude acceleration.

Across models based on accelerations from P12 recorded with both the Empatica E4 and the Polar H10, SHAP summary plots indicate that features describing movement magnitude and variability are the dominant contributors to model predictions. In Figure 9.17, the importance of the magnitude of the acceleration (AAI_{3D}) recorded with the Polar H10 remains prominent, although variability- and magnitude-based features from both sensor locations contribute more evenly. In particular, ASD_{3D} of Polar H10 and AAI_{3D} of Empatica E4 show substantial SHAP spreads, suggesting that both overall movement intensity and its variability are informative. The median acceleration MA_{3D} of Empatica E4 displays a comparatively narrow SHAP range, indicating a smaller but stable contribution. While in Figure 9.18, model predictions are driven almost exclusively by variability-related features. Both AAI_{3D} and ASD_{3D} of Polar H10 show strong positive SHAP contributions for higher values, whereas AA_{3D} of Empatica E4 exhibits negligible influence, with SHAP values tightly clustered around zero. Overall, ACC-based models rely primarily on movement intensity and variability, particularly with the Polar H10 sensor, rather than on central tendency measures. In Figure 9.19, the absolute acceleration AAI_{3D} of Polar H10 shows the greatest SHAP value dispersion, with higher feature values consistently associated with positive SHAP contributions, indicating a higher predicted outcome. The standard deviation of acceleration of Polar H10 ASD_{3D} exhibits a similar but weaker effect, while the average acceleration AA_{3D} of Empatica E4 contributes more moderately, with SHAP values distributed symmetrically around zero.

Therefore, the SHAP analysis reveals that the most relevant characteristics associated with the risk of falling for P7 are the average respiratory rate (*ABR*), the number/percentage of RR intervals that differ by more than 50 ms (*NN*50, *pNN*50), and the standard deviation of acceleration amplitude (ASD_{3D}). The SHAP analysis reveals that the most relevant characteristics associated with the risk of falling for P12 are the absolute integral (AAI_{3D}) and standard deviation (ASD_{3D}) of the acceleration amplitude of Polar H10 and the average/median (AA_{3D}, MA_{3D}) of the acceleration amplitude of Empatica E4.

9.6 Discussion

The discussion presented in this chapter consolidates the results obtained from two main data collection efforts: the laboratory study of activities of daily living (ADL) using multisensory wearable devices and the virtual reality (VR)-based environmental assessment for high fall risk scenarios.

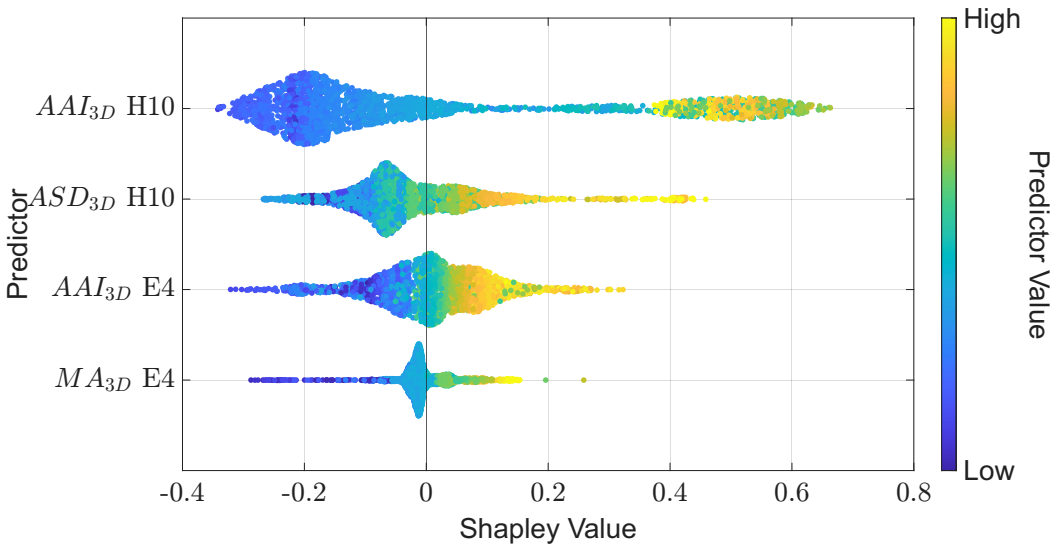


Figure 9.17 SHAP analysis of the Discriminant Analysis (DA) personalized classifier on the two levels of fall risk (low vs. high) trained using P12's acceleration signals recorded with Empatica E4 and Polar H10 during the Trip and Fall sequences, after feature selection. The variables are listed in descending order of importance, with the most important variable at the top. Legend: AAI_{3D} = absolute integral of magnitude acceleration, *ASD* = standard deviation of magnitude acceleration, and *MA* = median magnitude acceleration.

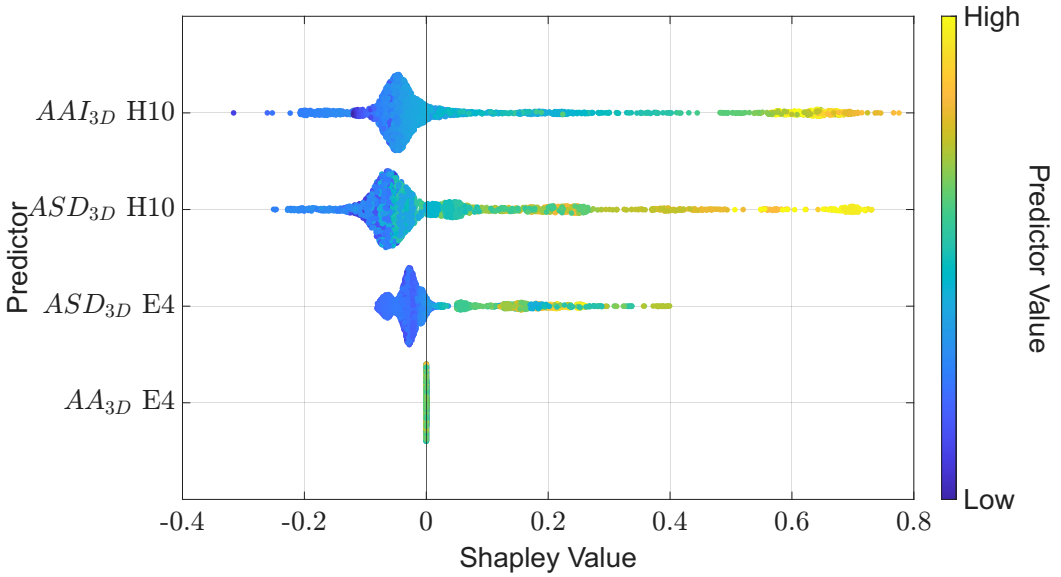


Figure 9.18 SHAP analysis of the Discriminant Analysis (DA) personalized classifier on the two levels of fall risk (low vs. high) trained using P12's acceleration signals recorded with Empatica E4 and Polar H10 during the Slip and Fall sequences, after feature selection. The variables are listed in descending order of importance, with the most important variable at the top. Legend: AAI_{3D} = absolute integral of magnitude acceleration, *ASD* = standard deviation of magnitude acceleration, and *AA* = average magnitude acceleration.

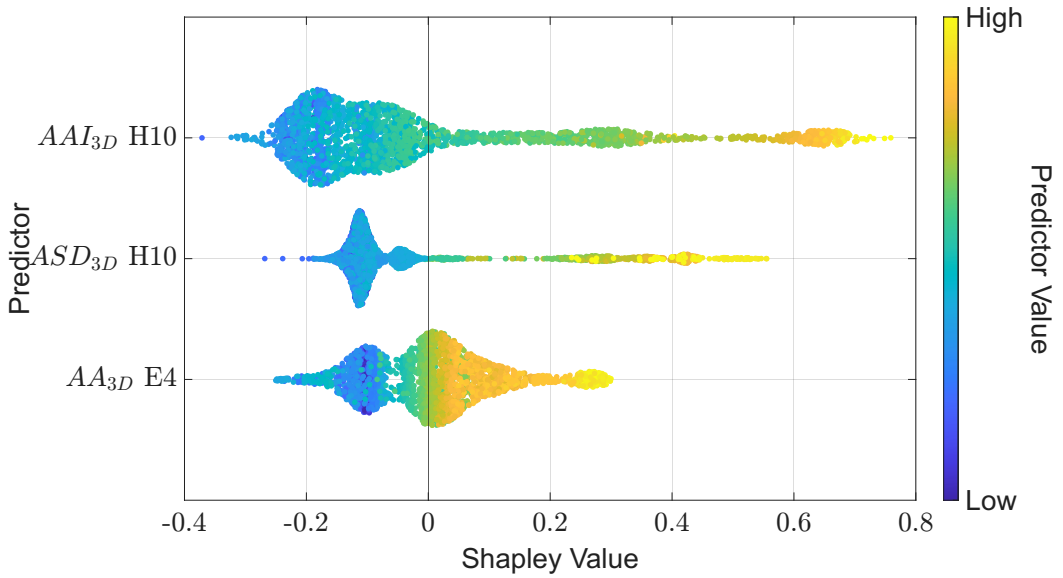


Figure 9.19 SHAP analysis of the Discriminant Analysis (DA) personalized classifier on the two levels of fall risk (low vs. high) trained using P12's acceleration signals recorded with Empatica E4 and Polar H10 during the Slip and Trip sequences, after feature selection. The variables are listed in descending order of importance, with the most important variable at the top. Legend: AAI_{3D} = absolute integral of magnitude acceleration, ASD = standard deviation of magnitude acceleration, and AA = average magnitude acceleration.

9.6.1 Activities of Daily Living

The findings successfully demonstrate that increasing speed leads to increased muscle activation in a near-linear fashion, and that high- and low-intensity activities elicit distinctive and characteristic patterns in heart and respiration rates. The results strongly suggest the potential utility of these biosensors for recognizing biomedical-based activities. The newly compiled dataset, along with these initial results, provides a strong foundation for developing reactive and proactive algorithms that can define and estimate levels of muscular fatigue, cognitive fatigue, and stress. Such algorithms could eventually assess the risk of an imminent fall before it occurs, enhancing personal safety.

9.6.2 VR-based environmental assessment study

The intrinsic motivation inventory results suggested that the participants progressively found the assembly activities more important and enjoyable, although they felt slightly pressured and lacked complete confidence in their competence. The results, in terms of presence, indicated that participants generally felt immersed in the virtual environment; however, lower spatial presence, involvement, and realism scores suggested that some elements reminded

them that they were in a simulation. For user experience, the participants found the virtual environment to be generally satisfactory and efficient; however, its effectiveness in helping users complete tasks successfully could be improved.

Regarding motion sickness, most participants experienced mild to moderate fatigue symptoms, and issues like eye strain and visual discomfort were common, indicating potential areas for improvement in user comfort. Usability results suggested that participants initially found the tasks easier and more satisfying; however, their ease, satisfaction, and efficiency slightly decreased as the experiment progressed. The task completion time results revealed improved task efficiency across activities, indicating a stable learning curve. Finally, the error report results indicated that the VR environment demonstrated high system accuracy with minimal crashes.

9.6.3 Fall-risk Estimation

The challenge of generalizing fall-risk classification models across subjects was evident in the low performance of the Leave-One-Subject-Out (LOSO) cross-validation. In the three-class problem (Low vs. Medium vs. High risk), test performance was often only slightly better than a random or majority classifier. This strongly suggests that a generalized, one-size-fits-all model cannot reliably capture the inter-subject variability in physiological and kinematic responses to fall risk.

Conversely, the personalized classification approach, trained and tested exclusively on individual subject data (k-fold cross-validation), demonstrated substantial and meaningful predictive power. For the most critical binary classification task (Low vs. High risk), personalized models achieved an average balanced accuracy of 90.39% for the Fall scenario, 89.83% for the Trip scenario, and 85.13% for the Slip scenario on the test set. This high level of performance indicates that the multimodal feature set (especially combining Polar H10 ACC/ECG and complete sensor fusion) successfully captures subject-specific signatures of high fall risk. This shift from generalized to personalized modeling is a crucial finding, suggesting that the path to real-world deployment of proactive fall risk systems lies in requiring an initial, personalized calibration period for each user.

9.7 Conclusions

This chapter comments on the findings of the comprehensive framework for estimating and proactively preventing fall risk, shifting the focus from reactive detection of an event to the anticipation of high-risk states.

9.7.1 Activities of Daily Living

The initial analysis of the XoADL dataset demonstrates the potential of wearable biosignals (EMG, HR, BR) for human activity recognition and the estimation of physical/cognitive states, confirming that different activity intensities elicit distinctive and predictable physiological patterns. These results provide the necessary groundwork for developing features in proactive algorithms that aim to predict fall risk based on fatigue and stress precursors.

9.7.2 VR-based environmental assessment study

The evaluation confirmed that the proposed VR environment successfully supported immersive and engaging user experiences suitable for fall-risk assessment and prevention training. Participants found the activities to be meaningful and enjoyable, although moderate decreases in usability and satisfaction suggest potential cognitive or visual fatigue with prolonged use. Presence scores confirmed a convincing sense of immersion, albeit with room for improvement in realism and spatial coherence.

Overall, the results highlight the system's technical robustness, user acceptability, and potential for realistic fall-risk evaluation, while identifying key areas for future enhancement, particularly in interaction accuracy, visual fidelity, and fatigue mitigation.

9.7.3 Fall-risk Estimation

The fall-risk estimation analysis concluded that generalized machine learning models (LOSO) fail to reliably classify fall-risk levels across unseen subjects due to high inter-subject variability. Conversely, the transition to a personalized classification framework (k-fold per subject) proved highly effective, particularly for the binary risk assessment (Low vs. High). This individualized approach, utilizing a rich set of multimodal features, yielded a high average balanced accuracy of over 85% on the test set for the most dangerous scenarios. This robust performance validates the novel XoFallVR dataset and the proposed multimodal framework for personalized, proactive fall-risk estimation in high-risk environments.

Part III

Conclusions

Chapter 10

Discussion

This chapter synthesizes the core findings presented in Part II, evaluating the interconnections among the developed reactive (detection), proactive (forecasting), and predictive (risk estimation) components of the fall safety framework. This discussion aims to contextualize the scientific merit and practical implications of the work in relation to the persistent challenges identified in Chapter 1.2.

10.1 Summary

The research successfully addressed the fundamental problem of fall safety by transitioning the focus from retrospective analysis to continuous, personalized, and proactive intervention.

Chapter 7 validated a Deep Learning architecture (FDNN) on the SisFall dataset, achieving a high True Negative Rate (TNR) of 94.4% across Activities of Daily Living (ADLs). This high specificity is crucial for user adoption in wearable systems, as it minimizes the occurrence of false alarms. However, the model's True Positive Rate (TPR) of 82.6% highlighted the inherent limitations of kinematic models alone, particularly in cases involving cushioned or incomplete falls.

This limitation led to the critical development in Chapter 8. The specialized Kolmogorov–Arnold Network (Fall-KAN) successfully estimated the Time-to-Impact with a superior mean Root Mean Squared Error (RMSE) of approximately 159 ms compared to traditional baselines. This level of temporal accuracy is a prerequisite for pre-impact mitigation systems such as airbag deployment. Furthermore, the Fall Forecasting Neural Network (FFNN), trained on an Inverted Pendulum model, validated the feasibility of trajectory forecasting, demonstrating a predictive lead time sufficient for actuation. This dual approach

validates the framework's ability to execute high-speed, safety-critical decisions during the fall event itself, successfully bridging the gap between detection and injury mitigation.

Chapter 9 addressed the root causes of falls by developing a multimodal and personalized framework for Fall Risk Estimation. The XoADL and XoFallVR datasets confirmed that biosignals (EMG, HR, EDA, and Pupillometry) capture distinctive physiological and psychological states (fatigue and stress) that precede instability. Crucially, the performance analysis using Leave-One-Subject-Out (LOSO) cross-validation demonstrated that generalized models fail to reliably classify fall risk due to high inter-subject variability. This significant finding necessitated a shift to a personalized modeling approach. This personalized strategy achieved a high test-set balanced accuracy of 90.39% for the binary (Low vs. High) Fall scenario. This robust performance validates the hypothesis that continuous, high-accuracy fall-risk prediction requires an individualized baseline and training period. It suggests that fall risk is fundamentally a subject-specific problem, a conclusion strongly supported by the VR-based data gathered in an ecologically valid and high-stress environment.

10.2 Limitations

Despite the demonstrated strengths, the work includes several limitations inherent to the field:

- Reliance on Simulated Kinematics: The kinematic data used for detection and forecasting (SisFall) involves simulated falls, which may not fully capture the complex, compensatory movements of real-world falls by vulnerable populations.
- Improvable performance: Although Fall-KAN's performance is superior to that of the reference models, there is still margin for improvement regarding the reliability of the model for full integration into a fall protection system.
- Idealized Biomechanical Model: The FFNN relied on the simplified Inverted Pendulum (IP) model. While effective for proof-of-concept, real human falls involve multi-segment dynamics and active muscle response that the current model idealizes.
- VR Environment Fidelity: While the VR environment proved engaging and effective at eliciting physiological responses, the observed drops in Spatial Presence and Experienced Realism suggest that the fidelity of the visual experience could be further optimized to reduce the mild reported motion sickness symptoms (UFS).

- Lack of generalization: personalized models have demonstrated significantly superior performance compared to generalized models, which nevertheless limit the use of the results obtained for large-scale model monitoring and complicate the practical application of the system.

Chapter 11

Future perspectives

The findings presented in this thesis open several promising research directions, particularly regarding the integration of the developed algorithms into robust real-world systems and the expansion of predictive, physiological, and multimodal modeling. The following sections outline the most relevant future perspectives, motivated by the limitations identified in the previous chapters and the opportunities emerging from recent scientific and technological developments.

11.1 Overcoming current limitations

The first step toward advancing this research is to address the constraints observed throughout the development and evaluation of the proposed framework. Strengthening each component of the system will not only improve performance but also broaden its applicability to real-world scenarios.

11.1.1 Multi-dataset validation

Currently, FDNN and Fall-KAN have been trained and validated exclusively on the SisFall dataset, which, despite its richness, contains only simulated falls and constrained ADLs. Incorporating additional datasets, such as those reviewed in Chapter 7.3, will improve generalization, reduce dataset-induced biases, and ensure robustness against diverse movement patterns. Moreover, expanding training to datasets with real-world falls and broader demographic variability would significantly strengthen the reliability of the proposed system. Parallel to this, the development of a fall-prediction model capable of estimating inverted-

pendulum parameters may simplify and speed up the process, offering a different or better approach to achieve the same result.

11.1.2 Advanced biomechanical modeling

The current inverted pendulum (IP) model has proven helpful for forecasting trajectories in simple forward or backward falls. However, human balance recovery involves multi-segment coordination, such as hip strategy, arm movements, and torso counter-rotation. Extending the biomechanical model to two- or three-link systems, or even developing a hybrid data-driven and physics-based model, would allow the forecasting framework to capture compensatory strategies and produce more realistic predictions. It will be essential to cover a broader spectrum of fall types, including lateral, rotational, and complex falls characteristic of occupational environments.

11.1.3 Kolmogorov-Arnold Network investigation

Kolmogorov–Arnold Networks (KANs) have rapidly become a highly active research topic, with over 3000 citations since their reintroduction in 2024. Their ability to learn functional decompositions with adaptive, spline-based activation functions makes them an attractive alternative to classical MLPs for interpretable regression and time-series forecasting. Investigating new KAN variants, such as TKAN, GKAN, Wav-KAN, and DeepOKAN, may provide substantial improvements for fall-time prediction, especially in noisy or highly non-linear conditions. A deeper exploration of KANs within the context of human movement and safety-critical forecasting is therefore an auspicious future direction.

11.1.4 XoFallVRv2: Next-Generation VR-Based fall-risk datasets

The thesis highlighted the value of immersive VR environments in eliciting physiological and psychological responses comparable to those in real fall-risk scenarios. A second-generation dataset (XoFallVRv2), designed to address the limitations identified in spatial presence, experiential realism, and motion-related discomfort, would represent an important contribution to the community. Making such datasets openly available would establish a shared benchmark for researchers developing fall-risk estimation and fall impact mitigation systems, thereby reducing participants' exposure to hazardous environments.

11.1.5 Integration of Cognitive Load and Attention

Initial work using pupillometry demonstrated the feasibility of tracking cognitive state during complex tasks. Future work should integrate metrics of cognitive load (such as dual-task performance and EEG spectral analysis) directly into the fall-risk estimation framework. Since many falls occur during moments of divided attention, quantifying the temporal interaction between cognitive fatigue, physiological arousal (as measured by EDA/HRV), and kinematic instability is crucial for developing a truly predictive warning system. It could lead to a system that, for instance, warns a worker when their combined physiological and cognitive fatigue makes working at height unsafe.

11.2 Real-world system deployment

The most immediate future perspective involves transitioning the personalized fall-risk estimation and mitigation models from the laboratory to edge-based real-world deployment.

The current work successfully used a fusion of commercial wearable biosensors (Empatica E4, Polar H10). Future research should focus on integrating the framework directly with a wearable mitigation system (such as an airbag vest). It involves rigorously testing the total system latency, from the sensor reading the fall detection to the Fall-KAN/FFNN prediction output, to guarantee the necessary millisecond lead time for actuator deployment.

The validated high performance of the personalized models necessitates a user-specific training phase. It requires developing Edge-AI solutions that can run complex Deep Learning models (like the FDNN and FFNN) locally on wearable devices. Future work should prioritize optimizing these models (such as utilizing quantization and pruning) to ensure they meet the stringent computational and energy constraints of battery-powered wearable devices.

Furthermore, integrating a Federated Learning paradigm would be crucial. It allows continuous learning and refinement of the subject-specific model (the "personalized layer") using new physiological data collected during daily life, without compromising user privacy by transmitting raw data. This mechanism will enable the models to adapt to changes in the user's health status, fitness level, and aging processes, a crucial capability for long-term clinical and occupational relevance.

The most immediate and impactful future direction involves transitioning the personalized fall-risk estimation and mitigation framework from controlled laboratory experiments to real-world, edge-based deployment. This requires integrating sensing, prediction, and actuation

into a unified wearable system that can operate reliably in both daily life and occupational environments.

The current work demonstrates the feasibility of multimodal physiological and kinematic monitoring using commercial wearable biosensors (Empatica E4, Polar H10). The next innovation step is the direct integration of the fall detection and forecasting pipeline with an active mitigation device, such as an airbag vest. Achieving this requires rigorous end-to-end latency characterization, from raw IMU/biomedical signal acquisition to inference from the FDNN/Fall-KAN models, ensuring a total response time compatible with the sub-100 ms deployment window required for injury-preventive actuation.

Since the models developed in this thesis rely on personalized components, real-world deployment will demand on-device adaptation and real-time resource management. It highlights the necessity for edge-AI strategies capable of executing deep learning models (such a as FDNN, FFNN, Fall-KAN) on embedded processors. Techniques such as model quantization, pruning, layer fusion, and hardware-aware optimization should be explored to meet the stringent constraints of wearable systems, including limited memory, low power consumption, and uninterrupted operation during physically demanding work tasks. To maintain long-term accuracy, the system must be able to adapt as users' physiological patterns evolve due to factors such as fatigue, health changes, aging, or variations in working conditions. Implementing Federated Learning (FL) is therefore a crucial future step. FL enables decentralized, privacy-preserving model refinement directly on users' devices, ensuring continuous personalization without transmitting sensitive raw data. It is particularly relevant in workplace scenarios, where privacy regulations (such as GDPR and HIPAA) and ethical constraints are crucial.

Finally, it is essential to recognize that deploying such a system outside the laboratory requires compliance with medical device, occupational safety, and personal protective equipment (PPE) certification frameworks. Standards such as ISO 13485, IEC 62304, ISO 80601-2-XX, EU MDR (2017/745), and PPE-specific regulations (such as EU Regulation 2016/425) will govern the validation, risk management, and life-cycle maintenance of the device. Addressing these regulatory pathways is not merely an administrative step; it will shape software architecture, data logging procedures, safety mechanisms, and human-machine interaction requirements.

In summary, future work should focus on system-level integration, embedded AI optimization, lifelong personalization, privacy-preserving learning, and regulatory certification, ultimately enabling the deployment of a reliable, safe, and field-ready fall-mitigation platform.

Chapter 12

Conclusions

This thesis aimed to advance the scientific foundation of fall detection, impact prediction, and fall-risk estimation by developing a unified, multimodal, and machine-learning-based framework that addresses the limitations of current technologies.

The first objective was to explore and benchmark data-processing techniques and machine-learning models for detecting slips, trips, and falls, estimating time-to-impact, and identifying fall-related indicators. This was accomplished through the development of a temporal deep neural model trained on complete IMU sequences, which demonstrated strong pre-impact fall detection performance on the SisFall dataset and addressed long-standing limitations of threshold-based or handcrafted feature approaches. The results showed that temporal representation learning is essential for recognizing the subtle onset of a fall and for minimizing false alarms, thereby confirming the feasibility of reliable real-time fall detection within wearable devices.

The second objective involved designing and collecting multimodal datasets that could support a comprehensive evaluation of fall dynamics, physiological responses, and human behavior under realistic conditions. This goal was achieved through the creation of two novel datasets: one capturing physiological and biomechanical signals during a broad range of daily activities, and another collected in immersive virtual-reality simulations designed to induce stress, height exposure, and balance perturbation safely. Together, these datasets provided a rich foundation for modeling fall-related mechanisms that cannot be replicated through traditional laboratory-based fall simulations, enabling the investigation of both reactive and predictive aspects of fall-related safety.

The third objective focused on developing individualized and adaptive models able to capture the evolving and subject-specific nature of fall risk. The research demonstrated that generic and population-level models fail to generalize across individuals, particularly when

physiological responses to stress, fatigue, or exposure to height vary widely. By transitioning to a personalized classification framework, the thesis demonstrated that fall-risk estimation can achieve substantially higher accuracy and robustness, with personalized models exhibiting strong and consistent performance in distinguishing between low- and high-risk episodes. This finding confirms that personalization is not a supplementary refinement, but a necessary condition for reliable physiological fall-risk estimation.

The fourth objective aimed to assess the usability, acceptability, and feasibility of multimodal fall-risk monitoring systems for simulated real-world occupational contexts. This was addressed through the integration of biosensing devices, wearable units, and VR-based tools within experimental studies that examined user comfort, workload, immersion, and overall acceptance. The results indicated that participants tolerated the sensing modalities well and were able to perform demanding tasks while being continuously monitored. The integration of these technologies proved feasible for realistic VR-based data collection and real-world deployment, suggesting that such systems can be effectively translated into operational environments where accurate, continuous monitoring is essential.

Taken together, the thesis successfully fulfilled all the objectives stated in Section 1.2. It demonstrated that reactive fall detection can be achieved reliably through temporal deep learning; that proactive impact prediction is feasible through models such as the Kolmogorov–Arnold Network and inverted-pendulum-driven forecasting; that personalized physiological modeling provides a powerful basis for predicting fall risk before an accident occurs; and that the proposed sensing and computational framework can operate in realistic occupational scenarios. In doing so, the work confirms the central hypothesis that fall-injury prevention requires unified, predictive, and personalized modeling approaches capable of capturing both the biomechanical and physiological dimensions of human stability.

References

[1] "Accidents at work statistics - statistics explained - eurostat."

[2] L. Colazzo, A. Cuteri, B. Martini, A. Sciarrone, S. Nava, P. Santia, G. Campo, A. Guglielmi, and M. Dei, "Le cadute dall'alto dei lavoratori," *Sistema di sorveglianza degli infortuni mortali sul lavoro*, vol. 2, pp. 1–6, 1 2017.

[3] R. Uppal, "Advanced mems accelerometers: Technical innovations and applications across industries," 11 2024.

[4] X. Wei, Y. Wang, and C. Sun, "Perdet: Machine-learning-based uav gps spoofing detection using perception data," *Remote Sensing*, vol. 14, no. 19, 2022.

[5] M. Karimi, E. Babaians, M. Oelsch, T. Aykut, and E. Steinbach, "Skewed-redundant hall-effect magnetic sensor fusion for perturbation-free indoor heading estimation," in *2020 Fourth IEEE International Conference on Robotic Computing (IRC)*, pp. 367–374, 2020.

[6] A. Schumann, M. Schmidt, M. Herbsleb, C. Semm, G. Rose, H. Gabriel, and K.-J. Bär, "Deriving respiration from high resolution 12-channel-ecg during cycling exercise," *Current Directions in Biomedical Engineering*, vol. 2, no. 1, pp. 171–174, 2016.

[7] R. K. Pandey and P. C.-P. Chao, "External temperature sensor assisted a new low power photoplethysmography readout system for accurate measurement of the bio-signs," *Microsystem Technologies*, vol. 27, no. 6, pp. 2315–2343, 2021.

[8] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi, "cvxeda: A convex optimization approach to electrodermal activity processing," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 797–804, 2016.

[9] B. L. Lussier, D. M. Olson, and V. Aiyagari, "Automated pupillometry in neurocritical care: research and practice," *Current neurology and neuroscience reports*, vol. 19, no. 10, p. 71, 2019.

[10] Safety and M. (SAM), *Step safely: strategies for preventing and managing falls across the life-course*. World Health Organization, 2021.

[11] T. Locklear, J. Kontos, C. A. Brock, A. B. Holland, R. Hemsath, A. Deal, S. Leonard, C. Steinmetz, and S. Biswas, "Inpatient falls: Epidemiology, risk assessment, and prevention measures. a narrative review," *HCA Healthcare Journal of Medicine*, vol. 5, no. 5, p. 517, 2024.

[12] J. D. S. et al., "Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017," *The Lancet*, vol. 392, no. 10159, pp. 1923–1994, 2018.

[13] R. Kakara, "Nonfatal and fatal falls among adults aged $\geq$ 65 years—united states, 2020–2021," *MMWR. Morbidity and Mortality Weekly Report*, vol. 72, 2023.

[14] M. F. Garnett, J. D. Weeks, and A. M. Zehner, "Unintentional fall deaths in adults age 65 and older: United states, 2023," *NCHS Data Brief*, vol. 532, pp. 1–11, 2023.

[15] "Cost of older adult falls," July 9, 2020.

[16] U. N. D. of Economic and S. Affairs, *World Population Ageing 2020: Highlights*. United Nations, 2021.

[17] E. van Boetzelaer, L. Rathod, P. Keating, U. Pellecchia, S. Sharma, J. Nickerson, J. van de Kamp, O. H. Franco, J. Smith, F. Escobio, and J. L. Browne, "Health needs of older people and age-inclusive health care in humanitarian emergencies in low-income and middle-income countries: a systematic review," *The Lancet Healthy Longevity*, vol. 6, no. 1, p. 100663, 2025.

[18] C. Mathews, "Population ageing in lower-and middle-income countries. celebrating 40 years of helpage international," *Journal of Population Ageing*, vol. 17, no. 1, pp. 1–4, 2024.

[19] S. F. Disability and R. (SDR), *World report on disability*. World Health Organization, 2021.

[20] E. Attaianese, G. Bufalo, R. d'Angelo, G. Duca, G. De Margheriti, P. De Joanna, A. Giglio, L. Mennella, and E. Russo, *Valutare il rischio di caduta in piano*, vol. Progetto RAS, Ricercare e Applicare la Sicurezza. INAIL, 2021.

[21] S. Usmani, A. Saboor, M. Haris, M. A. Khan, and H. Park, "Latest research trends in fall detection and prevention using machine learning: A systematic review," *Sensors*, vol. 21, no. 15, 2021.

[22] T. Wu, Y. Gu, Y. Chen, J. Wang, and S. Zhang, "A mobile cloud collaboration fall detection system based on ensemble learning," in *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '20, (New York, NY, USA), Association for Computing Machinery, 2020.

[23] A. Ramachandran and A. Karuppiah, "A survey on recent advances in wearable fall detection systems," *BioMed Research International*, vol. 2020, no. 1, p. 2167160, 2020.

[24] X. Wang, J. Ellul, and G. Azzopardi, "Elderly fall detection systems: A literature survey," *Frontiers in Robotics and AI*, vol. Volume 7 - 2020, 2020.

[25] M. Saleh and R. L. B. Jeannès, "Elderly fall detection using wearable sensors: A low cost highly accurate algorithm," *IEEE Sensors Journal*, vol. 19, no. 8, pp. 3156–3164, 2019.

[26] J. P. Queralta, T. N. Gia, H. Tenhunen, and T. Westerlund, "Edge-ai in lora-based health monitoring: Fall detection system with fog computing and lstm recurrent neural networks," in *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*, pp. 601–604, 2019.

[27] A. Chelli and M. Pätzold, "A machine learning approach for fall detection and daily living activity recognition," *IEEE Access*, vol. 7, pp. 38670–38687, 2019.

[28] T. Namba and Y. Yamada, "Risks of deep reinforcement learning applied to fall prevention assist by autonomous mobile robots in the hospital," *Big Data and Cognitive Computing*, vol. 2, no. 2, 2018.

[29] T. Namba and Y. Yamada, "Fall risk reduction for the elderly by using mobile robots based on deep reinforcement learning," *Journal of Robotics, Networking and Artificial Life*, vol. 4, pp. 265–269, 2018.

[30] C. Lord and D. Colvin, "Falls in the elderly: Detection and assessment," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society Volume 13: 1991*, pp. 1938–1939, 1991.

[31] A. Sucerquia, J. D. López, and J. F. Vargas-Bonilla, "Sisfall: A fall and movement dataset," *Sensors*, vol. 17, no. 1, 2017.

[32] L. Martínez-Villaseñor, H. Ponce, J. Brieva, E. Moya-Albor, J. Núñez-Martínez, and C. Peñafort-Asturiano, "Up-fall detection dataset: A multimodal approach," *Sensors*, vol. 19, no. 9, 2019.

[33] E. Casilari, J. A. Santoyo-Ramón, and J. M. Cano-García, "Umafall: A multisensor dataset for the research on automatic fall detection," *Procedia Computer Science*, vol. 110, pp. 32–39, 2017. 14th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2017) / 12th International Conference on Future Networks and Communications (FNC 2017) / Affiliated Workshops.

[34] C. Chatzaki, M. Pediaditis, G. Vavoulas, and M. Tsiknakis, "Human daily activity and fall recognition using a smartphone's acceleration sensor," in *Information and Communication Technologies for Ageing Well and e-Health* (C. Röcker, J. O'Donoghue, M. Ziefle, M. Helfert, and W. Molloy, eds.), (Cham), pp. 100–118, Springer International Publishing, 2017.

[35] T. R. Mauldin, M. E. Canby, V. Metsis, A. H. H. Ngu, and C. C. Rivera, "Smartfall: A smartwatch-based fall detection system using deep learning," *Sensors*, vol. 18, no. 10, 2018.

[36] D. Micucci, M. Mobilio, and P. Napoletano, "Unimib shar: A dataset for human activity recognition using acceleration data from smartphones," *Applied Sciences*, vol. 7, no. 10, 2017.

[37] X. Yu, J. Jang, and S. Xiong, "A large-scale open motion dataset (kfall) and benchmark algorithms for detecting pre-impact fall of the elderly using wearable inertial sensors," *Frontiers in Aging Neuroscience*, vol. Volume 13 - 2021, 2021.

[38] L. M. B. Soares, P. Rangel, G. Venturini, J. Oliveira, J. Gomes, E. Bezerra, and D. Brandão, "Ipqm-fall," 2024.

[39] J. Klenk, L. Schwickert, L. Palmerini, S. Mellone, A. Bourke, E. A. Ihlen, N. Kerse, K. Hauer, M. Pijnappels, M. Synofzik, *et al.*, "The farseeing real-world fall repository: a large-scale collaborative database to collect and share sensor signals from real-world falls," *European review of aging and physical activity*, vol. 13, no. 1, p. 8, 2016.

[40] N. Cartocci, A. E. Gkikakis, D. G. Caldwell, and J. Ortiz, "Real-time fall prevention system for the next-generation of workers," 2022.

[41] N. Cartocci, A. Gkikakis, D. Caldwell, and J. Ortiz, "Artificial intelligence-based wearable solution to prevent fall from heights injuries for the next generation of workers," in *Slips, Trips and Falls (STF) International Conference*, 06 2023.

[42] N. Cartocci, A. E. Gkikakis, D. G. Caldwell, and J. Ortiz, "Deep learning-based wearable device to prevent fall from height injuries," Oct. 2023.

[43] N. Cartocci, A. E. Gkikakis, F. Pera, M. T. Settino, D. G. Caldwell, and J. Ortiz, "Fall-kan: Fall impact time estimation kolmogorov-arnold network," in *2024 4th International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, pp. 1–6, 2024.

[44] N. Cartocci, A. E. Gkikakis, N. Kurvina, N. Takele, F. Pera, M. T. Settino, D. G. Caldwell, and J. Ortiz, "Recognition of physiological patterns during activities of daily living using wearable biosignal sensors," in *Proceedings of the 22nd Congress of the International Ergonomics Association, Volume 3* (S. Jin, J. H. Kim, Y.-K. Kong, J. Park, and M. H. Yun, eds.), (Singapore), pp. 447–454, Springer Nature Singapore, 2025.

[45] N. Cartocci, A. E. Gkikakis, R. F. Pitzalis, F. Pera, M. T. Settino, D. G. Caldwell, and J. Ortiz, "How can ai reduce fall injuries in the workplace?," 2025.

[46] R. F. Pitzalis, B. Lagomarsino, I. Ceroni, N. Cartocci, J. Ahmad, G. A. Albanese, J. Zenzeri, C. Di Natali, L. Monica, D. G. Caldwell, M. Casadio, and J. Ortiz, "Evaluating muscle fatigue with non-invasive approaches: A review of methods, metrics, and implications," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 33, pp. 2707–2723, 2025.

[47] G. Mariani, C. Lambranzi, N. Cartocci, G. Barresi, C. Di Natali, E. De Momi, and J. Ortiz, "Physiological measures of the mental workload in users of a lower limb exosuit: A comparison of subjective and objective metrics," in *2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 3948–3955, 2025.

[48] O. A. Moreno Franco, N. Cartocci, A. E. Gkikakis, G. Giurin, M. Carega, Y. Tefera, N. Deshpande, F. M. Fabiani, D. G. Caldwell, and J. Ortiz, "Telepresence in construction sites: Usability study of a vr-based environment for working in heights," in *2025 IEEE Conference on Telepresence*, 2025.

[49] W. H. Organization, *Global action plan on physical activity 2018-2030: more active people for a healthier world*. World Health Organization, 2019.

[50] W. H. O. Ageing and L. C. Unit, *WHO Global Report on Falls Prevention in Older Age*. Nonserial Publication Series, World Health Organization, 2008.

[51] J. Pillay, L. A. Gaudet, S. Saba, B. Vandermeer, A. R. Ashiq, A. Wingert, and L. Hartling, "Falls prevention interventions for community-dwelling older adults: Systematic review and meta-analysis of benefits, harms, and patient values and preferences," *Systematic Reviews*, vol. 13, no. 1, p. 289, 2024.

[52] A. F. Ambrose, G. Paul, and J. M. Hausdorff, "Risk factors for falls among older adults: A review of the literature," *Maturitas*, vol. 75, no. 1, pp. 51–61, 2013.

[53] L. Z. Rubenstein, "Falls in older people: epidemiology, risk factors and strategies for prevention," *Age and Ageing*, vol. 35, pp. ii37–ii41, 09 2006.

[54] R. W. Sattin, "Falls among older persons: A public health perspective," *Annual Review of Public Health*, vol. 13, no. Volume 13, 1992, pp. 489–508, 1992.

[55] S. L. James, L. R. Lucchesi, C. Bisignano, C. D. Castle, Z. V. Dingels, J. T. Fox, E. B. Hamilton, N. J. Henry, K. J. Krohn, Z. Liu, D. McCracken, M. R. Nixon, N. L. S. Roberts, D. O. Sylte, J. C. Adsuar, A. Arora, A. M. Briggs, D. Collado-Mateo, C. Cooper, L. Dandona, R. Dandona, C. L. Ellingsen, S.-M. Fereshtehnejad, T. K. Gill, J. A. Haagsma, D. Hendrie, M. Jürisson, G. A. Kumar, A. D. Lopez, T. Miazgowski, T. R. Miller, G. Mini, E. M. Mirrakhimov, E. Mohamadi, P. R. Olivares, F. Rahim, L. S. Riera, S. Villafaina, Y. Yano, S. I. Hay, S. S. Lim, A. H. Mokdad, M. Naghavi, and C. J. L. Murray, "The global burden of falls: global, regional and national estimates of morbidity and mortality from the global burden of disease study 2017," *Injury Prevention*, vol. 26, no. Suppl 2, pp. i3–i11, 2020.

[56] K. Ashby and M. Corbo, "Child fall injuries: an overview," *Victoria: Victorian Injury Surveillance & Applied Research System*, 2000.

[57] S. I. Hay, K. L. Ong, D. F. Santomauro, M. A. Aalipour, H. Aalruz, H. S. Ababneh, U. O. Abaraogu, B. B. Abate, C. Abbafati, N. Abbas, *et al.*, "Burden of 375 diseases and injuries, risk-attributable burden of 88 risk factors, and healthy life expectancy in 204 countries and territories, including 660 subnational locations, 1990–2023: a systematic analysis for the global burden of disease study 2023," *The Lancet*, 2025.

[58] J. Takala, P. Hämäläinen, R. Sauni, C.-H. Nygård, D. Gagliardi, and S. Neupane, "Global-, regional- and country-level estimates of the work-related burden of diseases and accidents in 2019," *Scandinavian Journal of Work, Environment & Health*, pp. 73–82, Mar 2024.

[59] F. Pega, H. Hamzaoui, B. Náfrádi, and N. C. Momen, "Global, regional and national burden of disease attributable to 19 selected occupational risk factors for 183 countries, 2000–2016: A systematic analysis from the who/ilo joint estimates of the work-related burden of disease and injury," *Scandinavian Journal of Work, Environment & Health*, pp. 158–168, Mar 2022.

[60] "A look at falls, slips, and trips in the construction industry : The economics daily: U.s. bureau of labor statistics."

[61] "Working at height - oshwiki | european agency for safety and health at work."

[62] M. F. Nasarwanji, "Contributing factors to slip, trip, and fall fatalities at surface coal and metal/nonmetal mines," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 60, no. 1, pp. 1666–1670, 2016.

[63] J. L. Bell, J. W. Collins, H. M. Tiesman, M. Ridenour, S. Konda, L. Wolf, and B. Evanoff, "Slip, trip, and fall injuries among nursing care facility workers," *Workplace Health & Safety*, vol. 61, no. 4, pp. 147–152, 2013. PMID: 23521142.

[64] T. K. Courtney, G. S. Sorock, D. P. Manning, J. W. Collins, and M. A. Holbein-Jenny, "Occupational slip, trip, and fall-related injuries can the contribution of slipperiness be isolated?," *Ergonomics*, vol. 44, no. 13, pp. 1118–1137, 2001. PMID: 11794761.

[65] T. Bentley and R. Haslam, "Identification of risk factors and countermeasures for slip, trip and fall accidents during the delivery of mail," *Applied Ergonomics*, vol. 32, no. 2, pp. 127–134, 2001. STFA 98: An International Conference on Slipping, Tripping and Falling Accidents.

[66] K. Kemmlert and L. Lundholm, "Slips, trips and falls in different work groups — with reference to age and from a preventive perspective," *Applied Ergonomics*, vol. 32, no. 2, pp. 149–153, 2001. STFA 98: An International Conference on Slipping, Tripping and Falling Accidents.

[67] S. Okada, K. Hirakawa, Y. Takada, and H. Kinoshita, "Age-related differences in postural control in humans in response to a sudden deceleration generated by postural disturbance," *European journal of applied physiology*, vol. 85, no. 1, pp. 10–18, 2001.

[68] A. J. CAMPBELL, M. J. BORRIE, G. F. SPEARS, S. L. JACKSON, J. S. BROWN, and J. L. FITZGERALD, "Circumstances and consequences of falls experienced by a community population 70 years and over during a prospective study," *Age and Ageing*, vol. 19, pp. 136–141, 03 1990.

[69] E. Vlaeyen, J. Coussement, G. Leysens, E. Van der Elst, K. Delbaere, D. Cambier, K. Denhaerynck, S. Goemaere, A. Wertelaers, F. Dobbels, E. Dejaeger, K. Milisen, and the Center of Expertise for Fall and Fracture Prevention Flanders , "Characteristics and effectiveness of fall prevention programs in nursing homes: A systematic review and meta-analysis of randomized controlled trials," *Journal of the American Geriatrics Society*, vol. 63, no. 2, pp. 211–221, 2015.

[70] J. Stevens, G. Ryan, and M. Kresnow, "Fatalities and injuries from falls among older adults - united states, 1993-2003 and 2001-2005," *Morbidity and Mortality Weekly Report*, vol. 55, pp. 1221–1224, 11 2006.

[71] S. Niemann, Y. A. Stürmer, L. Ellenberger, and D. Meier, "Statistica degli infortuni non professionali e del livello di sicurezza in svizzera," *UPI, Ufficio prevenzione infortuni*, pp. 1–76, 2025.

[72] C. S. Florence, G. Bergen, A. Atherly, E. Burns, J. Stevens, and C. Drake, "Medical costs of fatal and nonfatal falls in older adults," *Journal of the American Geriatrics Society*, vol. 66, no. 4, pp. 693–698, 2018.

[73] S. K. Verma, J. L. Willetts, H. L. Corns, H. R. Marucci-Wellman, D. A. Lombardi, and T. K. Courtney, "Falls and fall-related injuries among community-dwelling adults in the united states," *PLOS ONE*, vol. 11, pp. 1–14, 03 2016.

[74] K. A. Hartholt, E. F. van Beeck, S. Polinder, N. van der Velde, E. M. Van Lieshout, M. J. Panneman, T. J. van der Cammen, and P. Patka, "Societal consequences of falls in the older population: injuries, healthcare costs, and long-term reduced quality of life," *Journal of Trauma and Acute Care Surgery*, vol. 71, no. 3, pp. 748–753, 2011.

[75] M. E. Tinetti and C. Kumar, "The patient who falls: "it's always a trade-off"," *JAMA*, vol. 303, pp. 258–266, 01 2010.

[76] D. Carey and M. Laffoy, "Hospitalisations due to falls in older persons.," *Irish medical journal*, vol. 98, no. 6, pp. 179–181, 2005.

[77] B. S. Roudsari, B. E. Ebel, P. S. Corso, N.-A. M. Molinari, and T. D. Koepsell, "The acute medical care costs of fall-related injuries among the us older adults," *Injury*, vol. 36, no. 11, pp. 1316–1322, 2005.

[78] S. R. Cummings, S. M. Rubin, and D. Black, "The future of hip fractures in the united states: numbers, costs, and potential effects of postmenopausal estrogen," *Clinical Orthopaedics and Related Research®*, vol. 252, pp. 163–166, 1990.

[79] N. inform, "Causes of falls."

[80] S. N. K. Kodithuwakku Arachchige, H. Chander, A. J. Turner, and A. C. Knight, "Impact of prolonged exposure to a slippery surface on postural stability," *International Journal of Environmental Research and Public Health*, vol. 18, no. 5, 2021.

[81] W.-R. Chang, S. Leclercq, T. E. Lockhart, and R. Haslam, "State of science: occupational slips, trips and falls on the same level*," *Ergonomics*, vol. 59, no. 7, pp. 861–883, 2016. PMID: 26903401.

[82] H. HSIAO, "Fall prevention research and practice: A total worker safety approach," *Industrial Health*, vol. 52, no. 5, pp. 381–392, 2014.

[83] N. I. on Aging, "Falls and fractures in older adults: Causes and prevention," 2022.

[84] M. K. Appeadu and B. Bordoni, "Falls and fall prevention in older adults," in *StatPearls [Internet]*, StatPearls Publishing, 2023.

[85] J. E. Crews, "Falls among persons aged $\geq$ 65 years with and without severe vision impairment—united states, 2014," *MMWR. Morbidity and Mortality Weekly Report*, vol. 65, 2016.

[86] e. a. Montero-Odasso, "World guidelines for falls prevention and management for older adults: a global initiative," *Age and Ageing*, vol. 51, p. afac205, 09 2022.

[87] D. of Health, H. Services, and C. for Disease Control, "Fact sheet risk factors for falls," 2017.

[88] R. Vaishya and A. Vaish, "Falls in older adults are serious," *Indian journal of orthopaedics*, vol. 54, no. 1, pp. 69–74, 2020.

[89] Y. Li, L. Hou, H. Zhao, R. Xie, Y. Yi, and X. Ding, "Risk factors for falls among community-dwelling older adults: A systematic review and meta-analysis," *Frontiers in Medicine*, vol. Volume 9 - 2022, 2023.

[90] K. Lafontant, A. Blount, J. R. M. Suarez, D. H. Fukuda, J. R. Stout, E. M. Trahan, N. R. Lighthall, J.-H. Park, R. Xie, and L. Thiamwong, "Comparing sensitivity, specificity, and accuracy of fall risk assessments in community-dwelling older adults," *Clinical interventions in aging*, pp. 581–588, 2024.

[91] B. H. Shaw and V. E. Claydon, "The relationship between orthostatic hypotension and falling in older adults," *Clinical autonomic research*, vol. 24, no. 1, pp. 3–13, 2014.

[92] D. L. Sturnieks, A. Tiedemann, K. Chapman, B. Munro, S. M. Murray, and S. R. Lord, "Physiological risk factors for falls in older people with lower limb arthritis.," *The Journal of Rheumatology*, vol. 31, no. 11, pp. 2272–2279, 2004.

[93] Y.-C. Pai and J. Patton, "Center of mass velocity-position predictions for balance control," *Journal of Biomechanics*, vol. 30, no. 4, pp. 347–354, 1997.

[94] M. Woollacott and A. Shumway-Cook, "Attention and the control of posture and gait: a review of an emerging area of research," *Gait & Posture*, vol. 16, no. 1, pp. 1–14, 2002.

[95] S. N. Ramesh, M. Sarkar, M. Audette, and C. Paolini, "Efficient real-time fall prediction and detection using privacy-centric vision-based human pose estimation on the xilinx® kria™ k26 som," in *2023 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp. 1–5, 2023.

[96] "Dairlab - workair." https://dairlab.com. Accessed: 2025-11-25.

[97] "Protechto - mad work." https://www.protechto.it. Accessed: 2025-11-25.

[98] "Skylotec - skyvest." https://www.skylotec.com. Accessed: 2025-11-25.

[99] "Safeware - c3/c-light." https://en.safeware.co.kr/. Accessed: 2025-11-25.

[100] "Chutex - skyvest." https://chutex-innovation.com. Accessed: 2025-11-25.

[101] "S-airbag." https://sairbagpro.com/. Accessed: 2025-11-25.

[102] J. R. C. da Silva, *Machine learning applied to fall prediction and detection using wearable sensors*. PhD thesis, University of Porto, 2020.

[103] P. Groves, *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems, Second Edition*. Artech House Publishers, 03 2013.

[104] D. Titterton and J. L. Weston, *Strapdown inertial navigation technology*, vol. 17. IET, 2004.

[105] N. Barbour and G. Schmidt, "Inertial sensor technology trends," *IEEE Sensors Journal*, vol. 1, no. 4, pp. 332–339, 2001.

[106] H.-F. Liu, Z.-C. Luo, Z.-K. Hu, S.-Q. Yang, L.-C. Tu, Z.-B. Zhou, and M. Kraft, "A review of high-performance mems sensors for resource exploration and geophysical applications," *Petroleum Science*, vol. 19, no. 6, pp. 2631–2648, 2022.

[107] A. S. Algamili, M. H. M. Khir, J. O. Dennis, A. Y. Ahmed, S. S. Alabsi, S. S. Ba Hashwan, and M. M. Junaid, "A review of actuation and sensing mechanisms in mems-based sensor devices," *Nanoscale research letters*, vol. 16, no. 1, p. 16, 2021.

[108] F. Bagalà, C. Becker, A. Cappello, L. Chiari, K. Aminian, J. M. Hausdorff, W. Zijlstra, and J. Klenk, "Evaluation of accelerometer-based fall detection algorithms on real-world falls," *PLOS ONE*, vol. 7, pp. 1–9, 05 2012.

[109] T. G. Stavropoulos, A. Papastergiou, L. Mpaltadoros, S. Nikolopoulos, and I. Kompatsiaris, "Iot wearable sensors and devices in elderly care: A literature review," *Sensors*, vol. 20, no. 10, 2020.

[110] S. Patel, H. Park, P. Bonato, L. Chan, and M. Rodgers, "A review of wearable sensors and systems with application in rehabilitation," *Journal of neuroengineering and rehabilitation*, vol. 9, no. 1, p. 21, 2012.

[111] W. A. Gill, I. Howard, I. Mazhar, and K. McKee, "A review of mems vibrating gyroscopes and their reliability issues in harsh environments," *Sensors*, vol. 22, no. 19, 2022.

[112] D. Roetenberg, H. Luinge, C. Baten, and P. Veltink, "Compensation of magnetic disturbances improves inertial and magnetic sensing of human body segment orientation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 3, pp. 395–405, 2005.

[113] Y. Zhuang, X. Sun, Y. Li, J. Huai, L. Hua, X. Yang, X. Cao, P. Zhang, Y. Cao, L. Qi, J. Yang, N. El-Bendary, N. El-Sheimy, J. Thompson, and R. Chen, "Multi-sensor integrated navigation/positioning systems using data fusion: From analytics-based to learning-based approaches," *Information Fusion*, vol. 95, pp. 62–90, 2023.

[114] P. Ripka, *Magnetic sensors and magnetometers*. Artech house, 2021.

[115] M. A. Khan, J. Sun, B. Li, A. Przybysz, and J. Kosel, "Magnetic sensors-a review and recent technologies," *Engineering Research Express*, vol. 3, p. 022005, jun 2021.

[116] B. Fan, Q. Li, and T. Liu, "How magnetic disturbance influences the attitude and heading in magnetic and inertial sensor-based orientation estimation," *Sensors*, vol. 18, no. 1, 2018.

[117] D. Su, K. Wu, R. Saha, C. Peng, and J.-P. Wang, "Advances in magnetoresistive biosensors," *Micromachines*, vol. 11, no. 1, 2020.

[118] C. Paolini, D. Soselia, H. Baweja, and M. Sarkar, "Optimal location for fall detection edge inferencing," in *2019 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, 2019.

[119] R. Merletti and D. Farina, *Surface electromyography: physiology, engineering, and applications*. John Wiley & Sons, 2016.

[120] C. J. De Luca, "Surface electromyography: Detection and recording," *DelSys Incorporated*, vol. 10, no. 2, pp. 1–10, 2002.

[121] T. J. M. Dick, K. Tucker, F. Hug, M. Besomi, J. H. van Dieën, R. M. Enoka, T. Besier, R. G. Carson, E. A. Clancy, C. Disselhorst-Klug, D. Falla, D. Farina, S. Gandevia, A. Holobar, M. C. Kiernan, M. Lowery, K. McGill, R. Merletti, E. Perreault, J. C. Rothwell, K. Søgaard, T. Wrigley, and P. W. Hodges, "Consensus for experimental design in electromyography (cede) project: Application of emg to estimate muscle force," *Journal of Electromyography and Kinesiology*, vol. 79, p. 102910, 2024.

[122] H. Huang *, T. A. Kuiken, and R. D. Lipschutz, "A strategy for identifying locomotion modes using surface electromyography," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 1, pp. 65–73, 2009.

[123] G. Rescio, A. Leone, and P. Siciliano, "Supervised machine learning scheme for electromyography-based pre-fall detection system," *Expert Systems with Applications*, vol. 100, pp. 95–105, 2018.

[124] R. F. Pitzalis, N. Cartocci, C. Di Natali, D. G. Caldwell, G. Berselli, and J. Ortiz, "Development of a ml-control strategy for a wrist exoskeleton based on emg and force measurements with sensor strategy optimisation," in *2024 10th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*, pp. 1446–1453, 2024.

[125] M. Azeem, M. Shahid, I. Masin, and M. Petru, "Design and development of textile-based wearable sensors for real-time biomedical monitoring; a review," *The Journal of The Textile Institute*, vol. 116, no. 1, pp. 80–95, 2025.

[126] D. Xiong, D. Zhang, X. Zhao, and Y. Zhao, "Deep learning for emg-based human-machine interaction: A review," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 3, pp. 512–533, 2021.

[127] G. D. Clifford, F. Azuaje, P. McSharry, *et al.*, *Advanced methods and tools for ECG data analysis*, vol. 10. Artech house Boston, 2006.

[128] Y. Khan, A. E. Ostfeld, C. M. Lochner, A. Pierre, and A. C. Arias, "Monitoring of vital signs with flexible and wearable medical devices," *Advanced Materials*, vol. 28, no. 22, pp. 4373–4395, 2016.

[129] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in Public Health*, vol. Volume 5 - 2017, 2017.

[130] A. Shcherbina, C. M. Mattsson, D. Waggott, H. Salisbury, J. W. Christle, T. Hastie, M. T. Wheeler, and E. A. Ashley, "Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort," *Journal of Personalized Medicine*, vol. 7, no. 2, 2017.

[131] A. Pantelopoulos and N. G. Bourbakis, "A survey on wearable sensor-based systems for health monitoring and prognosis," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 1, pp. 1–12, 2010.

[132] D. Hernando, S. Roca, J. Sancho, A. Alesanco, and R. Bailón, "Validation of the apple watch for heart rate variability measurements during relax and mental stress in healthy subjects," *Sensors*, vol. 18, no. 8, 2018.

[133] V.-R. Xefteris, A. Tsanousa, G. Meditskos, S. Vrochidis, and I. Kompatsiaris, "Performance, challenges, and limitations in multimodal fall detection systems: A review," *IEEE Sensors Journal*, vol. 21, no. 17, pp. 18398–18409, 2021.

[134] R. Bailon, L. Sornmo, and P. Laguna, "A robust method for ecg-based estimation of the respiratory frequency during stress testing," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 7, pp. 1273–1285, 2006.

[135] G. B. Moody, R. G. Mark, A. Zoccola, and S. Mantero, "Derivation of respiratory signals from multi-lead ecgs," *Computers in cardiology*, vol. 12, no. 1985, pp. 113–116, 1985.

[136] P. H. Charlton, T. Bonnici, L. Tarassenko, D. A. Clifton, R. Beale, and P. J. Watkinson, "An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram," *Physiological Measurement*, vol. 37, p. 610, mar 2016.

[137] T. Penzel, J. W. Kantelhardt, R. P. Bartsch, M. Riedl, J. F. Kraemer, N. Wessel, C. Garcia, M. Glos, I. Fietze, and C. Schöbel, "Modulations of heart rate, ecg, and cardio-respiratory coupling observed in polysomnography," *Frontiers in Physiology*, vol. Volume 7 - 2016, 2016.

[138] B. Roy, A. Roy, J. K. Chandra, and R. Gupta, "i-prext: Photoplethysmography derived respiration signal extraction and respiratory rate tracking using neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021.

[139] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Measurement*, vol. 28, p. R1, feb 2007.

[140] T. Tamura, Y. Maeda, M. Sekine, and M. Yoshida, "Wearable photoplethysmographic sensors—past and present," *Electronics*, vol. 3, no. 2, pp. 282–302, 2014.

[141] F. Scardulla, G. Cosoli, S. Spinsante, A. Poli, G. Iadarola, R. Pernice, A. Busacca, S. Pasta, L. Scalise, and L. D'Acquisto, "Photoplethysmograhic sensors, potential and limitations: Is it time for regulation? a comprehensive review," *Measurement*, vol. 218, p. 113150, 2023.

[142] J. Lee, M. Kim, H.-K. Park, and I. Y. Kim, "Motion artifact reduction in wearable photoplethysmography based on multi-channel sensors with multiple wavelengths," *Sensors*, vol. 20, no. 5, 2020.

[143] P. H. Charlton, P. A. Kyriacou, J. Mant, V. Marozas, P. Chowienczyk, and J. Alastruey, "Wearable photoplethysmography for cardiovascular monitoring," *Proceedings of the IEEE*, vol. 110, no. 3, pp. 355–381, 2022.

[144] W. Boucsein, *Electrodermal activity*. Springer science & business media, 2012.

[145] M. Benedek and C. Kaernbach, "A continuous measure of phasic electrodermal activity," *Journal of Neuroscience Methods*, vol. 190, no. 1, pp. 80–91, 2010.

[146] H. F. Posada-Quintero and K. H. Chon, "Innovations in electrodermal activity data collection and signal processing: A systematic review," *Sensors*, vol. 20, no. 2, 2020.

[147] S. Mathôt, "Pupillometry: Psychology, physiology, and function," *Journal of cognition*, vol. 1, no. 1, p. 16, 2018.

[148] J. Beatty and B. Lucero-Wagoner, *The pupillary system*, ch. 1, p. 1. Handbook of psychophysiology, 10 2012.

[149] M. Meeker, R. Du, P. Bacchetti, C. M. Privitera, *et al.*, "Pupil examination: validity and clinical utility of an automated pupillometer," *Journal of Neuroscience Nursing*, vol. 37, no. 1, p. 34, 2005.

[150] N. Unsworth and M. K. Robison, "Tracking arousal state and mind wandering with pupillometry," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 18, no. 4, pp. 638–664, 2018.

[151] M. E. Kret and E. E. Sjak-Shie, "Preprocessing pupil size data: Guidelines and code," *Behavior research methods*, vol. 51, no. 3, pp. 1336–1342, 2019.

[152] M. Slater, "Immersion and the illusion of presence in virtual reality," *British journal of psychology*, vol. 109, no. 3, pp. 431–433, 2018.

[153] K. E. Laver, S. George, S. Thomas, J. E. Deutsch, and M. Crotty, "Virtual reality for stroke rehabilitation," *Cochrane database of systematic reviews*, no. 9, 2011.

[154] E. Carl, A. T. Stein, A. Levihn-Coon, J. R. Pogue, B. Rothbaum, P. Emmelkamp, G. J. Asmundson, P. Carlbring, and M. B. Powers, "Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials," *Journal of anxiety disorders*, vol. 61, pp. 27–36, 2019.

[155] M. C. Howard, "Virtual reality interventions for personal development: A meta-analysis of hardware and software," *Human–Computer Interaction*, vol. 34, no. 3, pp. 205–239, 2019.

[156] J. R. Lewis, "Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the asq," *SIGCHI Bull.*, vol. 23, p. 78–81, Jan. 1991.

[157] J. Sauro and J. R. Lewis, *Quantifying the user experience: Practical statistics for user research*. Morgan Kaufmann, 2016.

[158] B. Albert and T. Tullis, *Measuring the user experience: Collecting, analyzing, and presenting UX metrics*. Morgan Kaufmann, 2022.

[159] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence*, vol. 7, no. 3, pp. 225–240, 1998.

[160] E. McAuley, T. Duncan, and V. V. Tammen, "Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: A confirmatory factor analysis," *Research Quarterly for Exercise and Sport*, vol. 60, no. 1, pp. 48–58, 1989. PMID: 2489825.

[161] R. M. Ryan and E. L. Deci, "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being.," *American psychologist*, vol. 55, no. 1, p. 68, 2000.

[162] E. L. Deci and R. M. Ryan, "The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior," *Psychological Inquiry*, vol. 11, no. 4, pp. 227–268, 2000.

[163] V. Monteiro, L. Mata, and F. Peixoto, "Intrinsic motivation inventory: Psychometric properties in the context of first language and mathematics learning," *Psicologia: Reflexão e Crítica*, vol. 28, p. 434–443, Jul 2015.

[164] J. Choi, T. Mogami, and A. Medalia, "Intrinsic motivation inventory: An adapted measure for schizophrenia research," *Schizophrenia Bulletin*, vol. 36, pp. 966–976, 04 2009.

[165] N. Wenk, J. Penalver-Andres, K. A. Buetler, T. Nef, R. M. Müri, and L. Marchal-Crespo, "Effect of immersive visualization technologies on cognitive load, motivation, usability, and embodiment," *Virtual Reality*, vol. 27, no. 1, pp. 307–331, 2023.

[166] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Human Mental Workload* (P. A. Hancock and N. Meshkati, eds.), vol. 52 of *Advances in Psychology*, pp. 139–183, North-Holland, 1988.

[167] J. M. Noyes and D. P. J. Bruneau, "A self-analysis of the nasa-tlx workload measure," *Ergonomics*, vol. 50, no. 4, pp. 514–519, 2007. PMID: 17575712.

[168] J. C. Byers, "Traditional and raw task load index (tlx) correlations: are paired comparisons necessary?," *Advances in industrial ergonomics and safety*, pp. 481–485, 1989.

[169] S. G. Hart, "Nasa-task load index (nasa-tlx); 20 years later," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904–908, 2006.

[170] S. Rubio, E. Díaz, J. Martín, and J. M. Puente, "Evaluation of subjective mental workload: A comparison of swat, nasa-tlx, and workload profile methods," *Applied psychology*, vol. 53, no. 1, pp. 61–86, 2004.

[171] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[172] J. D. Morris, "Observations: Sam: The self-assessment manikin—an efficient cross-cultural measurement of emotional response," *Journal of Advertising Research*, vol. 35, no. 6, pp. 63–68, 1995.

[173] P. J. Lang, "The cognitive psychophysiology of emotion: Fear and anxiety," in *Anxiety and the anxiety disorders*, pp. 131–170, Routledge, 2019.

[174] T. Xie, M. Cao, and Z. Pan, "Applying self-assessment manikin (sam) to evaluate the affective arousal effects of vr games," in *Proceedings of the 2020 3rd International Conference on Image and Graphics Processing*, ICIGP '20, (New York, NY, USA), p. 134–138, Association for Computing Machinery, 2020.

[175] S. Reiss, R. A. Peterson, D. M. Gursky, and R. J. McNally, "Anxiety sensitivity, anxiety frequency and the prediction of fearfulness," *Behaviour Research and Therapy*, vol. 24, no. 1, pp. 1–8, 1986.

[176] S. Taylor, *Anxiety sensitivity: Theory, research, and treatment of the fear of anxiety*. Routledge, 2014.

[177] B. O. Olatunji and K. B. Wolitzky-Taylor, "Anxiety sensitivity and the anxiety disorders: a meta-analytic review and synthesis.," *Psychological bulletin*, vol. 135, no. 6, p. 974, 2009.

[178] R. J. McNally, "Anxiety sensitivity and panic disorder," *Biological psychiatry*, vol. 52, no. 10, pp. 938–946, 2002.

[179] S. Taylor, M. Zvolensky, B. Cox, B. Deacon, R. Heimberg, D. Roth Ledley, J. Abramowitz, R. Holaway, B. Sandín, S. Stewart, M. Coles, W. Eng, E. Daly, W. Arrindell, M. Bouvard, and S. Jurado, "Robust dimensions of anxiety sensitivity: Development and initial validation of the anxiety sensitivity index-3," *Psychological assessment*, vol. 19, pp. 176–88, 06 2007.

[180] R. E. Zinbarg, D. H. Barlow, and T. A. Brown, "Hierarchical structure and general factor saturation of the anxiety sensitivity index: Evidence and implications.," *Psychological assessment*, vol. 9, no. 3, p. 277, 1997.

[181] M. Hirota, K. Yada, T. Morimoto, T. Endo, T. Miyoshi, S. Miyagawa, Y. Hirohara, T. Yamaguchi, M. Saika, and T. Fujikado, "Objective evaluation of visual fatigue in patients with intermittent exotropia," *PLOS ONE*, vol. 15, pp. 1–16, 03 2020.

[182] M. Hirota, T. Morimoto, H. Kanda, T. Endo, T. Miyoshi, S. Miyagawa, Y. Hirohara, T. Yamaguchi, M. Saika, and T. Fujikado, "Objective evaluation of visual fatigue using binocular fusion maintenance," *Translational Vision Science & Technology*, vol. 7, pp. 9–9, 03 2018.

[183] J. F. Knight, D. Deen-Williams, T. N. Arvanitis, C. Baber, S. Sotiriou, S. Anastopoulou, and M. Gargalakos, "Assessing the wearability of wearable computers," in *2006 10th IEEE International Symposium on Wearable Computers*, pp. 75–82, 2006.

[184] J. Sauro and E. Kindlund, "Using a single usability metric (sum) to compare the usability of competing products," in *Human Computer Interaction International Conference*, 2005.

[185] J. Brooke *et al.*, "Sus-a quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.

[186] A. Bangor, P. T. Kortum, and J. T. Miller, "An empirical evaluation of the system usability scale," *International Journal of Human–Computer Interaction*, vol. 24, no. 6, pp. 574–594, 2008.

[187] J. Brooke, "Sus: a retrospective," *J. Usability Studies*, vol. 8, p. 29–40, Feb. 2013.

[188] K. Finstad, "The usability metric for user experience," *Interacting with Computers*, vol. 22, no. 5, pp. 323–327, 2010. Modelling user experience - An agenda for research and practice.

[189] K. M. Ellis, C. Norman, A. V. der Merwe, and M. Jeon, "Increasing patient compliance and satisfaction with physical therapy web-based applications," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 57, no. 1, pp. 1531–1535, 2013.

[190] T. Schubert, F. Friedmann, and H. Regenbrecht, "The experience of presence: Factor analytic insights," *Presence: Teleoperators and Virtual Environments*, vol. 10, pp. 266–281, 06 2001.

[191] V. Schwind, P. Knierim, N. Haas, and N. Henze, "Using presence questionnaires in virtual reality," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, (New York, NY, USA), p. 1–12, Association for Computing Machinery, 2019.

[192] D. Roth and M. E. Latoschik, "Construction of the virtual embodiment questionnaire (veq)," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 12, pp. 3546–3556, 2020.

[193] M. Gonzalez-Franco and T. C. Peck, "Avatar embodiment. towards a standardized questionnaire," *Frontiers in Robotics and AI*, vol. Volume 5 - 2018, 2018.

[194] K. Kilteni, R. Groten, and M. Slater, "The sense of embodiment in virtual reality," *Presence: Teleoperators and Virtual Environments*, vol. 21, pp. 373–387, 11 2012.

[195] Y. M. Kim and I. Rhiu, "Development of a virtual reality system usability questionnaire (vrsuq)," *Applied Ergonomics*, vol. 119, p. 104319, 2024.

[196] L. Bareišytė, S. Slatman, J. Austin, M. Rosema, I. van Sintemaartensdijk, S. Watson, and C. Bode, "Questionnaires for evaluating virtual reality: A systematic scoping review," *Computers in Human Behavior Reports*, vol. 16, p. 100505, 2024.

[197] T. Mitchell, *Machine Learning*. McGraw-Hill International Editions, McGraw-Hill, 1997.

[198] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[199] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[200] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.

[201] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, vol. 1. MIT press Cambridge, 2016.

[202] Z. Ghahramani, *Unsupervised Learning*, pp. 72–112. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004.

[203] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems)*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2005.

[204] R. S. Sutton, A. G. Barto, *et al.*, *Reinforcement learning: An introduction*, vol. 1. MIT press Cambridge, 1998.

[205] Øivind Due Trier, A. K. Jain, and T. Taxt, "Feature extraction methods for character recognition-a survey," *Pattern Recognition*, vol. 29, no. 4, pp. 641–662, 1996.

[206] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'95, (San Francisco, CA, USA), p. 1137–1143, Morgan Kaufmann Publishers Inc., 1995.

[207] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, p. 2825–2830, Nov. 2011.

[208] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," *Neural Computation*, vol. 10, pp. 1895–1923, 10 1998.

[209] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," *BMC bioinformatics*, vol. 7, no. 1, p. 91, 2006.

[210] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance," *Climate research*, vol. 30, no. 1, pp. 79–82, 2005.

[211] T. Chai and R. R. Draxler, "Root mean square error (rmse) or mean absolute error (mae)? – arguments against avoiding rmse in the literature," *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247–1250, 2014.

[212] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679–688, 2006.

[213] D. L. J. Alexander, A. Tropsha, and D. A. Winkler, "Beware of r2: Simple, unambiguous assessment of the prediction accuracy of qsar and qspr models," *Journal of Chemical Information and Modeling*, vol. 55, no. 7, pp. 1316–1322, 2015. PMID: 26099013.

[214] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Comprehensive review of neural network-based prediction intervals and new advances," *IEEE Transactions on Neural Networks*, vol. 22, no. 9, pp. 1341–1356, 2011.

[215] D. Powers, "Evaluation: From precision, recall and f-measure to roc, informedness, markedness & correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.

[216] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009.

[217] W. J. Youden, "Index for rating diagnostic tests," *Cancer*, vol. 3, no. 1, pp. 32–35, 1950.

[218] M. Kubat, R. C. Holte, and S. Matwin, "Machine learning for the detection of oil spills in satellite radar images," *Machine learning*, vol. 30, no. 2, pp. 195–215, 1998.

[219] T. Fawcett, "An introduction to roc analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006. ROC Analysis in Pattern Recognition.

[220] R. G. P. Jr and K. Si, "The total operating characteristic to measure diagnostic ability for multiple thresholds," *International Journal of Geographical Information Science*, vol. 28, no. 3, pp. 570–583, 2014.

[221] R. G. Pontius Jr and B. Parmentier, "Recommendations for using the relative operating characteristic (roc)," *Landscape Ecology*, vol. 29, no. 3, pp. 367–382, 2014.

[222] G. Seber and A. Lee, *Linear Regression Analysis*, vol. 1. Wiley, 01 2012.

[223] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, pp. 267–288, 12 2018.

[224] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification And Regression Trees*. Chapman and Hall/CRC, 10 2017.

[225] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 10 2001.

[226] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189 – 1232, 2001.

[227] R. A. FISHER, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[228] D. R. Cox, "The regression analysis of binary sequences," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 20, no. 2, pp. 215–232, 1958.

[229] P. Domingos and M. Pazzani, "On the optimality of the simple bayesian classifier under zero-oneloss," *Machine Learning - ML*, vol. 29, pp. 103–130, 01 1997.

[230] C. Cortes and V. Vapnik, "Support-vector networks," *Chem. Biol. Drug Des.*, vol. 297, pp. 273–297, 01 2009.

[231] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[232] A. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *Advances in Neural Information Processing Systems* (J. Platt, D. Koller, Y. Singer, and S. Roweis, eds.), vol. 20, Curran Associates, Inc., 2007.

[233] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, (New York, NY, USA), p. 785–794, Association for Computing Machinery, 2016.

[234] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), vol. 30, Curran Associates, Inc., 2017.

[235] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems* (F. Pereira, C. Burges, L. Bottou, and K. Weinberger, eds.), vol. 25, Curran Associates, Inc., 2012.

[236] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[237] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (A. Moschitti, B. Pang, and W. Daelemans, eds.), (Doha, Qatar), pp. 1724–1734, Association for Computational Linguistics, Oct. 2014.

[238] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.

[239] B. Krawczyk, "Learning from imbalanced data: open challenges and future directions," *Progress in artificial intelligence*, vol. 5, no. 4, pp. 221–232, 2016.

[240] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[241] H. He, Y. Bai, E. A. Garcia, and S. Li, "Adasyn: Adaptive synthetic sampling approach for imbalanced learning," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 1322–1328, 2008.

[242] C. Elkan, "The foundations of cost-sensitive learning," in *Proceedings of the 17th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'01, (San Francisco, CA, USA), p. 973–978, Morgan Kaufmann Publishers Inc., 2001.

[243] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "Rusboost: A hybrid approach to alleviating class imbalance," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 40, no. 1, pp. 185–197, 2010.

[244] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, *Learning from imbalanced data sets*, vol. 10. Springer, 2018.

[245] G. Williams, K. Doughty, K. Cameron, and D. Bradley, "A smart fall and activity monitor for telecare applications," in *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol.20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No.98CH36286)*, vol. 3, pp. 1151–1154 vol.3, 1998.

[246] A. Bourke and G. Lyons, "A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor," *Medical Engineering & Physics*, vol. 30, no. 1, pp. 84–90, 2008.

[247] Q. Li, J. A. Stankovic, M. A. Hanson, A. T. Barth, J. Lach, and G. Zhou, "Accurate, fast fall detection using gyroscopes and accelerometer-derived posture information," in *2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks*, pp. 138–143, 2009.

[248] K. Adhikari, H. Bouchachia, and H. Nait-Charif, "Activity recognition for indoor fall detection using convolutional neural network," in *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*, pp. 81–84, 2017.

[249] Y. Kong, J. Huang, S. Huang, Z. Wei, and S. Wang, "Learning spatiotemporal representations for human fall detection in surveillance video," *Journal of Visual Communication and Image Representation*, vol. 59, pp. 215–230, 2019.

[250] Q. Han, H. Zhao, W. Min, H. Cui, X. Zhou, K. Zuo, and R. Liu, "A two-stream approach to fall detection with mobilevgg," *IEEE Access*, vol. 8, pp. 17556–17566, 2020.

[251] Y. Wang, K. Wu, and L. M. Ni, "Wifall: Device-free fall detection by wireless networks," *IEEE Transactions on Mobile Computing*, vol. 16, no. 2, pp. 581–594, 2017.

[252] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, and S. Li, "Rt-fall: A real-time and contactless fall detection system with commodity wifi devices," *IEEE Transactions on Mobile Computing*, vol. 16, no. 2, pp. 511–526, 2017.

[253] K.-H. Chen, Y.-W. Hsu, J.-J. Yang, and F.-S. Jaw, "Evaluating the specifications of built-in accelerometers in smartphones on fall detection performance," *Instrumentation Science & Technology*, vol. 46, no. 2, pp. 194–206, 2018.

[254] A. Shojaei-Hashemi, P. Nasiopoulos, J. J. Little, and M. T. Pourazad, "Video-based human fall detection in smart homes using deep learning," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, 2018.

[255] V. Kumar, Y. Pathak, and M. Pandey, "A comparative evaluation of machine learning models for detecting human falls using wearable sensors," in *2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, vol. 3, pp. 1–6, 2025.

[256] H. Yhdego, J. Li, C. Paolini, and M. Audette, "Wearable sensor gait analysis of fall detection using attention network," in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 3137–3141, 2021.

[257] H. Yhdego, C. Paolini, and M. Audette, "Toward real-time, robust wearable sensor fall detection using deep learning methods: A feasibility study," *Applied Sciences*, vol. 13, no. 8, 2023.

[258] Y. Cho, E. Motta, O. Nocentini, M. Lagomarsino, A. Merello, M. Crepaldi, and A. Ajoudani, "Wifi based human fall and activity recognition using transformer based encoder decoder and graph neural networks," 2025.

[259] Y. Cho, G. Solak, O. Nocentini, M. Lorenzini, A. Fortuna, and A. Ajoudani, "Anticipatory fall detection in humans with hybrid directed graph neural networks and long short-term memory," 2025.

[260] W.-J. Chang, C.-H. Hsu, and L.-B. Chen, "A pose estimation-based fall detection methodology using artificial intelligence edge computing," *IEEE Access*, vol. 9, pp. 129965–129976, 2021.

[261] B.-S. Lin, T. Yu, C.-W. Peng, C.-H. Lin, H.-K. Hsu, I.-J. Lee, and Z. Zhang, "Fall detection system with artificial intelligence-based edge computing," *IEEE Access*, vol. 10, pp. 4328–4339, 2022.

[262] K. S. Arikumar, S. B. Prathiba, M. Alazab, T. R. Gadekallu, S. Pandya, J. M. Khan, and R. S. Moorthy, "Fl-pmi: Federated learning-based person movement identification through wearable devices in smart healthcare systems," *Sensors*, vol. 22, no. 4, 2022.

[263] P. Qi, D. Chiaro, and F. Piccialli, "Fl-fd: Federated learning-based fall detection with multimodal data fusion," *Information Fusion*, vol. 99, p. 101890, 2023.

[264] E. Casilari, J.-A. Santoyo-Ramón, and J.-M. Cano-García, "Analysis of public datasets for wearable fall detection systems," *Sensors*, vol. 17, no. 7, 2017.

[265] M. Musci, D. De Martini, N. Blago, T. Facchinetti, and M. Piastra, "Online fall detection using recurrent neural networks on smart wearable devices," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 3, pp. 1276–1289, 2021.

[266] M. Poluektov and A. Polar, "Construction of the kolmogorov-arnold representation using the newton-kaczmarz method," 2025.

[267] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljacic, T. Y. Hou, and M. Tegmark, "KAN: Kolmogorov-arnold networks," in *The Thirteenth International Conference on Learning Representations*, 2025.

[268] C. J. Vaca-Rubio, L. Blanco, R. Pereira, and M. Caus, "Kolmogorov-arnold networks (kans) for time series analysis," in *2024 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, 2024.

[269] Z. Bozorgasl and H. Chen, "Wav-kan: Wavelet kolmogorov-arnold networks," 2024.

[270] D. W. Abueidda, P. Pantidis, and M. E. Mobasher, "Deepokan: Deep operator network based on kolmogorov arnold networks for mechanics problems," *Computer Methods in Applied Mechanics and Engineering*, vol. 436, p. 117699, 2025.

[271] R. Genet and H. Inzirillo, "Tkan: Temporal kolmogorov-arnold networks," 2025.

[272] M. Kiamari, M. Kiamari, and B. Krishnamachari, "Gkan: Graph kolmogorov-arnold networks," 2024.

[273] A. Kolmogorov, "On the representation of continuous functions of several variables by superpositions of continuous functions of lesser variable count," in *Dokl. Akad. Nauk SSSR*, vol. 108, 1956.

[274] A. B. Givental, B. A. Khesin, J. E. Marsden, A. N. Varchenko, V. A. Vassiliev, O. Y. Viro, and V. M. Zakalyukin, eds., *On functions of three variables*, pp. 5–8. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

[275] J. Martinez, "Solving systems of nonlinear equations by means of an accelerated successive orthogonal projections method," *J. of Computational and Applied Mathematics*, vol. 16, no. 2, pp. 169–179, 1986.

[276] J. Martínez and R. J. De Sampaio, "Parallel and sequential kaczmarz methods for solving underdetermined nonlinear equations," *J. of Computational and Applied Mathematics*, vol. 15, no. 3, pp. 311–321, 1986.

[277] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of Bioinformatics and Computational Biology*, vol. 03, no. 02, pp. 185–205, 2005.

[278] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[279] I. D. Loram and M. Lakie, "Human balancing of an inverted pendulum: position control by small, ballistic-like, throw and catch movements," *The Journal of Physiology*, vol. 540, no. 3, pp. 1111–1124, 2002.

[280] D. A. Winter, *Biomechanics and motor control of human movement*. John wiley & sons, 2009.

[281] S. L. Delp, F. C. Anderson, A. S. Arnold, P. Loan, A. Habib, C. T. John, E. Guendelman, and D. G. Thelen, "Opensim: Open-source software to create and analyze dynamic simulations of movement," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 11, pp. 1940–1950, 2007.

[282] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014.

[283] T.-T. Nguyen, M.-C. Cho, and T.-S. Lee, "Automatic fall detection using wearable biomedical signal measurement terminal," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5203–5206, 2009.

[284] S. Parvaneh, B. Najafi, N. Toosizadeh, I. B. Riaz, and J. Mohler, "Is there any association between ventricular ectopy and falls in community-dwelling older adults?," in *43rd Computing in Cardiology Conference, CinC 2016*, vol. 43, p. 433 – 436, 2016. Cited by: 2; All Open Access, Bronze Open Access.

[285] S. Suessner, N. Niklas, U. Bodenhofer, and J. Meier, "Machine learning-based prediction of fainting during blood donations using donor properties and weather data as features," *BMC Medical Informatics and Decision Making*, vol. 22, no. 1, p. 222, 2022.

[286] G. Rescio, A. Leone, L. Giampetruzzi, and P. Siciliano, *Fall Risk Assessment Using New sEMG-Based Smart Socks*, pp. 147–166. Cham: Springer International Publishing, 2021.

[287] J. Cheng, X. Chen, and M. Shen, "A framework for daily activity monitoring and fall detection based on surface electromyography and accelerometer signals," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 1, pp. 38–45, 2013.

[288] X. Xi, M. Tang, S. M. Miran, and Z. Luo, "Evaluation of feature extraction and recognition for activity monitoring and fall detection based on wearable semg sensors," *Sensors*, vol. 17, no. 6, 2017.

[289] W. Saadeh, S. A. Butt, and M. A. B. Altaf, "A patient-specific single sensor iot-based wearable fall prediction and detection system," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 5, pp. 995–1003, 2019.

[290] V. Annese and D. De Venuto, "Fpga based architecture for fall-risk assessment during gait monitoring by synchronous eeg/emg," in *2015 6th International Workshop on Advances in Sensors and Interfaces (IWASI)*, pp. 116–121, 2015.

[291] H. Jebelli, S. Hwang, and S. Lee, "Eeg-based workers' stress recognition at construction sites," *Automation in Construction*, vol. 93, pp. 315–324, 2018.

[292] J. Jeon and H. Cai, "Multi-class classification of construction hazards via cognitive states assessment using wearable eeg," *Advanced Engineering Informatics*, vol. 53, p. 101646, 2022.

[293] B. M. Tehrani, J. Wang, and D. Truax, "Assessment of mental fatigue using electroencephalography (eeg) and virtual reality (vr) for construction fall hazard prevention," *Engineering, Construction and Architectural Management*, vol. 29, pp. 3593–3616, 08 2021.

[294] V. J. Kartsch, S. Benatti, P. D. Schiavone, D. Rossi, and L. Benini, "A sensor fusion approach for drowsiness detection in wearable ultra-low-power systems," *Information Fusion*, vol. 43, pp. 66–76, 2018.

[295] A. M. A. Handojoseno, J. M. Shine, T. N. Nguyen, Y. Tran, S. J. G. Lewis, and H. T. Nguyen, "Analysis and prediction of the freezing of gait using eeg brain dynamics," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 5, pp. 887–896, 2015.

[296] R. Rajagopalan, I. Litvan, and T.-P. Jung, "Fall prediction and prevention systems: Recent trends, challenges, and future research directions," *Sensors*, vol. 17, no. 11, 2017.

[297] R. N. Ferreira, N. F. Ribeiro, and C. P. Santos, "Fall risk assessment using wearable sensors: A narrative review," *Sensors*, vol. 22, no. 3, 2022.

[298] C.-T. C. for Construction Research, *The construction chart book: The US construction industry and its workers*. CPWR-The Center for Construction Research and Training, 2025.

[299] J. Tian, P. Mercier, and C. Paolini, "Ultra low-power, wearable, accelerated shallow-learning fall detection for elderly at-risk persons," *Smart Health*, vol. 33, p. 100498, 2024.

[300] H. Iwasawa, M. Nomura, N. Sakitani, K. Watanabe, D. Watanabe, and H. Moriyama, "Stretching after heat but not after cold decreases contractures after spinal cord injury in rats," *Clinical Orthopaedics and Related Research*®, vol. 474, no. 12, pp. 2692–2701, 2016.

[301] M. Mononen, "Ecg logger for polar h10."

[302] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, ICMI '18, (New York, NY, USA), p. 400–408, Association for Computing Machinery, 2018.

[303] L. R. Altimari, J. L. Dantas, M. Bigliassi, T. F. D. Kanthack, A. C. de Moraes, and T. Abrao, "Influence of different strategies of treatment muscle contraction and relaxation phases on emg signal processing and analysis during cyclic exercise," in *Computational Intelligence in Electromyography Analysis - A Perspective on Current Applications and Future Challenges* (G. R. Naik, ed.), ch. 5, London: IntechOpen, 2012.

[304] D. H. Spodick, "Survey of selected cardiologists for an operational definition of normal sinus heart rate," *The American journal of cardiology*, vol. 72, no. 5, pp. 487–488, 1993.

[305] S. M. Fox and J. P. Naughton, "Physical activity and the prevention of coronary heart disease," *Preventive Medicine*, vol. 1, no. 1, pp. 92–120, 1972.

[306] C. Ozemek, *ACSM's guidelines for exercise testing and prescription*. Lippincott Williams & Wilkins, 2025.

[307] C. Chourpiliadis and A. Bhardwaj, *Physiology, Respiratory Rate*, ch. 1, p. 1. StatPearls Publishing, 01 2019.

[308] E. L. Foundation, "Your lungs and exercise," *Breathe*, vol. 12, no. 1, pp. 97–100, 2016.

[309] S. Sguazza, A. Puiatti, S. Bernaschina, F. Faraci, G. Ramelli, V. D'Apuzzo, E. Rossini, and M. Papandrea, "Sensor data synchronization in a iot environment for infants motricity measurement," in *IoT Technologies for HealthCare* (N. M. Garcia, I. M. Pires, and R. Goleva, eds.), (Cham), pp. 3–21, Springer International Publishing, 2020.

[310] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biological Psychology*, vol. 84, no. 3, pp. 394–421, 2010. The biopsychology of emotion: Current theoretical and empirical perspectives.

[311] M. P. Tarvainen, J.-P. Niskanen, J. A. Lipponen, P. O. Ranta-aho, and P. A. Karjalainen, "Kubios hrv – heart rate variability analysis software," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 1, pp. 210–220, 2014.

[312] M. Malik, J. T. Bigger, A. J. Camm, R. E. Kleiger, A. Malliani, A. J. Moss, and P. J. Schwartz, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *European Heart Journal*, vol. 17, pp. 354–381, 03 1996.

[313] J. Choi, B. Ahmed, and R. Gutierrez-Osuna, "Development and evaluation of an ambulatory stress monitor based on wearable sensors," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 2, pp. 279–286, 2012.

[314] J. Healey and R. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 156–166, 2005.

[315] L.-M. Vortmann and F. Putze, "Combining implicit and explicit feature extraction for eye tracking: Attention classification using a heterogeneous input," *Sensors*, vol. 21, no. 24, 2021.

[316] E. Rey-Becerra, L. H. Barrero, R. Ellegast, and A. Kluge, "The effectiveness of virtual safety training in work at heights: A literature review," *Applied Ergonomics*, vol. 94, p. 103419, 2021.

[317] M. Cyma-Wejchenig, J. Tarnas, K. Marciniak, and R. Stemplewski, "The influence of proprioceptive training with the use of virtual reality on postural stability of workers working at height," *Sensors*, vol. 20, no. 13, 2020.

[318] H. Chander, A. Shojaei, S. Deb, S. N. K. K. Arachchige, C. Hudson, A. C. Knight, and D. W. Carruth, "Impact of virtual reality–generated construction environments at different heights on postural stability and fall risk," *Workplace Health & Safety*, vol. 69, no. 1, pp. 32–40, 2021. PMID: 32812846.

[319] C. Di Loreto, J.-R. Chardonnet, J. Ryard, and A. Rousseau, "Woah: A virtual reality work-at-height simulator," in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 281–288, 2018.

[320] S. Rokooei, A. Shojaei, A. Alvanchi, R. Azad, and N. Didehvar, "Virtual reality application for construction safety training," *Safety Science*, vol. 157, p. 105925, 2023.

[321] M. I. Al-Khiami and M. Jaeger, "Safer working at heights: Exploring the usability of virtual reality for construction safety training among blue-collar workers in kuwait," *Safety*, vol. 9, no. 3, 2023.

[322] M. Melo, G. Gonçalves, j. Vasconcelos-Raposo, and M. Bessa, "How much presence is enough? qualitative scales for interpreting the igroup presence questionnaire score," *IEEE Access*, vol. 11, pp. 24675–24685, 2023.

[323] M. Martínez Gutiérrez and J. Rafael Rojano-Cáceres, "Interpretation of the sus questionnaire in mexican sign language to evaluate usability an approach," in *2020 3rd International Conference of Inclusive Technology and Education (CONTIE)*, pp. 180–183, 2020.

Appendix A

Sample questionnaires

Self-Assessment Manikin (SAM)

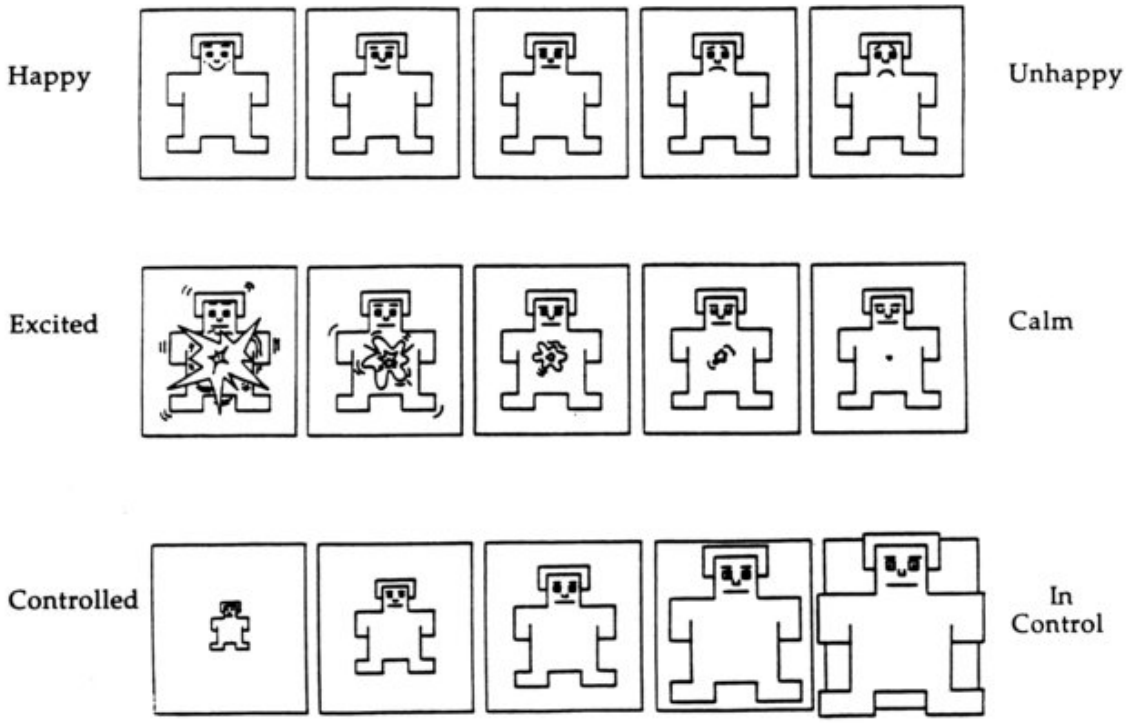


Figure A.1 The Self-Assessment Manikin (SAM) measurement scales (valence, arousal, and dominance; pole labels added).

After-Scenario Questionnaire (ASQ)

Study		Date	
Condition		Participant ID	

AFTER SCENARIO QUESTIONNAIRE (ASQ)

Lewis (1991) • Prepared by the Inclusive Interaction Lab

For each of the questions below, circle the answer of your choice.

1. Overall, I am satisfied with the ease of completing the tasks in this scenario.

1	2	3	4	5	6	7		NA
Strongly Agree						*Strongly Disagree*		*Not applicable*

Comments:

2. Overall, I am satisfied with the amount of time it took to complete the tasks in this scenario.

1	2	3	4	5	6	7		NA
Strongly Agree						*Strongly Disagree*		*Not applicable*

Comments:

3. Overall, I am satisfied with the support information (on-line help, messages, documentation) when completing the tasks?

1	2	3	4	5	6	7		NA
Strongly Agree						*Strongly Disagree*		*Not applicable*

Comments:

1 of 1

Figure A.2 UC Merced's version of the ASQ.

NASA Task Load Index (NASA-TLX)

Figure 8.6

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

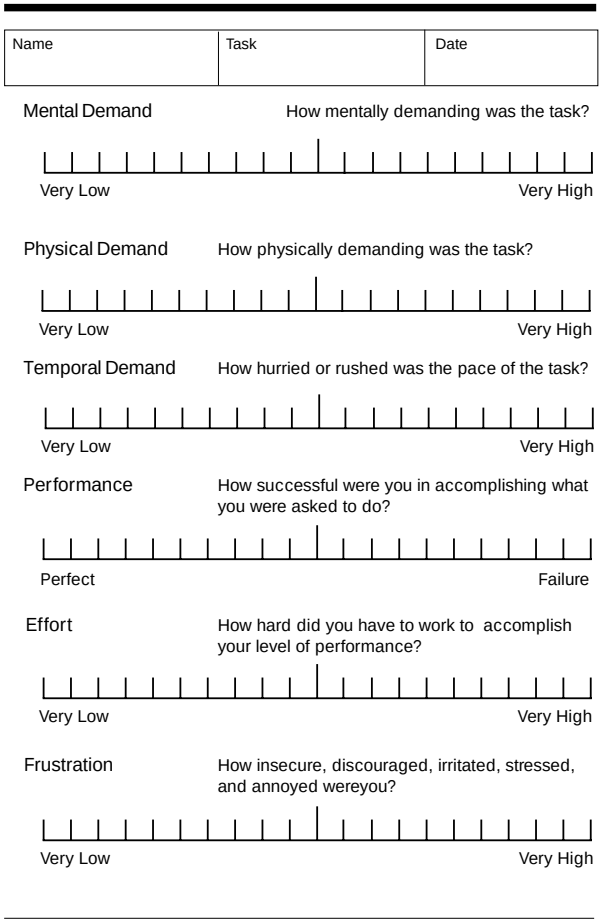


Figure A.3 Paper-and-pencil version of the NASA-TLX rating scale.

Anxiety Sensitivity Index (ASI)

Subject ID _____ Session _____ Study _____ Date ____/____/____

ASI-3

Enter the number from the scale below that best describes how typical or characteristic each of the 16 items is of ***you***, putting the number next to the item. You should make your ratings in terms of how much you agree or disagree with the statement as a ***general*** description of yourself.

0	1	2	3	4
very little	a little	some	much	very much

1. It is important for me not to appear nervous.
2. When I cannot keep my mind on a task, I worry that I might be going crazy.
3. It scares me when my heart beats rapidly.
4. When my stomach is upset, I worry that I might be seriously ill.
5. It scares me when I am unable to keep my mind on a task.
6. When I tremble in the presence of others, I fear what people might think of me.
7. When my chest feels tight, I get scared that I won't be able to breathe properly.
8. When I feel pain in my chest, I worry that I'm going to have a heart attack.
9. I worry that other people will notice my anxiety.
10. When I feel "spacey" or spaced out I worry that I may be mentally ill.
11. It scares me when I blush in front of people.
12. When I notice my heart skipping a beat, I worry that there is something seriously wrong with me.
13. When I begin to sweat in a social situation, I fear people will think negatively of me.
14. When my thoughts seem to speed up, I worry that I might be going crazy.
15. When my throat feels tight, I worry that I could choke to death.
16. When I have trouble thinking clearly, I worry that there is something wrong with me.
17. I think it would be horrible for me to faint in public.
18. When my mind goes blank, I worry there is something terribly wrong with me.

Figure A.4 Third version of the ASI.