Cycle XXXVII

A thesis presented for the degree of Doctor of Philosophy in Science and Technology for Electronic and Telecommunications Engineering

Integrated Biomedical Signal and Image Processing Techniques for Enhanced Disease Diagnosis and Clinical Decision Support

Supervisor:
Prof. Silvana Dellepiane

Candidate:
Giulia Iaconi

Coordinator of the PhD Course:
Prof. Maurizio Valle

Department of Electric, Electronic, and Telecommunication Engineering and Naval Architecture - Università degli Studi di Genova

b

Preface

The present thesis is prepared in fulfillment of the requirements for acquiring the degree of Doctor of Philosophy in Science and Technology for Electronic and Telecommunications Engineering.

The PhD in Science and Technology for Electronic and Telecommunications Engineering, is promoted by the Department of Electrical, Electronic, and Telecommunication Engineering and Naval Architecture (DITEN) - Università degli Studi di Genova.

Abstract

Human life has an invaluable worth, and its protection has always been a central priority for healthcare systems.

Recent advances in computerized medical image reconstruction, together with developments in analysis techniques and computer-assisted diagnosis, have significantly enhanced medical imaging, enabling more accurate diagnoses and more targeted treatment strategies.

In this context, Internet of Medical Things solutions, such as telerehabilitation, provide concrete answers to many healthcare needs, allowing remote patient support and ensuring continuity of care. The natural progression of this path calls for closer collaboration between clinicians and biomedical engineers, with the goal of developing innovative solutions and helping to transform the future of healthcare.

Digital processing of biomedical signals and images, once confined to research laboratories, is now an essential resource in medical applications such as early disease detection, monitoring, and treatment planning. The aim of this work is to integrate signal and image processing techniques, together with machine learning algorithms, into diagnostic practice and telerehabilitation pathways.

Both research directions are introduced simultaneously. Once the technical aspects have been outlined, their practical implementation is discussed through examples related to selected diseases.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The historical roots of digital signal processing (DSP) are profound, intertwining with ancient practices and evolving through significant technological advancements. According to Prandoni and Vetterli [2], they date back to the $25th$ century BC and they are related to the *Palermo stone* with earliest records of Nile's floods observed on the time base of 12 months (naive sampling). Processing of these records was concentrated to prediction of floods fundamental for watering fields. *Nilometers* used later were arabic buildings with instruments measuring the water level to predict climate conditions and to calculate taxes related to the prosperity of the country. The mathematical fundamentals of digital signal and image processing methods are based upon numerical analysis that predates the invention of modern computers by many centuries.

Over time, the convergence of mathematical theory, early computing, and dedicated electronic systems marked the official birth of DSP. These elements have fueled the ongoing digital transformation of society, profoundly influencing technological progress in multiple domains.

In today's fully connected world, DSP plays a central role in enhancing the quality and reliability of collected data, while also enabling efficient compression and transmission of information. Such capabilities have fostered the development of increasingly advanced systems for communication and storage. Among the most impactful application fields is healthcare, where DSP has proven to be one of the most versatile tools for ensuring safe, effective, and personalized patient care, spanning from diagnosis to follow-up. Rapid advances in medical image reconstruction and computer-aided diagnosis have boosted medical imaging into one of the most dynamic scientific fields, where key challenges include improving image quality, automating analysis, and refining predictive tools. At the same time, the growing adoption of Internet of Medical Things (IoMT) solutions has introduced new opportunities for patient monitoring and rehabilitation, allowing clinicians to design and continuously update personalized care plans.

This research project stems from the collaboration between the medical-clinical and engineering domains, with the overarching goal of leveraging DSP to support diagnosis and telerehabilitation.

The first research line focuses on ultrasound image analysis for hepatic fibrosis staging. The objective is to design a simple yet effective neural model capable of supervised learning based on multiple patient acquisitions. The proposed method integrates traditional image processing techniques during a preprocessing stage with machine learning algorithms for classification.

The second research line presents ReMoVES, an IoMT system developed at the DITEN Department of the University of Genoa [3] and lies within the framework of the STORMS project, dedicated to Multiple Sclerosis rehabilitation support.

ReMoVES employs a markerless approach to acquire multidimensional signals representing patients' joint movements, enabling the remote monitoring of prescribed exercises. To support this application, methods for signal segmentation and analysis are proposed.

In both contexts, signals are directly acquired from real-world sources: infrared sensors for motion analysis and ultrasound systems for imaging. However, such signals are affected by noise, variability, and nonlinear dynamics, which complicate their processing and interpretation.

The innovative contribution of this work lies in its strategy to address these challenges by working with contextualized signals, applying adaptive and nonlinear signal and image processing, segmentation methods, attentional focus mechanisms and artificial intelligence algorithms. Integrating these approaches improves the accuracy and efficiency of clinical decision-making, biomedical applications and telerehabilitation platforms, paving the way for more robust and personalized healthcare solutions.

A more detailed overview of aims of the thesis can be described as follows:

- To perform a comprehensive State of the Art (SoA) analysis of existing methodologies for the management, analysis, and interoperability of rehabilitation and diagnostic data, aiming to improve data integration and clinical workflow efficiency;
- To integrate the tele-rehabilitation system ReMoVES, for the remote treatment and monitoring of multiple sclerosis and its neurological conditions;
- To establish a generalized and adaptable data processing pipeline for clinical practice, incorporating signal and image processing techniques, to support enhanced medical decision-making, patient management, and remote therapeutic monitoring;
- To conduct an in-depth analysis of multidimensional and multiparametric signals acquired from 25 joints using the Kinect sensor [4] during the execution of motor-cognitive exercises, performed without the direct supervision of the physiotherapist, as a tool to support rehabilitation;
- To extract clinically significant parameters from these signals, considering their intrinsic noise and non-linearity, by applying adaptive and non-linear signal processing techniques followed by segmentation procedures, within a feasibility study framework;
- To perform a comprehensive SoA analysis of current clinical practices and technological solutions for the diagnosis and management of chronic liver disease;

- To analyze $2D$ biomedical image signals for diagnostic support applications, ensuring that the developed system functions solely as an assistive tool while preserving the indispensable role and clinical judgement of the specialist physician;
- To apply adaptive and non-linear image processing methodologies, incorporating focus-of-attention mechanisms for the intelligent contextualization, segmentation, and interpretation of biomedical images;
- To design and implement a supervised diagnostic framework that combines traditional image processing approaches with contemporary machine learning techniques;
- To apply the developed methodologies within the clinical context of chronic hepatic disease, specifically targeting the automated classification of liver fibrosis based on biomedical imaging data.

In Chapter 2, the SoA revision is presented, with respect to both research topics addressed in the study. In particular, the importance of images in the biomedical context is introduced, with a focus on the US and image-derived data processing, explaining the main processing techniques currently employed and the issues still open. The context of ICT solutions in healthcare is then analyzed, with a focus on signals from rehabilitation and assistive technologies. This is followed by a description of the pathologies under study, with an overview of their clinical framing. Finally, the integration of advanced computational techniques, such as machine learning alghoritm, within robust and scalable frameworks suitable for clinical use is presented. In this case too, there will be a focus on CNNs as the protagonists of the work. Chapters 3 and 4 provide an in-depth discussion of the methods and materials developed during the PhD course, outlining the main contributions of this work in diagnostic imaging and telerehabilitation, respectively. Their application to real-world scenarios is explored in Chapters 5 and 6, with a particular focus on the diseases introduced in Chapter 2. Finally, Chapter 7 offers both a discussion and insights into potential future developments of this research.

Chapter 2

State of the Art

2.1 Image Domain: Medical Image Processing

The role of images in the biomedical field is unquestionable. They provide visual representations of specific properties of the human body, generated by different types of scanners that exploit different physical principles. Medical images rely on varying shades of gray or colors to encode and transmit information. According to [5], one can define:

- *Image:* a representation that depicts a map of a measured property.
- *Medical Image:* an image generated to represent a property of the body or of an organ for medical purposes.

When an object is projected into the *Image Domain*, the significance lies less in the amount of information contained, and more in the quality and clarity with which it is conveyed. Different imaging modalities exhibit distinct characteristics and applications, offering clinicians a wide range of tools for accurate and reliable examination. Broadly, these modalities can be divided into two categories: anatomical imaging, primarily focused on organ morphology (X-ray, CT, MRI, US), and functional imaging, aimed at capturing metabolic or physiological activity (SPECT, PET, fMRI). In the present discussion, the main focus is on US, as its central role in the research project and will be covered in a dedicated section.

Image acquisition is the first step in digital image processing. In this step we get the image in digital form. Many different imaging techniques have been developed and are in clinical use. The type of image under analysis consistently affect the processing to be performed. Simplifying, we can say that the aim of medical image processing is to extract and quantify key features from imaging data, enabling clinicians to obtain a comprehensive understanding of underlying pathologies and make more informed decisions. Furthermore, these processed images often serve as inputs for other automated techniques such as image segmentation, classification, and disease progression modeling. Medical image processing deals with the development of problem-specific approaches to the enhancement of raw medical image data for the purposes of selective visualization as well as further analysis [6].

2.1.1 Ultrasound Imagery

Ultrasound is a widely used medical imaging tool valued for its non-invasiveness, safety to the human body, portability, accuracy, and cost-effectiveness. However, the quality of medical ultrasound, defined by image resolution and contrast, can be limited by several factors, both physical factors related to image acquisition and imperfections in the design of the imaging system.

Ultrasound is a mechanical wave with frequency higher than the acoustic range ; in clinical applications typical range of frequency is 1-15 MHz. The waves are generated by a transducers (or an array of transducers). As the US passes though the tissues, small part of its energy is back-scattered to the receivers due principally to the boundaries between different tissues. Based on the delay between the transmitted and received wave is possible to retrieve the location of a tissue (eventually also lesions). This technique share basically the same characteristic of a sonar. Due to the relatively high ultrasound speed of propagation, is possible to retrieve an entire image in fractions of second allowing real-time exams.

The US received signal is modified during its part depending from acoustic characteristic of the object under investigation. More in detail, speed of propagation, density and attenuation coefficient are properties specific for each medium and a wave that is propagating is affected by the change of one or more of this tissues properties.

Over the years, ultrasound has found application in numerous specialist fields: from obstetrics to cardiology [7], from abdominal diagnostics to the study of musculoskeletal [8] and vascular tissues and have also been used for many years for the evaluation of patients with CLD due to their ease of use and wide availability [9]. Conventional ultrasound in B-mode (Brightness mode) remains the most widely used imaging mode, providing two-dimensional grayscale images based on the intensity of reflected echoes. However, modern ultrasound includes a number of technological extensions, including Doppler [10] for blood flow evaluation, elastography [11] for measuring tissue stiffness and contrast echography (CEUS) [12] for the exploration of microvascular perfusion, particularly useful in oncology and the management of chronic liver disease. So, different structures are visualized using a grayscale imaging mode.

Despite the advantages, ultrasound images are known to suffer from structural limitations that compromise their quality and readability. In particular, the presence of speckle noise - a granular disturbance due to the constructive and destructive interference of reflected echoes - reduces contrast and definition of anatomical contours. To this is added a relatively low contrast resolution in the deep regions and a significant dependence on operator experience, which directly affects the quality and interpretation of acquired images.

In recent years, a number of techniques have been developed to address these challenges. Among these, speckle noise reduction algorithms based on median filters, anisotropic scattering and wavelet transforms have significantly improved

the visual quality of images [13] or to consider compound imaging that exploit different B-scan acquisition with different points of view [14]. At the same time, methodologies have been proposed for increasing contrast and for automatic segmentation of structures and lesions, using both traditional approaches (such as thresholds or active contours) and based on machine learning and deep learning techniques.

The integration of artificial intelligence (AI) and, in particular, deep learning has been a turning point in recent research on ultrasound imaging. Convolutional neural networks (CNN) and attention-based models have been successfully applied to automatic ultrasound image analysis, demonstrating significant potential in automated classification of lesions, in organ segmentation and noise reduction. These techniques have found application in specific clinical contexts, such as the staging of liver fibrosis, the characterization of thyroid nodules and the diagnosis of breast pathologies.

However, some significant weaknesses remain. The inherent variability of ultrasound images, due to different patient morphology, device settings and acquisition technique, is a challenge for the standardization and reliability of automated systems. In addition, the availability of high quality annotated datasets remains limited, hampering the development of generalizable and effective models in heterogeneous clinical contexts.

Future prospects focus on the integration of standardized acquisition protocols, multimodal data combination (for example integrating ultrasound with clinical, laboratory or other imaging data), and the development of real-time artificial intelligence-assisted ultrasound systems. Particular interest is given to the definition of new quantitative biomarkers derived from ultrasound images, which can contribute to more personalized and data-driven medicine.

2.1.2 Image-derived Data Processing

Image processing and analysis in pathology is a multidisciplinary area of research and development. In recent years, biomedical image processing has experienced unprecedented growth, emerging as a highly interdisciplinary research domain at the intersection of applied mathematics, computer science, engineering, physics, biology, and medicine. The integration of imaging technologies into clinical workflows has transformed diagnostic procedures, enabling clinicians to visualize, quantify, and interpret complex biological phenomena with remarkable precision [6].

Computer-aided diagnosis (CAD) is a vital computational method in the medical field, enhancing the efficiency and performance of radiologists, particularly in terms of sensitivity rate ([15]). Current research focuses on developing medical imaging and analysis systems that employ AI techniques and digital image processing tools. These systems aim to identify and classify abnormal features in medical images and provide visual confirmations to radiologists ([16]).

To have better services and better identification of diseases, it is necessary

to double medical images such as magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography (PET), ultrasound and advanced microscopy. But some of these images contain low intensity and contrast which is difficult for doctors to diagnose diseases by observing that the images are unclear. At the time of surgery, it is very important to identify problems. For this reason, the improvement of image intensity and edge detection of objects of interest is a concern in the medical sector. For diagnostic and research applications, image extraction is required. Correct edge detection and image optimization are important steps to this end. pixel, local and global are the three main features of the feature extraction process for an image, where the visual image feature is mainly based on the pixel value. On the other hand, edge detection reduces the amount of data and filters out unnecessary data in an image [17]. The improvement of image greatly increases man's visual perception. The combined effect of these two processes in a medical image can be of great help in the field of medicine. So, various moderns researches [18, 19, 20] are going on to improve the medical imaging system.

It should also be borne in mind that the increase in imaging modes has introduced new challenges in processing and interpreting huge volumes of heterogeneous imaging data. These data sets, often multimodal and multidimensional, require computational methods to extract meaningful information for disease diagnosis, treatment planning and monitoring.

The image enhancement process consists of several techniques used to enhance the visual appearance of an image or to transform it so that it can be better represented by a human being or a machine. Image processing has two main objectives. The first is that it will be useful for both people and doctors to identify patients' problems by looking at noise-free images, while the second is that the computer can calculate and process data faster than the human brain. The detection of the edges of an image is the most important process in image processing, as these edges contain the fundamental characteristics of the image. The Sobel operator is a well-established method of edge detection [21]. Islam and Mondal [22] perform a combination of techniques in which the Laplacian and Sobel operators and power transformation has been used to achieve a better result.

Image contrast enhancement is also used for a variety of applications. Most images, such as medical images, remote sensing images, aerial images and real life photographs, have poor contrast. Again, the main purpose of contrast enhancement is to improve image quality or clarity or enhance interpretability for human vision. Neighborhood operations are those that combine a small area or neighborhood of pixels to generate an output pixel. The average filter is used for smoothing image. Bilateral retinex and sigmoid function are both simple techniques in which contrast of an image is enhanced [23]. A critical review is used to classify image enhancement techniques into two categories: techniques are like spatial domain enhancement and frequency domain enhancement [24]. Ganesan et al. in [25] presented a hybrid approach to contrast enhancement

for a medical image. Applying an efficient approach for an adaptive anisotropic diffusion algorithm, color images and medical images are enhanced using a new edge –stopping function for an efficient adaptive anisotropic diffusion algorithm to improve the performance of the an efficient adaptive anisotropic diffusion filter [26]. Another study, propose a fast algorithm to increase the contrast of an image locally using singular value decomposition (SVD) approach and attempt to define some parameters which can give clues related to the progress of the enhancement process.[27]

Despite significant progress, biomedical image processing continues to face critical challenges [6], including image enhancement and restoration, automated and accurate segmentation of features of interest, classification of image features, namely characterization and typing of structures, quantitative measurement of image features and an interpretation of the measurements, development of integrated systems for the clinical sector, limited annotated datasets, and the need for real-time, clinically interpretable outputs. Addressing these obstacles requires the integration of advanced computational techniques—such as deep learning, image registration, segmentation, and multimodal data fusion—within robust, scalable frameworks suitable for clinical deployment.

2.1.3 Main Traditional Feature-based Techniques for Liver Staging

B-mode ultrasound is one of the most widely used tools for assessing liver disease, thanks to its non-invasive nature, reduced operational complexity, and widespread availability in clinical settings [28]. However, staging liver disease using traditional feature-based methods presents numerous challenges, particularly in the early stages of fibrosis. One of the main challenges inherent in ultrasound imaging is the presence of speckle noise, which is not a direct representation of tissue microstructure but results from the interference of backscattered signals from multiple scatterers. Despite this, the local distribution of speckle intensity is related to the echogenicity of the underlying tissue and, in the case of the liver, partly reflects the organization of the hepatic lobules.

Numerous studies have shown that the speckle pattern varies in the presence of fibrosis, suggesting the possibility of detecting microarchitectural alterations that are not immediately apparent through visual observation of conventional B-mode images. In this context, quantitative analysis approaches of the acoustic structure, such as Acoustic Structure Quantification (ASQ), have been proposed to statistically model ultrasound signal deviations associated with diffuse pathological processes [29]. However, the assessment of liver parenchyma structure remains largely subjective and characterized by limited sensitivity, particularly for the detection of cirrhosis. Retrospective studies have also shown that conventional ultrasound is not a reliable predictor of early or intermediate stages of fibrosis, especially in patients with chronic viral hepatitis [30].

Diagnostic accuracy can be partially improved by combining echotexture

analysis with additional morphological features, such as liver surface nodularity and liver margin configuration [31, 32]. In particular, surface nodularity has been shown to be more accurate than echotexture alone in the diagnosis of cirrhosis, achieving high sensitivity and specificity values. However, this approach is mainly effective in advanced stages of the disease, limiting its usefulness in early staging. Better visualization of the liver surface can be achieved in the presence of ascites, which provides a fluid-tissue interface favorable for morphological analysis [33]. However, the appearance of ascites is generally associated with already advanced cirrhosis, reducing the diagnostic value of this condition in screening or early diagnosis contexts.

In the absence of ascites, the use of the lumen of the hepatic veins as an internal fluid-tissue interface has been proposed, hypothesizing that the architectural distortion induced by cirrhosis is reflected in alterations in vascular morphology [34]. Although preliminary studies have suggested high accuracy for this approach, subsequent results have shown lower sensitivity than hepatic surface nodularity. These discrepancies are probably due to differences in the acquisition technique and the strong dependence on the operator, which limit the reproducibility of the results.

Further traditional approaches include the use of Doppler ultrasound for the evaluation of hepatic hemodynamic changes associated with the progression of fibrosis and cirrhosis [35]. Although the theoretical logic is based on alterations in intrahepatic and portal blood flow, Doppler-derived indices have shown poor reproducibility and a weak correlation with disease stage. Furthermore, parameters such as portal vein diameter are characterized by high interindividual variability and are strongly influenced by technical and respiratory factors, making it difficult to acquire reliable and repeatable measurements.

A further advancement is represented by contrast-enhanced ultrasound (CEUS), which uses microbubbles as kinetic tracers for the evaluation of hepatic perfusion [36]. Ultrasound contrast agents have a good safety profile and allow for a dynamic analysis of the microcirculation; however, the intrinsic differences between the various agents limit their interchangeability and increase their application complexity and costs, reducing their large-scale adoption.

Finally, the growing evidence of a correlation between hepatic parenchymal pathology and tissue stiffness has led to the development and dissemination of ultrasound elastography techniques, such as transient elastography, acoustic radiation force impulse imaging (ARFI), shear wave elastography and strain elastography [37, 38]. These approaches provide a quantitative estimate of liver stiffness and have demonstrated good accuracy in the assessment of fibrosis [39, 40]. However, elastographic techniques are also subject to limitations, including the influence of confounding factors and variability between different technologies and systems.

Overall, while traditional feature-based ultrasound techniques offer advantages in terms of accessibility and safety, they suffer from significant limitations in terms of accuracy, robustness, and ability to detect pathological changes early.

These challenges have progressively driven research toward advanced quantitative methodologies and approaches based on machine learning and deep learning, with the aim of overcoming the limitations of conventional approaches and improving the noninvasive staging of liver diseases [41, 42, 43, 44].

2.1.4 The AI Revolution in Medical Imaging

Medical image classification plays an essential role in clinical treatment, as doctors often need to use medical images to help diagnose whether there are diseases in the human body and to quantify the severity of diseases [45]. Before the emergence of deep learning, handcrafted features such as color, texture, image shape, etc., were widely used in medical image classification [46]. After feature extraction and selection, traditional machine learning methods such as Support Vector Machines (SVM), Logistic Regression (LR), Random Forests (RF), etc. were used for classification [47]. However, manual feature extraction and selection requires a lot of time and effort. Furthermore, the artificial feature-based method has reached its limit in terms of performance [48]. Deep learning technology, particularly Convolutional Neural Networks (CNNs), is an emerging machine learning method that has proven its potential for different classification tasks [49]. Notably, the CNNs dominates with the best results on varying biomedical image classification tasks [50, 51] and for study of different organs such a liver [52, 51, 53, 54]. They are an effective tool for automatic feature extraction, as highlighted in the pioneering work of LeCun et al. [55]. These architectures allow discriminative representations to be learned and selected automatically directly from the input data, integrating feature extraction and classification processes into a single pipeline [56]. This approach reduces dependence on manual feature engineering procedures, which are often complex and computationally expensive, and can support the diagnostic process by providing results in a short time. In general, CNN-based methods show superior performance compared to traditional approaches, thanks to their ability to effectively learn relevant features from raw data. For these reasons, CNNs, as a fundamental component of deep learning, have been widely adopted in numerous computer vision problems [57], with particular relevance in the field of medical imaging.

CNNs are specialized ANNs that are designed to solve pattern recognition tasks via machine learning (Figure 2.1). Thus, rather than receiving scalar input, as with dense networks, CNNs receive matrix input, such as images. The basis for modern CNNs was laid by the neocognitron by Fukushima in 1980 [58] and the time delay neural networks by Waibel in 1987 [59]. One of the first widely recognized networks was LeNet, a CNN for the recognition of postal zip codes, designed by LeCun et al. in 1989 [60]. Since then, several variations of CNN architecture have emerged (AlexNet [61], VGGNet [62], GoogLeNet [63], ResNet [64]) with the introduction of new datasets, such as MNIST and CIFAR-10, and competitions such as ILSVRC (ImageNet Large Scale Visual Recognition Challenge). CNNs are composed of three main components: con-

volutional, downsampling/pooling and dense layers. In contrast to dense layers, convolutional layers perform convolution, which means each neuron calculates weighted sums of a predefined set of inputs for each input rather than forming a weighted sum for all inputs. The size and weighting of the area is defined by a convolution kernel, which is shared between all neurons of a layer. This allows convolutional layers to perform image processing tasks, such as edge and corner detection. Per convolutional layer, multiple convolution kernels are trained to perform different processing tasks. To reduce the input dimensionality as well as to abstract it, each convolutional layer is followed by a downsampling layer. Whilst different methods for pooling are available, the most commonly used is maximum pooling, where the maximum of the specified area is used as the output. In addition, reducing the output dimension of a convolutional layer and thus subsequently the complexity of the network, it can also help to prevent overfitting by reducing the availability of raw input information. To be compatible with the dense part of the network, the output of the last downsampling layer is vectorized before passing. Although CNN has been successfully applied in many computer vision tasks, its application in medical image classification still faces many difficulties, such as unbalanced sample categories, small data sets, small structural lesions, poor model generalization ability, lack of explainability, etc. Furthermore, their image segmentation capabilities are limited: due to their dense layer, a convolutional network can output a certainty if an object is contained in an image but not where the object is located. Another limitation is the detection of multiple different objects in the same image.

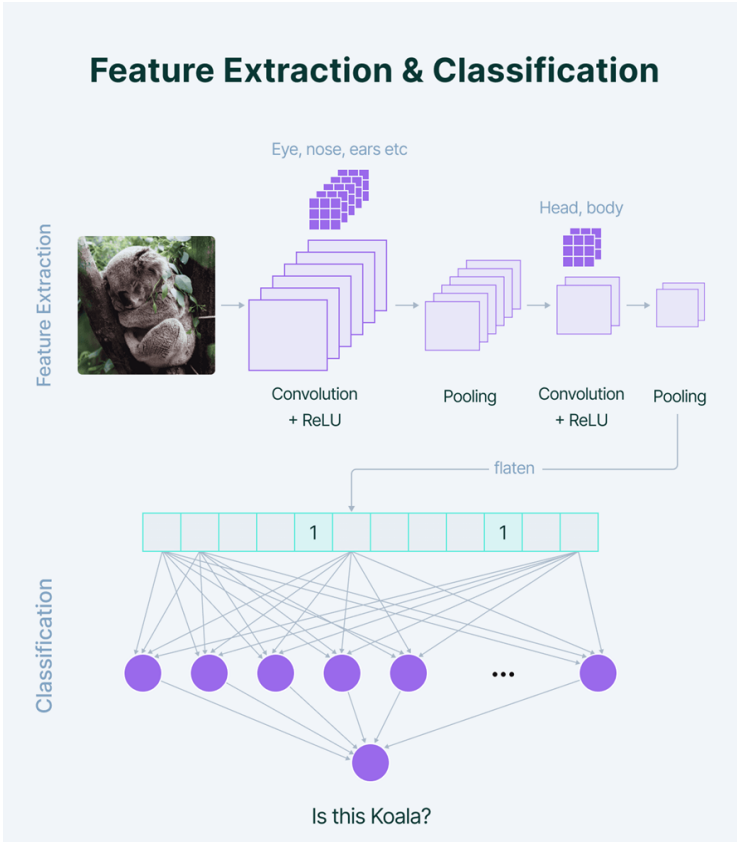


Figure 2.1: Convolutional Neural Networks is a type of Feed-Forward Neural Networks with hidden layers of convolutional layers.At a higher level, convolutional layers detect these patterns in the image data with the help of filters. The higher-level details are taken care of by the first few convolutional layers.

Faced with these challenges, the research community is constantly exploring related solutions, with the goal of building AI models that are more consistent with medical image classification. Gheshlaghi et al. [65] used the superpixel segmentation technique to overcome dimensionality problems for multiple sclerosis lesion detection. Fang et al. [66] proposed a superpixel segmentation algorithm to segment two-dimensional bone images and three-dimensional brain images [67, 68]. In comparison to the older dense neural network, neural networks with convolutional layers have different advantages: due to the weight sharing, convolutional layers can process much more input parameters. The performed convolution also allows each neuron to consider local neighborhood parameter

relationships rather than weighing each parameter individually, which allows the network to recognize features, such as edges, corners, and patterns. Additionally, the added pooling forces the network to become less reliant on the input data as part of it is cut at each pooling layer, which might help the network to generalize better. A major drawback of convolutional neural networks is their dependence on huge datasets for training, as they are neither rotation-, translation- nor color-invariant. This requires multiple images of the object of interest in different positions and rotations as well as lighting conditions. This could partially be alleviated by the introduction of random rotation/translation into the dataset (data augmentation), but might introduce difficult to detect artificial artifacts, which could negatively influence the classification results. The State-of-the-Art therefore highlights how neural networks continue to evolve, with ongoing research focused on improving their efficiency, robustness, and applicability to a wider range of problems.

Recently, deep learning approaches appear promising in the domain of medical image analysis. However, in the medical imaging, acquiring the required amount of data for training the deep learning models remains to be a significant challenge when compared to natural images. When deep learning models are trained with the limited medical data available, deep CNN tends to suffer from overfitting, and there arise the convergence problems. Because the collection of numerous samples is cost prohibitive, data augmentation methods have commonly been used [69, 70]. To address these issues, also a transfer learning approaches have been applied to deep CNN's thereby enabling their use in medical imaging applications with smaller datasets [71, 72, 73, 74]. Therefore, the specific feature is an essential factor to improve accuracy; only a network model with proper size and other effective methods preventing overfitting. It is therefore necessary to find effective methods both to stabilize the training process and to direct the neural network to extract correct features.

2.2 Internet of Medical Things

A new network infrastructure is being planned and proposed considering continuous increment of the number of connected devices. The academy and industry proposed a new vision of the Internet with the Internet of Medical Things (IoMT), considering the next generation of the Internet. IoMT offers intelligence to objects by adding them the capacity to collect and store data from different types of sensors, to perform actions autonomously based on actuators, coordinate functions, and share information considering the connectivity among nodes, for example, [75, 76].

Supporting IoT are remote controls and machine-to-machine (M2M) communication in various industrial sectors [77, 78, 79, 80] including healthcare [81]. Many research companies present perspectives and trends for the future of IoT as well as research community proposes a Future Internet of Things [82]. Inter-

national Data Corporation (IDC) predicts that IoT will reach about a US$ 1.7 trillion market by 2020. Gartner expects 25 billion connected devices by 2020 while Cisco mentions about 50 billion. And Harvard Business Review expects 28 billion *things* connected to the Internet [83].

Healthcare industry is among the fastest to embrace IoT-based solutions. It is being considered one of the key industry drivers and a special concept for it, considering the IoT application on e-Health, aka Internet of Health Things (IoHT). The key benefit of the IoT in the medical domain is connected technology. Devices are used for assessing patients' conditions, and monitoring and supporting rehabilitation, so that a personalized plan of care can be defined and kept updated. This also fosters continuity of care, enabling a patient to be supervised by a multidisciplinary team even after dehospitalization. This solution includes a network architecture that allows the connection between a patient and healthcare. The data input from patients can be collected through sensors and processed by applications developed for a user terminal, such as computers, smart phones, smart watches or, even, a specific embedded device. Healthcare providers may analyze the data from patients in remote locations and request emergency assistance if necessary.

Among the several implications of using IoT/ICT solutions in the healthcare domain, both clinical and practical consequences stand out. Traditionally, a significant portion of physical therapy, rehabilitation, and clinical assessment has relied heavily on the clinician's expertise, often based on visual observation and subjective judgment. This approach, while valuable, can lack the precision and consistency needed for quantitative analysis and large-scale monitoring.

In recent years, however, advances in sensor technology, wearable devices, and computing power have enabled the integration of motion capture systems into clinical practice. These systems support objective evaluation, continuous performance monitoring, and accurate Range-of-Movement (ROM) measurements. Compared to conventional supervision by rehabilitation specialists, such systems offer multiple advantages: they allow for precise tracking of joint angles over time, enable the visualization of complex movements in three dimensions ($3D$), support remote or home-based rehabilitation programs, and provide access to large datasets that can be used for machine learning, pattern recognition, and longitudinal analysis.

Within this context, it becomes possible to analyze motion and posture quantitatively through the acquisition of a temporal feature signal, denoted as $\Psi(t)$. This signal can represent raw sensor data (e.g., acceleration, angular velocity) or derived biomechanical parameters (e.g., joint torque, segment orientation). Assuming that the subject is in a defined postural state, Γ, specific characteristics of interest can be extracted. To structure the analysis, we introduce two key states: a *resting state*, denoted as R_β, which corresponds to static postures (e.g., standing, seated rest), and a *motion state*, denoted as M_α, which captures specific dynamic activities (e.g., walking, reaching, lifting). Based on these definitions, we distinguish between: *Instant Features*, expressed as $\Psi(t)|R_\beta$,

which are extracted during resting conditions and can reflect stability, postural alignment, or involuntary tremor and *Dynamic Features*, expressed as $\Psi(t)|M_\alpha$, which are derived from specific movements and can capture performance metrics such as velocity profiles, joint coordination, or movement symmetry.

These classifications provide a structured framework for feature extraction and analysis, supporting greater personalization of treatment plans, early detection of motor impairments, increased engagement in therapeutic activities, continuous patient monitoring across clinical and non-clinical environments (dehospitalization). Finally, IoT is bringing innovations to many segments of the industry. The world of IoT, includes the development of architectures and platforms, security requirements, policies and regulations that should be taken into account in future research especially if other technologies, such as big data and cognitive systems, could fit into this context.

2.2.1 Exergames

Exergames [84] are digital games designed to promote physical exercise and support rehabilitation practices. As already highlighted in the 2015 Canadian Stroke Best Practice Recommendations [85], virtual reality, including both immersive and non-immersive technologies such as gaming devices, can serve as an adjunct to traditional rehabilitation therapies, offering additional opportunities for engagement, feedback, repetition, intensity, and task-oriented training. Indeed, the *gamification* [86] determines a motivating and engaging environment in order to keep the patient busy without inducing boredom or fatigue, with consequent frustration and abandonment of therapy. The collection of data during the exercises is the necessary feature to give the therapist the opportunity to understand the patient's behavior, assess fatigue and possibly correct erroneous attitudes.

2.2.2 Telerehabilitation

Over the last years, the impact of telemedicine is still growing thanks to the adoption of the concept of *continuity of care* by the Health Community. Operators have access to all patients' information, regardless of where they are located, and this helps to improve access to health services for end users [87]. Along with the societal changes, health-care sector is facing a quick and smart evolution, both in Europe and in the rest of the world. Some of the big issues, such as aging population, chronic diseases, and spread of disabilities as stroke and all degenerative diseases have a great impact on health care strategies and care solutions. Advances in internet technologies and the widespread availability of broadband connections in homes and workplaces have made it possible to provide telerehabilitation services that were previously economically or logistically unsustainable. Telerehabilitation can be defined as the application of telecommunications technologies to support rehabilitation services [88].

The type of communication technologies and IT infrastructure used in telerehabilitation is strongly influenced by the specific nature of the rehabilitation services provided. The telerehabilitation paradigm is based on the assumption that reducing geographical barriers improves access to care, opening up new possibilities for the design and implementation of intervention strategies across the entire care continuum. The mitigation of distance constraints can be achieved through different modes of telecommunication, including voice communications, video systems, and virtual reality environments [89].

In its early stages of development, telerehabilitation was mainly associated with synchronous interaction models, as opposed to solutions based on asynchronous interaction mechanisms (asynchronous interaction) [90]. However, the sector is characterized by a dynamic relationship between technological infrastructure and the services offered: the availability of new infrastructure enables the introduction of innovative services, while the emergence of new clinical and operational needs can in turn stimulate the development of dedicated infrastructure solutions. Thanks to the new technology solutions, the patient is more and more attentive and aware of his state of health, trying to maintain the highest quality of life level [91]. The digital innovation that has involved people's daily life, plays a fundamental role in the field of medical rehabilitation and many medical clinics start to include it as a fundamental therapeutic and commercial service.

Telerehabilitation is mainly used in physical therapy [92, 93], while neurological rehabilitation programs have been used, in particular, to monitor the rehabilitation progress of stroke patients [94]. Various telerehabilitation techniques make use of virtual reality-inspired environments [95, 96], as well as solutions for the rehabilitation of neurological disorders based on the use of robotic systems and game-based approaches [97]. In numerous application contexts, telerehabilitation has also been integrated with technologies not strictly related to rehabilitation, such as remote monitoring of cardiovascular parameters, including electrocardiogram (ECG), blood pressure, and oxygen saturation, in patients with chronic diseases [98]. These solutions fall within a distinct area of telemedicine known as telemonitoring, which has undergone extensive development and significant diffusion in recent years. Finally, some studies have analyzed the economic aspects associated with the adoption of telerehabilitation, highlighting its potential contribution to reducing hospitalization costs [99].

In a telerehabilitation context, the patient is provided with systems capable of remotely giving indications on the prescribed exercises while medical professionals can remotely observe the activity performed. An example of this approach is the markerless/contactless ReMoVES system [100], based on Microsoft Kinect technology [4], which is capable of acquiring a large amount of data and signals from which the skeletal joints of the patient's body are estimated and information of clinical interest is extracted. Designed to provide remote tests for the assessment of patients' impairments and to provide exercises to support motor/cognitive rehabilitation, ReMoVES has already been tested

on neurological [101, 102, 103] and rheumatic patients [104] as well as on the elderly [100].

2.2.3 Technologies Employed in Multiple Sclerosis

The compromise of cognitive functions affects up to 70% of the population diagnosed with Multiple Sclerosis (MS) and involves specific deficits in domains rather than a uniform global cognitive decline. Patients with MS (PWMS) may have difficulties in information processing speed, attention, learning and episodic memory, executive functions, and visuospatial skills. Cognitive impairment (CI) can also occur in the early stages of the disease, and about half of the individuals with disability report minimal or mild cognitive difficulties within the first years after diagnosis. CI has significant consequences for everyday life and is the main cause of occupational disability and difficulties in activities of daily living.

Neurorehabilitation is becoming a therapeutic option for patients with multiple sclerosis, and there is increasing use of technology to make therapeutic programs more accessible, convenient, and suitable for self-administration. Furthermore, after the COVID-19 pandemic emergency, healthcare professionals are increasingly integrating tele-rehabilitation techniques to allow patients to participate in rehabilitation programs remotely. This approach not only provides a safe alternative, but can also increase flexibility and accessibility for patients who may have difficulty attending rehabilitation sessions in person.

The importance of using technology in the treatment of MS has now been recognized, so much so that several solutions addressing diagnosis, monitoring, and rehabilitation can be found in the literature [105, 106]. The findings from recent reviews suggest that rehabilitative exercise can be viewed as a significant complementary therapy for MS [107, 108]. The results of recent reviews indicate that rehabilitative exercise can be considered an important complementary therapy for MS, inducing neuroprotective phenomena [109]. The use of new technologies such as Virtual Reality (VR) and exergames has emerged as a reinforcement tool for the rehabilitative treatment of people with MS [110]. Systematic reviews on VR in rehabilitation across various neurological conditions, including MS [111], suggest that VR serves as a motivating and engaging rehabilitation method, potentially enhancing therapeutic compliance [112]. Additionally, by allowing for the selection of different exercises and levels of complexity, VR can adapt to the wide variability in patients' conditions and disease progression [113]. The use of markerless sensors has recently emerged as the most easily feasible approach that avoids the use of wearable systems and can be easily used for the assessment of disabilities and as a support for home rehabilitation by the patient in full autonomy [114]. In this context, the Microsoft Kinect sensor has proven to be reliable and effective in telerehabilitation applications also for PWMS [115].

Based on such a past experience, new stakeholder roles have been defined in close coordination between medical and technical developers to address the new MS use case target and update related goals and actions. Based on the

clinical needs of MS patients and the particular types of CI they are affected by, new activities delivered in form of exergames have been developed to address functional abilities of attention, memory, executive functions, and information processing speed. Several studies have shown that patients have a better understanding of their goals and their physical and mental well-being thanks to the improved feedback provided by technology, which leads to better practice and greater involvement in therapy [110]. In addition, technology support also favors intense and repetitive training that yields effective results in functional recovery for MS patients [116, 117]. Of note, the IoMT technologies can support patients in taking control of their own MS disease and collaborate more effectively with the clinical staff [118].

Despite the large interest towards assistive technology in MS, solutions are still not as widespread as they may be, because of some barriers for patients in terms of usability and feasibility, and also because of the high costs of some devices [119]. Indeed, MS patients can experience difficulties in dealing with technological devices, as well as poor skills in using it [120]. In addition, technological solutions made up of wearable devices or controllers may require external support from caregivers, limiting the independent use. Furthermore, the high cost of solutions hinders the possibility of a large home-based usage.

2.3 Clinical Domains

In this section, a brief description of the diseases taken into account for the experimental phase is provided. They refer to different areas, chronic hepatic disease, cognitive diseases, and degenerative diseases.

Chronic Hepatic Diseases

Chronic liver disease (CLD) is one of the common diseases and about two million people die from liver disease every year worldwide [121, 122]. The severity of liver fibrosis has an important impact on the treatment of CLD. In the clinic, the gold standard for fibrosis staging is mainly based on liver tissue biopsy [123], which, however, presents several contraindications. Therefore, rapid and effective treatments are needed for the staging of fibrosis with inexpensive and non-invasive methods.

Hepatic diagnosis and staging is critical for assessing liver health, diagnosing diseases such as cirrhosis, and planning treatment. Noninvasive methods are increasingly preferred to reduce the risks associated with invasive procedures and include: liver elastography in which techniques such as FibroScan (vibration pulse elastography) or MRI/US elastography (magnetic resonance imaging or ultrasonography), measure liver stiffness (an indicator of the degree of fibrosis or cirrhosis) are rapid and widely used in clinical practice, blood tests (Biomarkers) used as an alternative or complement to elastography and finally Advanced

radiological imaging. Modern medical practices for liver staging are based on a combination of non-invasive and invasive techniques. Traditional methods take a lot of time and effort to extract and select classification features.

Typical methods are elastography-based techniques [124, 125, 126, 127, 128], are the technologies most used for this purpose, but may depend on the operator's experience and therefore they may lead to inconsistent results. In fact, Shear Wave Elastography (SWE) is a medical imaging technique used to quantify tissue elasticity by analyzing the propagation velocity of shear (transverse) waves in the region of interest. SWE generates transverse waves, and their propagation speed is related to tissue stiffness (kPa). Faster propagation indicates a stiffer fabric. The measurement protocol involves a minimum of 10 measurements (is recommended) to ensure statistical reliability. The median value of these measurements is calculated to represent the stiffness of the tissue. Then, the corresponding Measurement Reliability Index (RMI) is obtained, which quantifies the confidence in the measurement. A value greater than 0.5 is necessary for a reliable reading. Only measurements with RMI values greater than 0.5 and IQR/med (interquartile range/median) less than 30 percent are considered reliable for liver elasticity.

Another method, FibroScan (transient elastography), is a highly accurate and practical tool for assessing fibrosis in chronic liver diseases, especially advanced fibrosis and cirrhosis. However, confounding factors such as inflammation, steatosis, or obesity should be considered, and alternative methods may be needed for borderline or inconclusive cases.

Other problems in this area relate to access to medical datasets, which are often more difficult to obtain than other datasets. First, because annotation of medical images requires significant professional medical knowledge, which makes annotation very time-consuming, expensive, and rare. Second, medical data are private and cannot be used publicly without first obtaining certain permissions. Because of these limitations, it is necessary to have other types of tools.

Multiple sclerosis

Multiple Sclerosis (MS) is a chronic, inflammatory and demyelinating disorder of the central nervous system (CNS). It is characterized by a deficit of neurological functions, including motor, sensory, and cognition, which can be relapsing and/or progressive in nature [129]. Cognitive impairment (CI) can affect up to 70% of the MS population. Persons with MS can experience difficulties in several cognitive domains, including processing speed, sustained and selective attention, learning and episodic memory, with executive functions compromised in more advanced, progressive stages [130]. Other cognitive deficits such as visuospatial problems and difficulties with social functioning can also be present [131]. While more severe cognitive impairment is more likely in persons with secondary progressive MS, signs of cognitive involvement can be present early in the disease process. Within the first year of diagnosis, about half of persons

with MS report having either minimal or mild cognitive difficulties, with greater complaints over the first decade. Although uncommon, some persons with MS present with cognitive impairment as their primary symptom. In addition, cognitive issues may be present preclinically [132]. These impairments correlate closely with brain pathology and can have major consequences for everyday life. Further, CI is the leading predictor of occupational disability in these patients. Treatment options remain extremely limited, however, despite an increase in available interventions. In principle, disease-modifying therapies might improve cognition in people with MS as these agents are primarily designed to arrest the disease and prevent relapses, but whether they directly improve cognition remains speculative [133].

Cognitive and behavioral rehabilitation strategies are designed to enhance an individual's capacity to process and interpret information and to function in all aspects of family and community life. Although the focus on designing and testing effective cognitive rehabilitation programs for people with MS is a relatively recent phenomenon, the growth in research studies addressing this need has been substantial over the past decade. As a result, the Italian National MS Society has recommended remedial interventions and accommodations that can be made to manage cognitive impairment and improve everyday functioning in both adult and pediatric MS populations. Such recommendations include more comprehensive assessment for anyone who tests positive for cognitive impairment on cognitive screening or demonstrates substantial cognitive decline, as well as neuropsychological evaluation for any unexplained change in academic performance or behavioral functioning in school-aged children with MS. Evidence suggests that cognitive rehabilitation has a long-term impact well beyond the treatment period and might enhance cognition in the face of future brain changes [134]. Such sustained effects have been documented in the literature on aging, in which cognitive rehabilitation not only improved everyday life activities, but also resulted in a 29% reduction in dementia risk ten years after treatment [134].

2.4 The World of Machine Learning

The aim of this section is not to provide a formal definition or exhaustive classification of machine learning methods, nor to outline their theoretical foundations in detail. These aspects are widely covered in the reference literature and are beyond the scope of this work. Rather, the intent is to introduce the reader to the conceptual context in which ML is situated, highlighting its role as a computational paradigm oriented toward extracting knowledge from data and modeling complex phenomena.

2.4.1 The Concept

The biological vision system of humans is an extraordinary mechanism, capable of performing intricate tasks such as analysis, interpretation, recognition, and classification of visual patterns with remarkable ease. Emulating these capabilities through computer vision and artificial intelligence has been a long-standing goal of medical image processing research. By treating medical images as structured matrices of data and applying multi-level processing algorithms, it becomes possible to approximate human-level perceptual analysis and extract diagnostically relevant features.

The artificial neural network (ANN)–a machine learning technique refers to a system of interconnected nodes, inspired by biological brain neurons, designed to mimic human cognition and learning Fig.2.2 – came into prominence in the mid-20th century, with the development of the perceptron [135]. However, the ANN was previously limited in its ability to solve actual problems, due to the vanishing gradient and overfitting problems with training of deep architecture, lack of computing power, and primarily the absence of sufficient data to train the computer system. Interest in this concept has lately resurfaced, due to the availability of big data, enhanced computing power with the current graphics processing units, and novel algorithms to train the deep neural network. Different types of network architectures have been developed over time to address different types of data and problems. The first developed types were fully connected neural networks, followed by convolutional neural networks (CNN). Currently, more complicated networks, such as U-Nets or Generative Adversarial Neural Networks are also abundant.

Regarding the increase in computing power of graphics processing units (GPUs), numerous frameworks for a variety of programming languages can be exploited to use ANNs (Table 2.1). This includes prominent examples, such as TensorFlow and Torch, and lesser known examples, such as scikit-learn (Table 2.1). Most frameworks are designed to be used without in-depth knowledge of the underlying functionality and concepts. One example of this approach is the TensorFlow Estimator classes, which enable model creation and training without the precise definition of learning methods, network layers and activation functions.

Table 2.1: Selected machine-learning frameworks and their supported programming languages.

Framework	**Programming Languages**
Tensorflow [136]	Python, R, Java, C++, Go
pyTorch [137]	Python
sklearn [138]	Python
caffe [139]	C++, Python, Matlab
Keras [140]	Python, R
SparkMLlib [141]	Java, Scala, Python, R

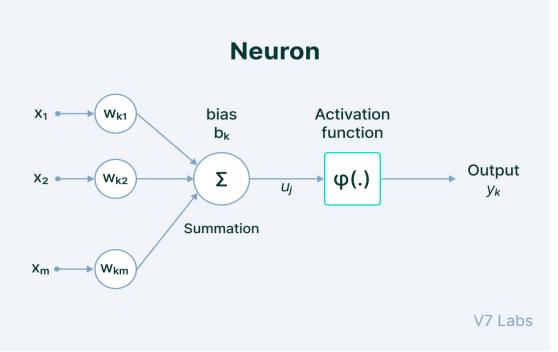


Figure 2.2: The input is the set of features that are fed into the model for the learning process. Weights allow you to give importance to those features that contribute the most to the learning. The job of the transfer function is to combine multiple inputs into an output value so that the activation function can be applied. This is done by simply adding all the inputs to the transfer function. Activation function introduces nonlinearity into the operation of the perceptrons to account for variable linearity with the inputs. Without this, the output would just be a linear combination of input values and would not be able to introduce nonlinearity into the network. Finally, the role of bias is to shift the value produced by the activation function.

Instead, with regard to new deep neural network (DNN) training algorithms we have: supervised learning that involves generating a predictive function by inferring patterns from labeled training data. In this approach, training data consist of numerical or nominal feature vectors representing the characteristics of input data, accompanied by corresponding output labels. When the output variable is continuous, the task is referred to as regression, while in cases where the output assumes discrete, categorized values, the task is classified as classification. Conversely, unsupervised learning focuses on identifying hidden structures or intrinsic patterns in unlabeled data, without reference to predefined outputs. Since the input examples lack explicit labels, the performance evaluation relies on indirect validation methods rather than objective accuracy assessments. Semi-supervised learning, combining supervised and unsupervised learning. In this type of learning, both labeled and unlabeled data are used to train the AI models needed for classification and regression operations.

Recent studies on this technology suggest its potentially to perform better than humans in some visual and auditory recognition tasks, which may portend its applications in medicine and healthcare, especially in medical imaging, in the foreseeable future.

The ability of a neural network to generalize and model real-world problems

comes mainly from the following aspects:

- Network Structure: Neural networks are composed of layers of artificial neurons. Neurons in one layer are connected to neurons in subsequent layers through weighted connections.
- Weights and Biases: Each connection between neurons has an associated weight, which determines the importance of the input. Each neuron also has a bias, an additional value that helps adjust the neuron's output. The weights and biases are the parameters that the network learns during the training process.
- Activation Function: Each neuron applies an activation function to its weighted input to produce an output. Activation functions introduce non-linearity into the model, allowing the network to learn and represent complex relationships between inputs and outputs.
- Backpropagation: During training, the network uses an algorithm called backpropagation to update the weights and biases. This process involves calculating the error between the network's predicted output and the target value, propagating the error back through the network, and adjusting the weights and biases to minimize this error.
- Optimization: The network uses optimization techniques, such as the gradient descent algorithm, to find the optimal values of the weights and biases that minimize the loss function (i.e., the error measure).
- Generalization: The ability of the network to generalize, that is, to perform well on new data not seen during training, is crucial. This is achieved by balancing the number of parameters and the volume of training data, and using techniques such as regularization to prevent overfitting.

In summary, through training, a neural network continuously adjusts its weights and biases to reduce error and move closer to an optimal solution. This process allows it to model complex real-world problems and make accurate predictions or decisions based on the data provided.

2.4.2 Machine Learning as a Support Tool

Machine Learning (ML) is defined as a set of methods that automatically detect patterns in data, and then utilize the uncovered patterns to predict future data or enable decision making under uncertain conditions [142]. ML is a subset of artificial intelligence (AI). In general, there are three approaches to AI: symbolism (rule based, such as IBM Watson [143]), connectionism (network and connection based, such as deep learning or artificial neural net), and Bayesian (based on the Bayesian theorem [144]). The most representative characteristic

of ML is that it is driven by data, and the decision process is accomplished with minimum interventions by a human. The program can learn by analyzing training data, and then make a prediction when new data is put in. This thesis focuses on the second type mentioned. Solving many limitations due to the current availability of big data, the increased computational power of graphics processing units (GPUs) and new deep neural network (DNN) training algorithms, these deep learning approaches have shown impressive performance in mimicking humans in various fields, including medical imaging. Since the 1980*s*, numerous ML algorithms with different implementations, mathematical bases, and logical theories have been executed to perform such classification tasks. Accordingly, several computer-aided detection (CAD) systems were developed and introduced in the clinical workflow in the early 2000*s* [145]. Supervised methods, such as Support Vector Machines (SVM), Random Forests, and traditional neural networks, represented a first step in the application of machine learning to imaging, demonstrating effectiveness in classification and prediction tasks based on manually extracted features. At the same time, unsupervised approaches, based on clustering and dimensionality reduction techniques, have made it possible to identify latent structures and correlations in the data, proving particularly useful when expert annotations are scarce or absent.

The rapid development of AI technology requires radiologists to be informed about this technology in order to understand the capabilities of AI and how it could change and affect radiological practice in the near future. In fact, machine learning methods have taken on a leading role in biomedical image analysis, offering innovative tools to address the growing complexity of diagnostic and research data. Images produced by different acquisition modalities, such as magnetic resonance imaging, computed tomography, ultrasound, microscopy, or digital histopathology, are characterized by high dimensionality and information content that often exceeds the capabilities of manual interpretation. In this context, machine learning allows complex patterns to be extracted, modeled, and interpreted, automating tasks traditionally entrusted to the physician or researcher's experience.

It is important to emphasize that the introduction of AI into the clinical setting should not be regarded as a replacement for human expertise, but rather as a support tool. The intention is not to substitute radiologists, but to automate specific tasks, thereby enhancing the overall efficiency and effectiveness of radiological practice. In this sense, AI serves as a complement to the irreplaceable and highly valuable human skills involved in clinical decision-making. Relying entirely on them could be a risky choice, as one of the main challenges associated with neural networks remains their inherent complexity, which often hinders the interpretability of the results they produce, commonly referred to as the *black box* problem. In recent years, some efforts have been made to make neural networks more interpretable (e.g., through reverse engineering or various visualization techniques, [146, 147]); however, as neural networks grow in size and complexity, the problem is likely to become worse.

2.4.3 Soft Computing: Machine Learning in Biomedical Context

In the healthcare field, there is a demand to create an environment where patients can receive optimal medical and nursing care in a sustainable and efficient manner. This was definitely the right push to insert soft computing techniques in healthcare due to their ability to efficiently diagnose diseases, propose suitable treatments, and deliver superior results compared to conventional approaches.

These techniques exhibit adaptable behavior, which allows them to adapt their strategies based on the specific problem at hand. Soft computing techniques, including machine learning, have become increasingly popular in healthcare due to their effectiveness in diagnosing diseases, proposing treatments, and delivering superior outcomes compared to traditional approaches [148]. These adaptable techniques allow them to tailor their strategies to the specific problem. ML, a subfield of artificial intelligence, plays a crucial role in soft computing. It enables systems to learn and improve from experience without explicit programming [149]. ML algorithms make informed decisions by leveraging data, observations, or previous experiences.

Recent advances in ML have highlighted the increasingly central role of these techniques in clinical processes, particularly with regard to the early diagnosis and management of chronic and infectious diseases [150, 151]. In this context, ML has become a key tool for disease prediction, with numerous studies analyzing a wide range of approaches, including models based on laboratory biomarkers, symptom data, contextual information, and integrated architectures for managing multiple diseases. Machine learning models based on laboratory test results have been extensively investigated to support the diagnosis and treatment of noncommunicable diseases [152, 153]. Although these studies confirm the importance of clinical biomarkers, many of these models have been validated on controlled and homogeneous datasets, raising significant questions about their ability to generalize effectively across heterogeneous populations and different clinical settings. At the same time, symptom-based data are becoming increasingly relevant, especially in scenarios where standardized clinical biomarkers are not available [154, 155]. However, despite encouraging results, symptom-centric approaches are particularly sensitive to noise and data inconsistency, as patient-reported information can vary considerably between different demographic groups and healthcare systems. These studies therefore highlight both the predictive potential of these methods and their limitations in terms of reproducibility and cross-context robustness. In this context, the definition and adoption of robust and clinically relevant evaluation metrics remain crucial for the effective integration of machine learning techniques into medical practice [156]. In recent years, there has also been growing interest in the development of integrated predictive models for the simultaneous management of multiple diseases [157, 158, 159]. Despite these advances, many of these approaches are evaluated exclusively on limited case studies; moreover, performance tends to

degrade when datasets are extended to include heterogeneous disease types or less structured data. It is important to note that most of these solutions aim to build a single unified predictive model for multiple diseases. Despite the progress described, a fundamental critical issue remains: the effectiveness of machine learning models depends significantly on the nature, quality, and context of the data used. Much of the existing literature focuses on specific diseases or highly homogeneous datasets, limiting the transferability of predictive performance to real-world clinical scenarios. As highlighted in [160], there is a substantial gap between the theoretical development of models and their actual clinical applicability.

Deep learning, a subfield of machine learning, is considered a part of soft computing techniques. Deep learning methods are highly effective when the number of available data is large during a training stage. Particularly, great improvements in computer vision inspired the use of deep learning in medical image analysis (e.g., segmentation, object detection, classification, prognosis prediction, and microscopic imaging analysis). Among computerized tools, deep learning-based applications are proving to be the State-of-the-Art foundation, leading to improved accuracy, which allowed new frontiers in medical image analysis.

Medical images are different from other pictures where they depict distributions of various physical features of the human body. These medical images carry different information, and many analytical tasks are related to the individual quantification of entities in the human body. Medical images provide information on the anatomy and physiology of different organs, which is required to precisely diagnose lesions. Recently, the use of big data analytics has substantially increased in healthcare, with medical imaging playing a key role in it. Machine learning have dramatically improved the State-of-the-Art in medical image analysis [161, 162, 163]. Software applications are beginning to be certified for clinical use [164, 165]. Particularly, deep convolutional neural networks (CNNs) automatically learn mid- and high-level abstractions from raw data, which are powerful tools for computer vision tasks, including medical image analysis. Moreover, the demand for reducing manual costs in various industries has promoted the growth of automation and computer-aided technologies. Additionally, the availability of low-cost graphical processing units (GPUs) and memory from the video-game industry has made it possible to introduce CNNs with several layers and kernels [166]. Deep learning methods are highly effective when the amount of data is large for training [167]. However, as for medical image analysis, usually, the amount of available data from medical institutions is limited. Additionally, because medical image labeling can be performed by professional medical practitioners, correctly labeled data are even more scarce. These factors prohibited the expansion of CNNs in medical image analysis. Therefore, one of the main challenges in applying deep learning to medical image analysis is the limited number of available training and test datasets to develop models with high accuracy without suffering from overfitting [167]. To

reduce overfitting, there are good algorithmic techniques to better train models (e.g., initialization and momentum [168, 169], rectified linear unit (ReLU) [170], denoising [171], dropout [172] and dropconnect [173], and batch normalization [174, 175]). Data augmentation techniques are often employed to expand and balance datasets, enhancing training efficiency as well. For image classification, augmentations such as rotations, noise addition, translations, and blurring are crucial [176]. However, selecting the appropriate augmentation strategy is essential; excessive modifications can hinder training, as the performance of CNNs is strongly influenced by the training set. Moreover, to overcome the aforementioned obstacles, transfer learning is providing great benefits for deep learning models in medical image analysis [177, 178]. Transfer learning can help promote deep learning models in less-developed application areas, as well as less technically developed geographical areas, even when not much labeled data are available in medical image analysis. Since deep learning techniques have achieved benchmark performance in numerous medical applications, their use to achieve further improvements represents a promising direction in medical image analysis, with potentially significant implications for both clinical practice and scientific research. However, the increase in scientific output in this field does not necessarily translate into actual clinical advancement. In many cases, the intensification of research activity is more aligned with academic incentives than with the real needs of doctors and patients. This can lead, for example, to a proliferation of studies reporting State-of-the-Art performance on reference datasets, without bringing concrete benefits to the resolution of the underlying clinical problem [179, 180].

Chapter 3

Image processing techniques for medical diagnosis

Central to biomedical computing is the use and development of mathematical models that can capture and predict biological behavior based on image-derived data. Modern medical images are no longer limited to static visual records, but are structured and quantitative data arrays on spatial and temporal scales. This evolution has expanded the potential for complex, data-driven modeling, offering new insights into disease mechanisms and therapeutic responses. This chapter introduces the theoretical foundations of the methods developed. In particular, after a brief introduction on the concept of digital images, their characteristics and image processing, details on connected filtering, enhancement, edge detection, segmentation and threshold techniques and recognition problem are reviewed.

3.1 Definitions and Features of Digital Image

A digital image is a two-dimensional array of dots called *image elements* or *pixels*. Each pixel is associated with a discrete value that expresses an intensity. The physical meaning associated with the intensity of each pixel depends on the type of sensor that captured the image. In fact, when a machine (scanner) produces a medical image, it is mapping a certain property of the object. The physical effect of these changes in response to variations is measured by the sensor, which converts the physical property into electrical signals. These are commonly characterized by changes in voltage, current, or frequency. The scanning system has a hardware component that works by adopting a tolerable limit for distortions (system's *dynamic range*) of the image, but also a range of physical levels to produce useful signals needed for image reconstruction. Basically, the minimum possible visual distortion is black, while the maximum brightness is white. These two values represent the two thresholds (minimum and maximum) within which intermediate levels (*gray levels*) appear. Setting the number of gray levels displayed affects the global quality of the image. Gray level and dynamic range are certainly among the fundamental characteristics of a digital image.

In image representation we are concerned with the characterization of the quality that each pixel represents. In general, any two-dimensional feature that contains information can be considered an image. An important consideration in image representation is the criterion of fidelity or intelligibility to measure

the quality of an image or the performance of a processing technique.

3.2 Image Pre-processing

The term *pre-processing* indicates the set of operators that make an image easier to understand for a human operator or more suitable for an automatic processing. In particular we consider two main types of operation:

- **Restoration** devoted to eliminate or reduce image degradations (the information contained in the output image is improved or completed):
  - Correction of geometric distorsions and non-linearities introduced by the sensor;
  - Reconstruction (interpolation) for incomplete data or data affected by noise;
  - Noise filtering (additive, multiplicative, impulsive noise, etc.).
- **Enhancement** aiming to improve some characteristics of the image for subsequent analysis or display (emphasizes characteristics already present in the image):
  - Contrast enhancement (histograms, look-up table);
  - Pseudocolor (False colors LUT);
  - High Frequencies enhancement (sharpening).

In image analysis, pre-processing is often considered a crucial first step, since the quality of subsequent tasks (such as segmentation, feature extraction, or recognition problem) strongly depends on how well the input data has been prepared.

3.2.1 SRAD

In the context of medical imaging, filtering techniques can be used to improve the clarity and detail of medical images. Filtering is the process of manipulating an image to alter its appearance. Ultrasound imaging devices due to speckle noise, it exhibits lower resolution and image quality than other than medical diagnostic devices. An algorithm using speckle reducing anisotropic diffusion (SRAD) was then created following [181]. It is possible to summarize the various steps:

- compute derivative approximation and Laplacian approximation;
- compute the diffusion coefficient;
- compute the divergence of diffusion function;

- restore image.

The speckle noise generated in the ultrasound image represents a different character from the conventional noise, so this feature is difficult to remove speckle noise with existing filtering techniques [182]. Since speckle noise in ultrasound images interferes with the diagnosis of lesions through the image, this noise reduction technique in the ultrasound image is an important technology in medical image processing.

3.2.2 Image Enhancement

In enhancement techniques, the goal is to accentuate certain image features for later analysis or display. Image enhancement itself does not increase the intrinsic information content of the data and is useful for feature extraction, image analysis, and visual information visualization. Let a $2D$ digital image be represented by a matrix I of discrete points, which is a functional mapping $I(x, y)$:

$$\mathcal{I} : \mathcal{Z}\mathrm{x}\mathcal{Z} \rightarrow \mathcal{N}$$

where $\mathcal{Z}$ is the set of integer numbers and $\mathcal{N}$ is the set of natural numbers.

The discrete domain is (x, y), where

$$x = 0, ..., M - 1$$

is the column number and

$$y = 0, ..., N - 1$$

is the row number. The discrete co-domain is the pixel echographic intensity U, where

$$U \in [0, L - 1],$$

with L being equal to 256 in the present case of a format comprising 8 bits per pixel.

Punctual-based processing refers to point operators where the value of each pixel in the output image depends solely on the value of the corresponding pixel in the input image, using a linear or non-linear function. These operations are often based on appropriate transformations of the image histogram, which is the graphical representation of the occurrence frequency of gray levels:

$$h(U) = \frac{\#Pixel(U)}{M\mathrm{x}N}$$

It is evident that $\sum_{U=0}^{L-1} h(U) = 1$. Moreover, the histogram serves as an estimate of the statistical distribution of the image's intensity and tracks the occurrences of gray levels.

The use of point operators makes it possible for the value of each pixel in the output image to depend only on the value of the corresponding pixel in the

input image, via a linear or nonlinear function. Considering the histogram of an image, it describes the statistical distribution of gray levels, i.e. it traces the occurrences of gray levels: $h(l) = numPixel$ with $l = 0 \div L - 1$.

It is usually displayed as a bar graph where each bar (corresponding to one of the possible intensity values) is the count of pixels that in the image have that associated value.

The distribution histogram is modified by appropriate functions, in order to enhance certain features (Fig. 3.1). A classical histogram operations is contrast stretching that through piecewise linear functions (windowing, thresholding, etc...) map each gray level to another gray level by a predetermined transformation (Eq.3.1):

$$V = g(U) = \begin{cases} \alpha U, & 0 \leq U < a \\ \beta(U-a)V_a, & a \leq U < b \\ \gamma(U-b)V_b, & b \leq U < L-1 \end{cases} \tag{3.1}$$

where the slopes α, β, γ determine the relative contrast stretching.

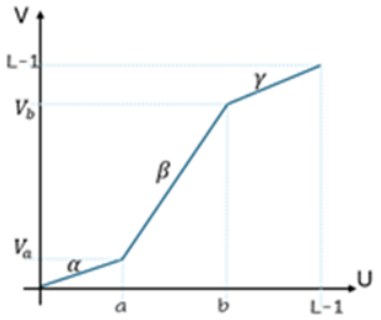


Figure 3.1: Piecewise histogram contrast transform. In abscissa, the original gray levels $I_{in} : (x, y) \rightarrow U$ are shown; in the ordinate axis, the output intensity $I_{out} : (x, y) \rightarrow V$ is shown.

Contrast stretching enhances image quality by adjusting intensity values to span the full range allowed by the image format. This process modifies the distribution of pixel intensity values, making finer details more distinguishable, especially those that may be difficult for the human eye to perceive. The greatest difficulty in image enhancement is quantifying the criterion for enhancement.

In this thesis, contrast enhancement is used to improve the clarity of blurred images and overall image quality.

3.3 Symbolic Features

3.3.1 Edge Detection Operators

The edge of objects forms the most important part of the structure hierarchy that links the raw data of an image to its interpretation. The detection operation consists of two main phases:

1. Each pixel is associated with a value that represents the possibility of being an edge point;
2. A binarization step is then performed according to more or less complex criteria.

Edge models are theoretical constructs used to describe and understand the different types of edges that can occur in an image. These models help in developing algorithms for edge detection by categorizing the types of intensity changes that signify edges. The basic edge models are *Step*, *Ramp* and *Roof*. A step edge represents an abrupt change in intensity, where the image intensity transitions from one value to another in a single step. A ramp edge describes a gradual transition in intensity over a certain distance, rather than an abrupt change. A roof edge represents a peak or ridge in the intensity profile, where the intensity increases to a maximum and then decreases.

In order to measure dissimilarities between neighbors pixels the smallest window is made of two pixels. In this way, it's possible measure the difference between the two pixels gray levels (Eq.3.2 and Eq.3.3). When the difference is large we say we have an edge passing between the two pixels.

$$\frac{d}{dx}I(x,y) \cong I(x+1,y) - I(x,y) \tag{3.2}$$

$$\frac{d}{dy}I(x,y) \cong I(x,y+1) - I(x,y) \tag{3.3}$$

which translate respectively into the masks (o kernel): $\begin{bmatrix} +1 & -1* \end{bmatrix}$ $\begin{bmatrix} +1 \\ -1* \end{bmatrix}$

Depending upon the kernel matrix, there are various methods of edge detection techniques.

The *first derivative* of an image measures the rate of change in pixel intensity. It is useful for detecting edges because edges are locations in the image where the intensity changes rapidly. In image processing this all translates into using differential operators, which are based on the use of FIR filters implemented with appropriate derivative masks that enhance the high-frequency content of the image. First order Differential (Gradient) methods for edge-detection are non-linear methods, based on indipendent convolutions with two directional masks, approximate calculation of the gradient magnitude and direction, thresholding

step. These techniques utilize gradient information and direction to detect edges within images. How said before, filters play a crucial role in edge detection. Gradient-based methods such as Roberts, Prewitt, and Sobel leverage specific filter matrices to detect edges by analyzing changes in intensity or color gradients across the image.

This approach is better than linear high-pass filtering since it performs a balancing between the results of the two directional filters, thus reducing the influence of noise.

The gradient magnitude is calculated for each point of the image, then a thresholding is applied to decide which pixels are to be considered edges. Main problem of methods based on thresholding is the difficulty of establishing a proper threshold value; a too low value implies a greater influence of noise on the final result, as well as an excessive thickness of the contours, viceversa, a too large threshold value affects the identification of the feeble contours.

In addition to Gradient-based edge detection techniques, there are also Gaussian-based techniques that first use smoothing filters to smooth the image pixels and then make them noise-resistant. The Canny operator and Gaussian Laplacian are widely used for edge detection.

Edge detection is widely used for object recognition, that is, it helps to identify objects in images (e.g., car, face or text detection), for feature extraction, which is essential for artificial intelligence, machine learning and computer vision, in medical imaging to detect tumors and abnormalities, in motion detection they help to track objects in videos (e.g., in surveillance cameras), and finally in autonomous vehicles it is used for lane detection and obstacle recognition.

3.3.2 Segmentation

Image segmentation is defined as a partitioning of an image into regions that are meaningful for a specific task; it is a labeling problem. A region could correspond to an object in the image, or a significant part of it. The information contained in an image can be seen as a composition of spatial events, where each event is a set of pixels. Segmentation is one of the first steps leading to image analysis and interpretation. The goal is easy to state, but difficult to achieve accurately. Image segmentation approaches can be classified according to both the features and the type of techniques used. Features include pixel intensities, edge information, and texture, etc. Techniques based on these features can be broadly classified into structural and statistical methods. Structural methods are based on the spatial properties of the image, such as edges and regions. Various edge detection algorithms have been applied to extract boundaries between different brain tissues. However such algorithms are sensitive to artifacts and noise. Region growing is another popular structural technique. In this approach, one begins by dividing an image into small regions, which can be considered as *seeds*. Then, all boundaries between adjacent regions are examined. Strong

boundaries (in terms of certain specific properties) are kept, while weak boundaries are rejected and the adjacent regions merged. The process is carried out iteratively until no boundaries are weak enough to be rejected. However, the performance of the method depends on seed selection and whether the regions are well defined, and therefore is also not considered robust. Statistical methods label pixels according to probability values, which are determined based on the intensity distribution of the image. Gray-level thresholding is the simplest, yet often effective, segmentation method. In this approach, structures in the image are assigned a label by comparing their gray-level value to one or more intensity thresholds. A single threshold serves to segment the image into only two regions, a background and a foreground. Sometimes the task of selecting a threshold is quite easy, when there is a clear difference between the gray-levels of the objects we wish to segment.

3.3.3 Thresholding

Thresholding is the simplest segmentation technique. Tipically, the image histogram is analyzed by considering the point of minimum between two peaks: this value is chosen as the threshold value. The value 1 is assigned to pixels having a value greater or equal to the threshold value; the remaining pixels are associated with the 0 value. Given the original image $Input(i, j)$, and the image after the thresholding phase, $Output(i, j)$, we have:

$$Output(i,j) = \begin{cases} 1 & is \quad Input(i,j) \geq threshold \\ 0 & is \quad Input(i,j) < threshold \end{cases} \tag{3.4}$$

3.4 Recognition Problem

The *Recognition Problem* refers to the challenge faced in distinguishing background and foreground images accurately under less deal conditions.

The steps to follow within this problem are as follows:

- *Conditioning:* The image is modified to remove information that does not interest you or to improve some characteristics of the images.

- *Labeling:* The labeling operation determines for each pixel the event in which it is involved. Image labeling is a general framework in image processing, consisting of associating each pixel in the image with a label from a finite set. The meaning of such label depends on the problem under analysis, and could either be semantic, as in classification tasks, or numerical, as in segmentation tasks, where it represents the belonging region. A typical example of labeling on grayscale images is the thresholding operation which assigns a label value of 1 to all pixels that have a gray level greater

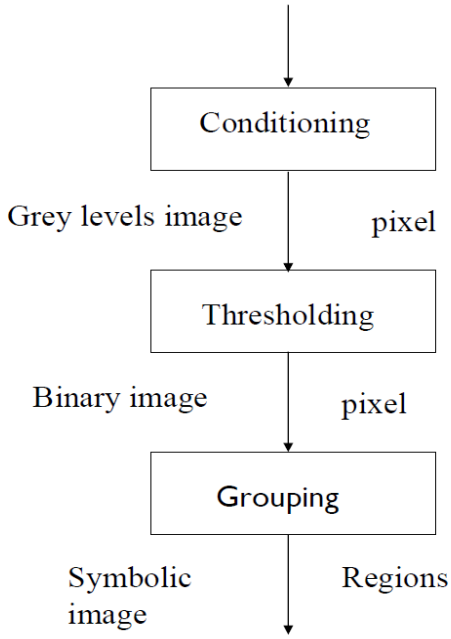


Figure 3.2: A zoom workflow of recognition problem.

than a threshold value. We get a binary image in which the value 1 indicates the pixels belonging to the object of interest while all the other pixels are assigned the label 0 (Fig. 3.2).

- *Grouping:* After the labeling operation that tags pixels as pertaining or not to an object, the operation of grouping follows. Connected pixels with the same label are grouped. This operation carries out a transformation on the elementary processing unit: from single pixels we arrive at a region, i.e, a set of neighbouring pixels. Regions are said *connected components*, and are located using a suitable operator. Such operators have a binary image as input and a symbolic image as output in which the label assigned to every pixel is an integer which identifies the region to which the pixel belongs. Two pixels p and q with the same label f belongs to the same connected component C if there is a sequence of points $(p_0, p_1, ..., p_n)$ of value f belonging to C where $p_0 = p$ and $p_n = q$ and p_i is neighbour to $p_(i-1)$ for $i = 1, ..., n$. Based on the operator we are using there are 4-connected or 8-connected regions (Fig.3.3).

- *Feature Extracting:* For each region identified through grouping, various

properties (features) can be calculated (area, spatial moments, etc.); they depend on the grey levels of the pixels belonging to the region and their spatial arrangement. For each region a features vector is created.

- *Matching:* A meaning can be given to a region when there is a perceptual grouping which allows to interpret the set of pixels as an already known object. Matching operation determines the interpretation of a set of pixels, associating 3-D objects or 2-D shapes.

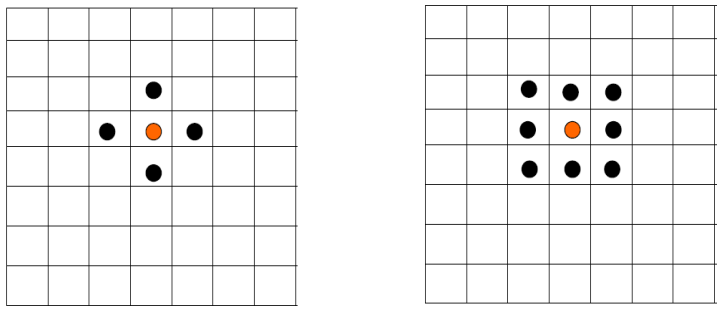


Figure 3.3: Example of 4-connected or 8-connected regions.

3.5 Model Evaluation

In Machine Learning, the efficacy of a model is not just about its ability to make predictions but also to make the right ones. As described in Section 2.4.3, it is essential to select the appropriate augmentation strategy; excessive modifications can hinder training, as CNN performance is strongly influenced by the training set.

In the current work, we have found that traditional data augmentation methods are insufficient, and the direct application of a CNN does not yield satisfactory performance, especially when the database is of a limited size. After training and validating the supervised classification model, its performance was evaluated by calculating accuracy, precision, and recall for binary and three-class classification. Accuracy is a measure of the overall correctness of a model across all classes. The most intuitive metric is the proportion of true results in the total pool. True results include true positives and true negatives. Accuracy may be insufficient in situations with unbalanced classes or different error

costs. Precision and recall fill this gap. Precision measures how often predictions for the positive class are correct. Recall focuses on what you are interested in capturing or rather measures the model's ability to find all positive cases in the dataset. This makes recall crucial when false negatives are expensive or dangerous. In addition, the mean absolute error (MAE) was calculated. This metric was chosen because it is easy to interpret and resistant to outliers. The confusion matrix maps predicted results to actual labels, with $C(p, c)$ denoting predicted class p versus actual class c.

The following metrics were used:

- **Mean Absoulte Error of the predicted classes:**

$$MAE = \frac{1}{N_t} \sum_{i=1}^{N_t} \left| cc_i - pc_i \right| \tag{3.5}$$

  where cc_i is the correct class number of the i_{th} element, pc_i is its predicted class number, N_t is the size of the test set, as defined in Section 5.3 in the description of *Data Augmentation*.

- **Overall Accuracy:**

$$accuracy = \left| \frac{\sum_p C(p,p)}{N_t} \right| \tag{3.6}$$

  is the ratio between the confusion matrix trace and the total tested cases (proportion of correct predictions).

- **Precision** for each class p is

$$Precision = \frac{TP(p)}{TP(p) + FP(p)} \tag{3.7}$$

  where $TP(p)$ is the True Positive value from the confusion matrix (i.e, $TP(p)$=C(p,p)) and $FP(p)$ is the False Positive occurrence:

$$FP(p) = \sum_{c, c \neq p} C(p, c) \tag{3.8}$$

- **Recall** is a measure of how many of the positive cases of a class c are correctly predicted, as compared with the positive cases in the class.

$$Recall = \frac{TP(c)}{TP(c) + FN(c)} \tag{3.9}$$

  where $FN(c)$ is the False Negative occurrence:

$$FN(c) = \sum_{p, c \neq p} C(p, c) \tag{3.10}$$

- **F1-score** is a measure combining both precision and recall and is generally the harmonic mean of the two.

3.6 Machine Learning

Neural networks are a subset of machine learning, and they are at the heart of deep learning algorithms. They are comprised of node layers, containing an input layer, one or more hidden layers, and an output layer. Each node connects to another and has an associated weight and threshold. If the output of any individual node is above the specified threshold value, that node is activated, sending data to the next layer of the network. Otherwise, no data is passed along to the next layer of the network.

3.6.1 Convolutional Neural Network

Convolutional neural networks use two and three-dimensional data for image classification and object recognition tasks. Prior to CNNs, manual, time-consuming feature extraction methods were used to identify objects in images. However, convolutional neural networks now provide a more scalable approach to image classification and object recognition tasks, leveraging principles from linear algebra, specifically matrix multiplication, to identify patterns within an image.

They have three main types of layers, which are: Convolutional layer, Pooling layer and Fully-connected (FC) layer. The convolutional layer is the first layer of a convolutional network. While convolutional layers can be followed by additional convolutional layers or pooling layers, the fully-connected layer is the final layer. With each layer, the CNN increases in its complexity, identifying greater portions of the image. Earlier layers focus on simple features, such as colors and edges. As the image data progresses through the layers of the CNN, it starts to recognize larger elements or shapes of the object until it finally identifies the intended object.

In a convolutional layer a feature detector, also known as a kernel or a filter, which will move across the receptive fields of the image, check if the feature is present. This process is known as a convolution. The feature detector is a two-dimensional (2-D) array of weights, which represents part of the image. While they can vary in size, the filter size is typically a $3x3$ matrix; this also determines the size of the receptive field. The filter is then applied to an area of the image, and a dot product is calculated between the input pixels and the filter. This dot product is then fed into an output array. Afterwards, the filter shifts by a stride, repeating the process until the kernel has swept across the entire image. The final output from the series of dot products from the input and the filter is known as a feature map, activation map or a convolved feature.

Note that the weights in the feature detector remain fixed as it moves across the image, which is also known as parameter sharing. Some parameters such as

the weight values, adjust during training through the process of backpropagation and gradient descent. However, there are three hyperparameters which affect the volume size of the output that need to be set before the training of the neural network begins. These include:

1. The number of filters affects the depth of the output.
2. Stride is the distance, or number of pixels, that the kernel moves over the input matrix.
3. Zero-padding is usually used when the filters do not fit the input image.

After each convolution operation, a CNN applies a Rectified Linear Unit (ReLU) transformation to the feature map, introducing nonlinearity to the model.

The pooling of layers, also known as downsampling, reduces the number of parameters in the input. While a lot of information is lost in the pooling layer, it also has a number of benefits to the CNN. They help to reduce complexity, improve efficiency, and limit risk of overfitting.

The name of the full-connected layer aptly describes itself. As mentioned earlier, the pixel values of the input image are not directly connected to the output layer in partially connected layers. However, in the fully-connected layer, each node in the output layer connects directly to a node in the previous layer. This layer performs the task of classification based on the features extracted through the previous layers and their different filters.

3.6.2 Support Vector Machine

A SVM is a supervised learning technique from the field of machine learning applicable to both classification [183] and regression. Such a model can be used either for binary and multi-class classification, by switching to the *one-to-one* or *one-to-rest* formulation[184]. The SVM algorithm is widely used in machine learning because of its ability to handle both linear and nonlinear classification tasks.

Support Vector Machine (SVM) Terminology:

- Hyperplane: A decision boundary separating different classes in feature space, represented by the equation $wx + b = 0$ in linear classification.
- Support Vectors: The closest data points to the hyperplane, crucial for determining the hyperplane and margin in SVM.
- Margin: The distance between the hyperplane and the support vectors. SVM aims to maximize this margin for better classification performance.
- Kernel: A function that maps data to a higher-dimensional space, enabling SVM to handle non-linearly separable data.

- Hard Margin: A maximum-margin hyperplane that perfectly separates the data without misclassifications.
- Soft Margin: Allows some misclassifications by introducing slack variables, balancing margin maximization and misclassification penalties when data is not perfectly separable.
- C: A regularization term balancing margin maximization and misclassification penalties. A higher C value enforces a stricter penalty for misclassifications.
- Hinge Loss: A loss function penalizing misclassified points or margin violations, combined with regularization in SVM.
- Dual Problem: Involves solving for Lagrange multipliers associated with support vectors, facilitating the kernel trick and efficient computation.

The formulation of the standard SVM [185] is defined as a maximum margin classifier, that is, a classifier whose decision function is a hyperplane that maximally separates samples from different classes. Notationally, given a labeled training data set:

$$\{x_i, y_i\}_{\mathrm{i}=1^n}$$

where

$$x_i \in \mathbb{R}^N$$

and

$$y_i \in \{-1, +1\}$$

and given a nonlinear mapping $\varphi(.)$, the SVM method solves the following:

$$\min_{\omega,\xi,b} = \left\{ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \xi_i \right\}, \tag{3.11}$$

constrained to:

$$\begin{aligned} y_i(\langle \phi(xi) \rangle, w + b) &\geq 1 - \xi_i \qquad \forall i = 1, ..., n \\ \xi_i &\geq 0 \qquad \forall i = 1, ..., n \end{aligned} \tag{3.12}$$

where w and b define a linear classifier in $\mathbb{R}^N$ since x_i are in $\mathbb{R}^N$ and ξ_i are positive slack variables enabling to deal with permitted errors.

Note that minimization functional contains two terms with clear meaning: we try to minimize the committed errors, $\sum_{i=1}^{n} \xi_i$, while minimizing the norm of the model weights, $\|w\|^2$, which can be shown to be equivalent to the maximization of the margin (separation between classes). Note that one could just maximize the margin without paying attention to committed errors. This would give rise to the maximum margin SVM. By including the slack variables ξ_i one

relaxes the problem and the solution is called the *soft-margin* SVM, which minimizes the training error traded off against the margin. An appropriate choice of the nonlinear mapping f guarantees that the transformed samples are more likely to be linearly separable in the (higher dimension) feature space. The regularization parameter C controls the generalization capability of the classifier, and it must be selected by the user. The problem is solved using its dual problem counterpart, and the decision function for any test vector x^* is finally given by

$$f(x_*) = sgn(\sum_{i=1}^{n} y_i \alpha_i K(x_i, x_*) + b), \tag{3.13}$$

where α_i are the Lagrange multipliers corresponding to the constraints in Eq. 3.12, being the support vectors (SVs) those training samples xi with a nonzero Lagrange multiplier $\alpha_i \neq 0$; $K(x_i, x_*)$ is a function in x_* with parameter x_i; and the bias term b is calculated using any of the constraints corresponding to an unbounded Lagrange multiplier, as

$$b = \frac{1}{k} \sum_{i=1}^{n} (y_i - \langle \phi(xi), w \rangle), \tag{3.14}$$

where k is the number of unbounded Lagrange multipliers (i.e., $0 \leq \alpha_i < C$ being $\alpha_i \neq 0$) and $w = \sum_{i=1}^{n} y_i \alpha_i \phi(x_i)$.

To sum up, a SVM constructs a hyper-plane or set of hyper-planes in a high or infinite dimensional space, which can be used for classification, regression or other tasks. Intuitively, a good separation is achieved by the hyper-plane that has the largest distance to the nearest training data points of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier (Figure 3.4).

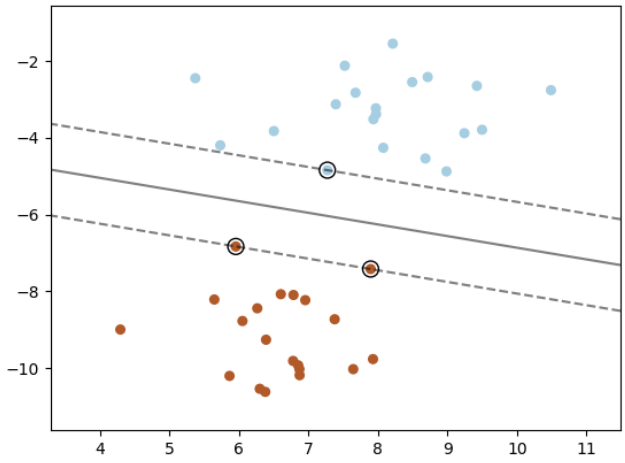


Figure 3.4: The figure shows an example of the decision function for a linearly separable problem, with three samples on the margin boundaries, called *support vectors*.

Linear SVMs are used with linearly separable data; this means that the data does not require any transformation to be separated into different classes. Mathematically, this separating hyperplane can be represented as:

$$wx + b = 0, \tag{3.15}$$

where w is the weight vector, x is the input vector, and b is the bias term. There are two approaches to calculate the margin, or the maximum distance between classes, which are *hard-margin* classification and *soft-margin* classification. If we use a hard-margin SVM, the data points will be perfectly separated outside the support vectors. This is represented by the formula:

$$(wx_j + b)y_j \geq a, \tag{3.16}$$

and then the margin is maximized, which is represented as:

$$\max_{\gamma} = \frac{a}{\|w\|}, \tag{3.17}$$

where a is the margin projected onto w.

Soft-margin classification is more flexible and allows some misclassification through the use of reserve variables. The hyperparameter C adjusts the margin; a larger value of C narrows the margin for minimal misclassification, while a smaller value of C widens it, allowing more tolerance for misclassification data.

However, most data in real-world scenarios are not linearly separable, and that's where nonlinear SVMs come in. To make data linearly separable, preprocessing methods are applied to the training data to transform it into a multidimensional feature space. That said, multidimensional spaces can create more complexity, increasing the risk of overfitting the data and becoming computationally expensive. The *kernel trick* helps reduce some of that complexity, making the computation more efficient, and it does so by replacing the dot product calculations with an equivalent kernel function.

There are several types of kernels that can be applied to classify data. Some popular kernel functions include: Polynomial kernel,RBF kernel (also known as Gaussian kernel) and Sigmoid kernel. When training an SVM with the Radial Basis Function (RBF) kernel, two parameters must be considered: C and γ. The C parameter, common to all SVM kernels, balances the misclassification of training examples with the simplicity of the decision surface. A low C makes the decision surface smooth, while a high C aims to correctly classify all training examples. Γ defines how much influence a single training example has. The larger the γ, the closer the other examples must be to be affected.

Chapter 4

Signal Processing Techniques for Telerehabilitation

In this chapter, will be examined at the types of signals that can capture patient movements, along with a quick overview of the kinds of noise that often show up in these recordings. Will also be go over the key features that can be pulled from the collected signals. The $Kinect\_v2$ sensor utilizes time-of-flight (ToF) technology, recognized as the standard for understanding human motion.

4.1 Microsoft Kinect

The Microsoft $Kinect\_v2$ is a motion sensing markerless input device based on a high-resolution RGB camera and an infrared ToF camera for depth analysis. A depth sensing camera automatically detects the presence of any nearby object and measures its distance to it. One of the most popular and commonly used depth technologies is the ToF; this term refers to the time taken by light to travel a given distance. ToF cameras work based on this principle where the distance to an object is estimated using the time taken by the light emitted to come back to the sensor after reflecting off the object's surface. A ToF camera, like Microsoft Kinect, can ensure patient anonymity in remote monitoring setups as their movements can be detected by acquiring depth data only and not RGB camera data, thus ensuring their privacy. The Microsoft Kinect offers a wide field of view (70x60 degrees) and recognition up to 4.5 [m] from the device [186]. Several studies have demonstrated that through the Microsoft $Kinect\_v2$ spatio-temporal parameters can be validly obtained [187, 188] and can be also a satisfactory tool for rehabilitation due to its low cost and adequate spatial accuracy (with an order of magnitude of centimeters) [189]. The set-up of the interface between the Microsoft $Kinect\_v2$ and the Unity3D engine is effortless because the manufacturer provides a Software Development Kit (SDK) and a Unity add-on, which gives developers access to body joint positions and orientations that can be used directly in rehabilitation game development. In ReMoVES the data and signals from the tracked user's body are recorded at a frequency of 10 Hz.

4.2 Kinect Joint Signals

A stochastic (or random) process is a mathematical description that links each possible outcome of an experiment to a function that changes over time. Every time the experiment is repeated — called a trial — it produces one possible version (or realization) of the process. In this context, the $3D$ signal representing joint $j-th$ in space (x, y, z), captured by the Kinect infrared depth sensor, is expressed as:

$$\{\mathbf{v}_j(t)\} = \begin{bmatrix} \{x_j(t)\} \\ \{y_j(t)\} \\ \{z_j(t)\} \end{bmatrix}, \tag{4.1}$$

with $j \in 1, \ldots, 25$, Figure 4.1.

More specifically, this is a digital signal recorded at discrete time intervals: $v_j(t) = v_j(kT_s)$, where the sampling frequency is $f_s = 30$ Hz and the sampling period is $T_s = 0.03$ s.

From these basic joint signals, many other useful measurements can be derived to analyze human movement. For example, it's possible to calculate displacements and trajectories in $3D$ space, on a specific plane, or along an axis, as well as determine speed, acceleration, and more. The distance, speed, and acceleration of individual points — or groups of points — can be measured in a similar way. Additionally, angles between joints can be computed from their $3D$ coordinates.

A time-dependent feature, denoted as $\psi(t)$, can be defined to represent either one of these basic signals or any quantity derived from them.

When analyzing movement, it's common to consider the person in a specific postural state, labeled Γ. The statistical properties of a random process — like mean or variance — are only meaningful if the process is stationary, meaning its behavior doesn't change over time. For complex, real-world signals like human motion, stationarity typically holds only when the signal is conditioned to a specific state, such as being at rest or performing a particular action.

To reliably analyze these signals and develop appropriate models for both the movement and any noise in the data, it's essential to set clear operating conditions. These conditions are equally important for all later stages of signal processing and analysis.

The three components of the signal $v_j(t)$ are considered independent, and under the assumption of stationarity, their average power is denoted by S_{ij}^2. For example, along axis i, the mean power of joint j, conditioned to a state Γ, is: $S_{ij}^2|\Gamma = E\{v_{ij}^2(t)|\Gamma\}$.

Stationarity is also a key requirement to verify ergodicity, a property that allows using time averages as substitutes for expected statistical values.

Finally, using advanced reconstruction techniques, markerless sensors like Kinect can generate $3D$ signals representing various points on the human body

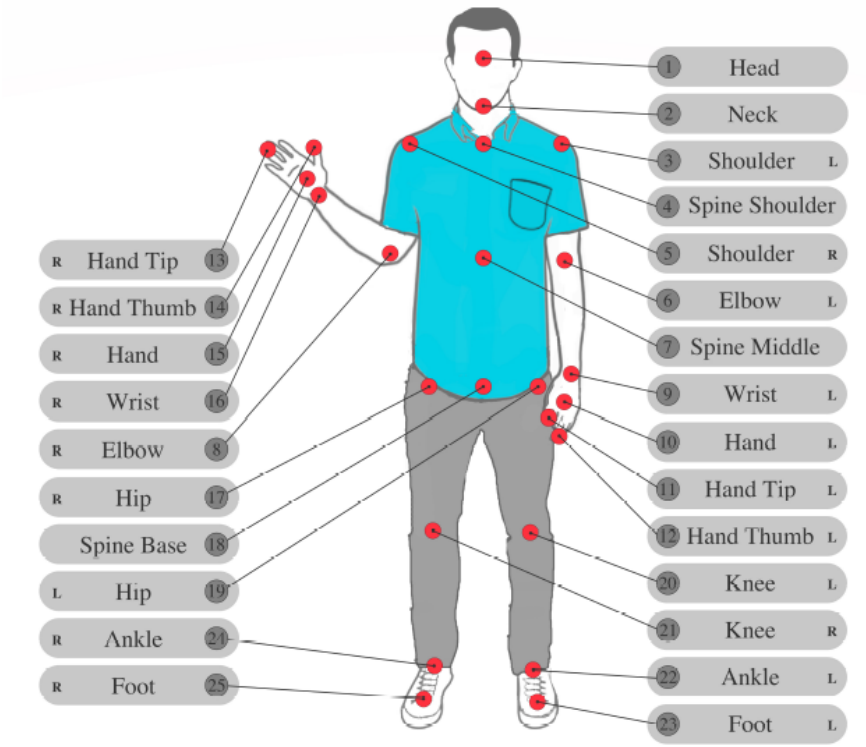


Figure 4.1: Joints tracked by Microsoft Kinect sensor.

— points that may correspond not only to the outer surface but also to an internal skeletal model.

4.3 Noise Model

The major signal noise sources in real applications are classical additive noise and multiplicative noise, as described in Equation 4.1 for the $j-th$ joint acquired by a generic sensor as an example:

$$\{\mathbf{v}_j(t)\} = \{v_{ij}(t)\} = \{(\mathbf{v}_{Bj}(t) + \mathbf{n}_j(t))\} \cdot \{\mathbf{m}_j(t)\}, \tag{4.2}$$

where $\mathbf{v}_j(t)$ is the acquired signal, $\mathbf{v}_{Bj}(t)$ is the noise-free signal, $\mathbf{n}_j(t)$ is the additive noise and $\mathbf{m}_j(t)$ is the multiplicative noise used to describe jitter and nonlinearities.

4.3.1 Additive Noise

The multidimensional additive noise, denoted as $\{n_{ij}(t)\}$, represents a zero-mean 3-component vector signal for each joint j:

$$\vec{n}_j(t) = \begin{bmatrix} n_{xj}(t) \\ n_{yj}(t) \\ n_{zj}(t) \end{bmatrix} = \{n_{ij}(t)\} \tag{4.3}$$

This type of noise is typically independent of the signal itself and is often assumed to be independent and identically distributed (IID). This implies that the covariance matrix is expressed as:

$$\Sigma_n = \mathbb{E}\left[\vec{n}_j(t) \cdot \vec{n}_k(t)^T\right] = \mathbb{E}\left[n_{ij}(t) \cdot n_{lk}(t)\right] \qquad \forall i, j, l, k \tag{4.4}$$

where it results in a diagonal form:

$$\sigma^2_{n_{ij}} = \text{constant} \qquad \forall i, j \tag{4.5}$$

4.3.2 Impulsive Noise

Among the various models for impulsive noise, one of the most widely used is the Poisson-Gaussian model. In this case, a sequence of Dirac pulses is generated according to a Poisson point process. Let $\{b(t)\}$ be this sequence, which is then amplitude modulated by an independent Gaussian random process $\{a(t)\}$.

This modulated sequence is then fed into a linear time-invariant (LTI) system, whose impulse response is defined as a rectangular pulse of duration T:

$$h(t) = \Pi\left(\frac{t}{T}\right) \tag{4.6}$$

The resulting output can be expressed as:

$$\{w(t)\} = \left\{ \sum_{m=-\infty}^{+\infty} a \cdot h(t - t_m) \right\} = \left\{ \sum_{m=-\infty}^{+\infty} a \cdot \Pi \left(\frac{t - t_m}{T} \right) \right\}, \tag{4.7}$$

where the pulse occurrences t_m are randomly distributed in time, following a Poisson process with a rate parameter λ. The amplitudes a are drawn from a Gaussian distribution, and the pulses have a fixed rectangular shape and duration.

4.4 The Kinect noise model

Occlusion caused by the body itself or individual joints, incorrect posture, scenarios in which skeletal joints fall outside the sensor field of view, or the distance between the human subject and the sensor are all factors that can negatively impact the accuracy of the Kinect measurement. In such situations, the estimation of joint positions becomes unreliable, leading to irregular tracking behavior and incorrect skeletal joint positioning.

Additionally, the computational power and memory capacity of the acquisition computer have a significant effect on the variable frame rate at which Kinect data is acquired. By combining a low-cost, portable sensor with appropriate signal processing techniques, it is possible to develop systems capable of achieving satisfactory performance for most medical and rehabilitation applications.

Furthermore, the additive and multiplicative noise characteristics of the Kinect have been empirically investigated, and based on these studies — as well as existing literature — the following conclusions can be drawn.

4.4.1 Additive Noise in the Kinect Sensor

Additive noise affecting the Kinect sensor is independent of the underlying signal, with its components being independent but not necessarily identically distributed. As a result, the covariance matrix is diagonal, although its diagonal elements may differ, unlike in the case of IID noise. Specifically, for any joints j, k and any axes i, l, the following holds:

$$\sigma^2_{n_{ij}} \neq \sigma^2_{n_{lk}}, \tag{4.8}$$

According to this definition, $\sigma^2_{n_{ij}}$ represents the variance, and therefore the average power of the noise at a given time instant.

4.4.2 Impulsive Noise in the Kinect Sensor

A statistical analysis of the Kinect signal, as supported by the literature, reveals that signal losses may randomly occur at certain independent time instants. These events can be modeled as a type of multiplicative noise.

A suitable statistical model for this behavior involves an unmodulated Poisson process. The sequence $\{b(t)\}$, consisting of ideal Dirac pulses following a Poisson distribution, is used as input to a system whose impulse response is a rectangular pulse of random duration d. Given the impulse response:

$$h(t) = \Pi\left(\frac{t}{d}\right) \tag{4.9}$$

the randomness is associated with the duration of the pulse, while the amplitude remains fixed (typically at unit value). The resulting signal can then be expressed as:

$$\{w(t)\} = \left\{ \sum_{m=-\infty}^{+\infty} \Pi\left(\frac{t - t_m}{d}\right) \right\}, \tag{4.10}$$

where t_m represents the random occurrence times of the impulses.

The loss of signal along a given axis i of joint j can be modeled using a multiplicative noise factor:

$$m_{ij}(t) = \{1 - w(t)\}, \tag{4.11}$$

Finally, by construction, this noise process behaves as a binary stationary process, characterized by the following probability density function (PDF):

$$f_m(M) = P_0 \cdot \delta(M) + P_1 \cdot \delta(M - 1), \tag{4.12}$$

where P_0 and P_1 represent the probabilities associated with the two possible states of the process:

$$P_0 = P\{m(t) = 0\} = P\{w(t) = 1\}. \tag{4.13}$$

4.5 Preprocessing and Kinect Noise Filtering

The $Kinect\_v2$ sensor employs a machine learning approach—specifically, a random forest classifier trained on a large dataset of human poses—to estimate $3D$ skeletal points from depth camera data. However, skeletal tracking can be affected by several common noise-related issues, including jitter, flickering, and missing samples. Jitter refers to small, rapid fluctuations in joint positions, typically caused by sensor noise or misclassifications, which result in shaky and unstable motion. Flickering involves sudden, noticeable jumps in joint locations, often due to occlusions or inconsistencies in the classifier, leading

to erratic tracking behavior. Missing samples occur when joint data are absent for certain frames, causing breaks in the continuity of the skeletal data, with joints intermittently disappearing and reappearing. Most research on filtering Kinect signals focuses primarily on reducing jitter and smoothing the skeletal motion data to improve overall tracking quality [190, 191, 192, 193].

To address the noise issues affecting the Microsoft Kinect, the following procedure has been proposed. Initially, the zero-value sequences within the signal are automatically identified. These missing or invalid samples are then reconstructed using spline interpolation. Since the primary form of noise present in the signal is impulsive, a non-linear filtering approach is necessary. However, a standard median filter proves ineffective in this context due to the high density of noise peaks.

As a solution, an adaptive filtering technique was developed, combining outlier detection with median-based substitution. This method adjusts the denoising process according to the statistical characteristics of the signal. Specifically, once outliers are identified, they are replaced with the median value computed within a moving window of size fifteen. This approach effectively suppresses the noise while preserving the integrity of the original signal as much as possible.

Following the removal of impulsive noise, the resulting error can be expressed as:

$$\sigma^2_{e_{ijF}} = P_1 \cdot \sigma^2_{n_{ij}} \tag{4.14}$$

The expected reduction in error is reflected by the error reduction factor:

$$\frac{\sigma^2_{e_{ijF}}}{\sigma^2_{e_{ij}}} = \frac{P_1}{1 + P_0 \cdot (SNR - 1)} \tag{4.15}$$

where the Signal-to-Noise Ratio (SNR) is defined as:

$$SNR = \frac{\sigma^2_s}{\sigma^2_{n_{ij}}} \tag{4.16}$$

It can be observed that the error decreases more consistently when both the SNR and P_0 values are large.

After addressing the non-linearities, a final low-pass filtering step is applied to further smooth the signal.

4.6 Interpolation

Interpolation is a mathematical technique used to estimate unknown values within the range of a discrete set of known data points on a Cartesian plane.

Given a sequence of N distinct real values x_k and their corresponding values y_k, the goal is to determine a function f that models the relationship between these two sets in such a way that:

$$f(x_k) = y_k \qquad \text{for } k = 1, \ldots, N \tag{4.17}$$

Each pair (x_k, y_k) is called a data point, while f is referred to as the interpolating function. The various interpolation methods differ in terms of accuracy, computational complexity, required number of data points, and the smoothness of the resulting curve.

In this work, spline interpolation was employed for the analysis of Kinect signals, as detailed below.

4.6.1 Spline Interpolation

A spline function $f_m : [a, b] \to \mathbb{R}$ is a piecewise polynomial function of degree m, such that:

$$f_m \in C^{m-1}([a, b]) \tag{4.18}$$

which means its derivatives up to order $m-1$ are continuous across the entire interval $[a, b]$. Given a set of data points (x_k, y_k) ordered such that $x_k < x_{k+1}$ for $k = 1, \ldots, N$, the spline must satisfy:

$$f_m(x_k) = y_k \qquad \text{for } k = 1, \ldots, N \tag{4.19}$$

In particular, when $m = 3$, cubic splines are obtained. These are functions composed of polynomials of degree 3 in each subinterval $[x_i, x_{i+1}]$ $(i = 1, \ldots, N-1)$ and are globally twice continuously differentiable, i.e. they belong to the class $C^2([a, b])$. In each subinterval, the cubic spline takes the form:

$$f_3(x) = a_i x^3 + b_i x^2 + c_i x + d_i \qquad \text{for } x \in [x_i, x_{i+1}] \tag{4.20}$$

In this study, sections of the signal affected by zero runs were automatically identified, and classical spline interpolation was applied to reconstruct the missing values. The results obtained proved to be reliable and precise, as illustrated in Figure 4.2. A possible future enhancement could involve the adoption of probabilistic interpolation approaches, such as those proposed by the authors in [194].

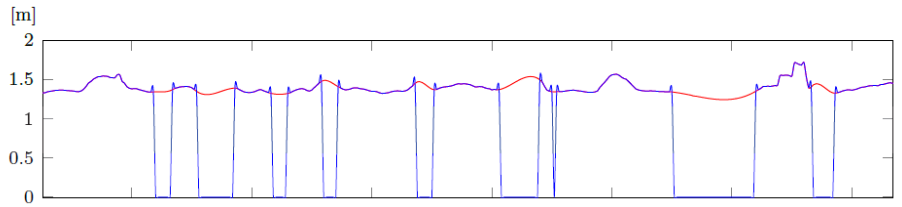


Figure 4.2: In blue the original signal, in red the interpolation

4.7 Chebyshev outlier detection method

Impulsive fluctuations were observed in all three spatial components of the signals, probably due to temporary interruptions in the Kinect joint tracking system. Since significant or high-frequency calibration errors can distort signal trajectories and introduce inaccuracies in the computation of clinical metrics, it was necessary to apply appropriate filtering techniques prior to any further analysis.

To manage outliers, values whose deviation from the signal mean exceeded λ times the standard deviation were removed. As the precise statistical distribution of the signals is not fully known, a conservative approach based on the Chebyshev inequality [195, 196] was adopted to determine a lower bound for the probability:

$$P(|x - \mu| \leq \lambda\sigma) \geq 1 - \frac{1}{\lambda^2} \tag{4.21}$$

where x represents a generic signal sample, and μ and σ^2 are the estimated mean and variance, respectively. Setting $1 - \frac{1}{\lambda^2} = 0.89$, values differing by more than $\lambda = 3$ times the standard deviation from the mean are classified as outliers.

Additionally, the Chebyshev inequality provides an upper bound for the probability of observing a deviation greater than $\lambda\sigma$:

$$P(|x - \mu| \geq \lambda\sigma) \leq \frac{1}{\lambda^2} \tag{4.22}$$

The application of this outlier detection strategy follows recommendations from previous studies [197]. Specifically, lower and upper Outlier Detection Values (ODVs) were computed, and signal values exceeding these thresholds were flagged and handled as outliers.

To determine which values are outliers, the following steps are performed:

1. The expected probability p of seeing outlier is decided.
2. The probability p value is used to find λ using Equation 4.22:

$$\lambda = \frac{1}{\sqrt{p}} \tag{4.23}$$

3. The upper and lower ODVs are calculated using the following equations:

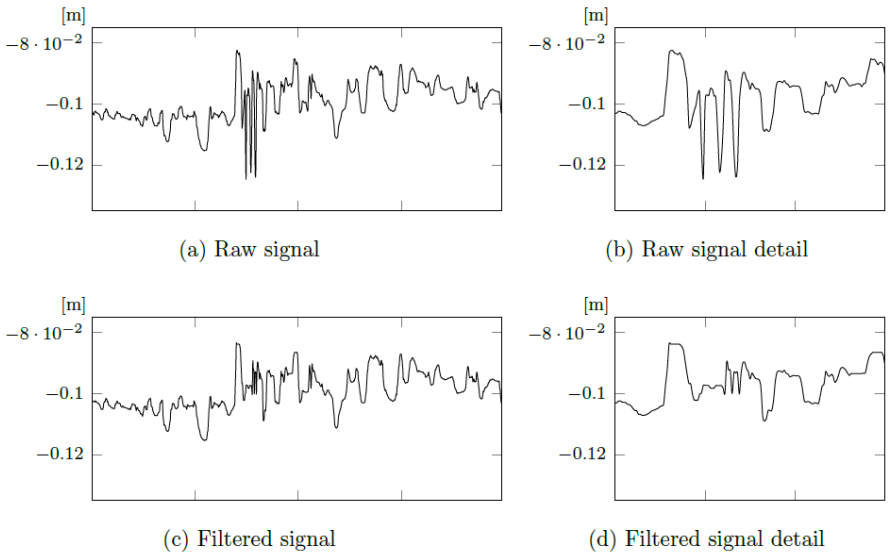
$$ODV_u = \mu + \lambda * \sigma, \tag{4.24}$$

$$ODV_l = \mu - \lambda * \sigma, \tag{4.25}$$

   All data larger than ODV_u or smaller than ODV_l are considered outlier.

4. The outlier values are replaced by the median value in a window of dimension fifteen.

An example of outlier detection operation is provided in Figure 4.3.



(a) Raw signal

(b) Raw signal detail

(c) Filtered signal

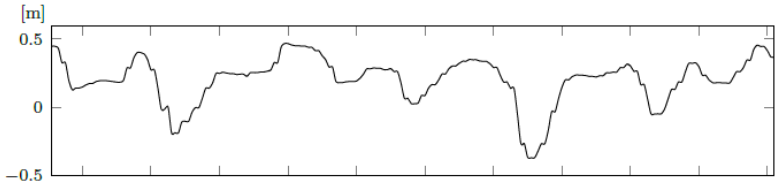
(d) Filtered signal detail

Figure 4.3: Signal filtering action through the Chebyshev method - full signal and frame detail.

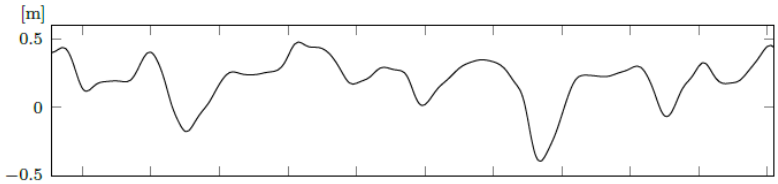
4.8 Butterworth filter

After identifying and replacing the outliers, a low-pass Butterworth filter is applied to the signal, Figure 4.4. The data acquired from the sensor exhibit stationarity around their mean and contain no relevant information at higher frequencies. The purpose of the Butterworth filter [198] is to maintain a maximally flat frequency response within the passband.

On a Bode plot, the filter's frequency response decreases linearly towards negative infinity (Figure 4.5, [199]). The attenuation rate for a first-order Butterworth filter is 20 [dB] per decade, increasing to 40 [dB] per decade for a second-order filter, and becoming progressively steeper with higher-order filters.



(a) Portion of the raw signal.



(b) Portion of the filtered signal.

Figure 4.4: Signal filtering action.

The frequency response of an n^{th}-order Butterworth filter is given by:

$$|T(j\omega)| = \frac{1}{\sqrt{1+\omega^{2n}}} \tag{4.26}$$

where n is the order of the filter, ω is the ratio between the frequency of the signal and the cut-off frequency. For signal regularization, after few tests a 2^{nd} order Butterworth filter with cut-off frequency of 2 [Hz] was implemented. An example of such filtering operation is provided in Figure 4.6.

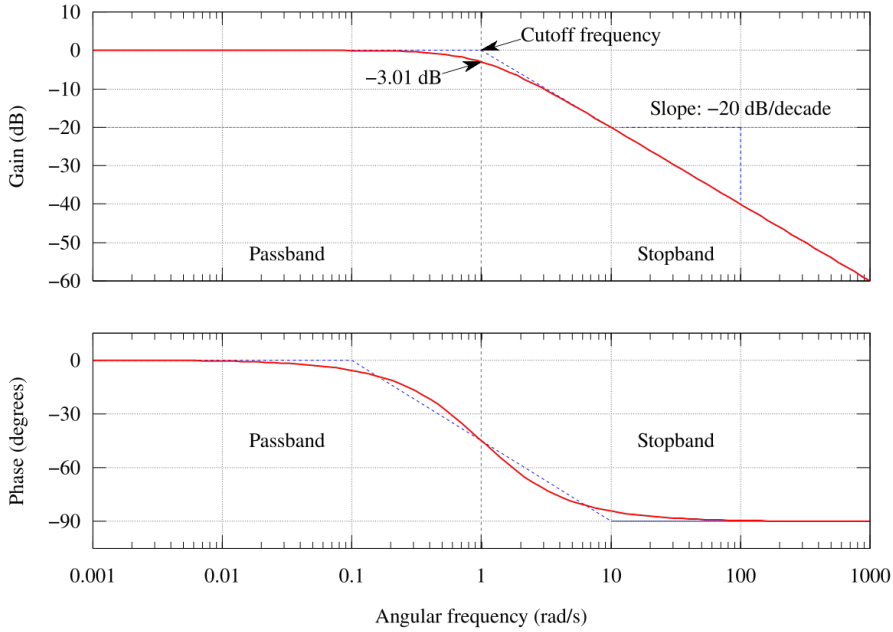
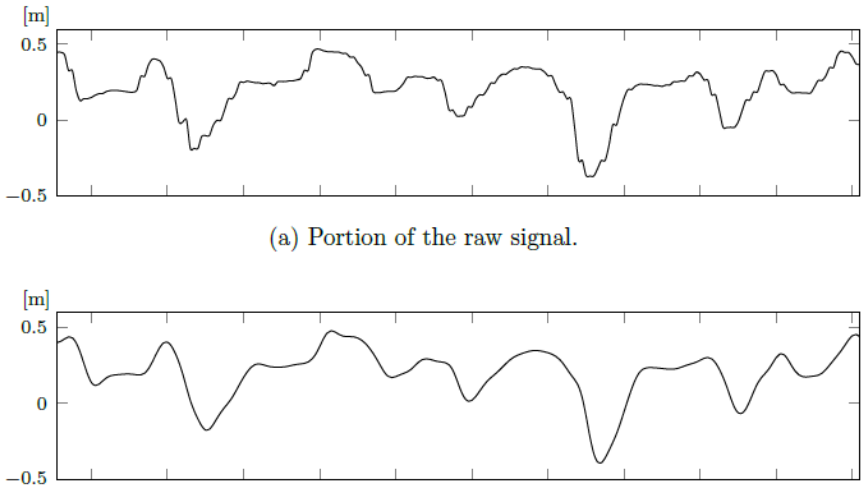


Figure 4.5: The frequency response of a first-order low-pass Butterworth filter.



(a) Portion of the raw signal.

(b) Portion of the filtered signal.

Figure 4.6: Signal filtering action with the low-pass Butterworth filter.

4.9 Motion Segmentation

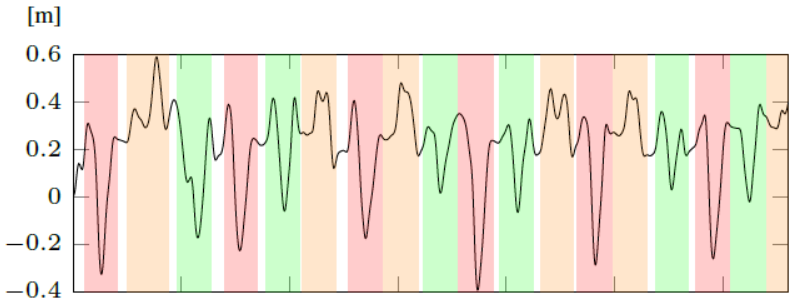
Rehabilitation exercises typically involve repetitive movements, resulting in signals that exhibit a certain degree of pseudoperiodicity. These movements are often composed of two or more elementary actions or phases, such as flexion-extension, adduction-abduction, or transitions between standing and sitting. For this reason, a signal segmentation phase is essential to ensure a reliable and meaningful analysis. This phase focuses on isolating stationary portions of the signal corresponding to homogeneous states, denoted as Γ. Through segmentation, complex motor sequences are broken down into their elementary actions, and the segmentation strategy must be specifically adapted to each exergame, according to the required movements and the joints involved.

The execution of a generic movement M_α generates a realization of the stochastic process associated with that particular experimental trial, giving rise to the multidimensional random process $M_{\alpha,j}(t)$. As each trigger event corresponds to a different movement that the patient is prompted to perform by the system, this information is crucial for identifying the precise instant when a specific movement should begin (e.g., α, β, etc.). Exploiting the pseudorepetitive nature of rehabilitation exercises, the segmentation process focuses on detecting these trigger events, which are closely linked to the structure of the exergame being analyzed. In this way, the anticipated movements can be represented as $W_\alpha(t)$, $W_\beta(t)$, and so on, indicating the expected signal trajectories for the correct execution of each movement.

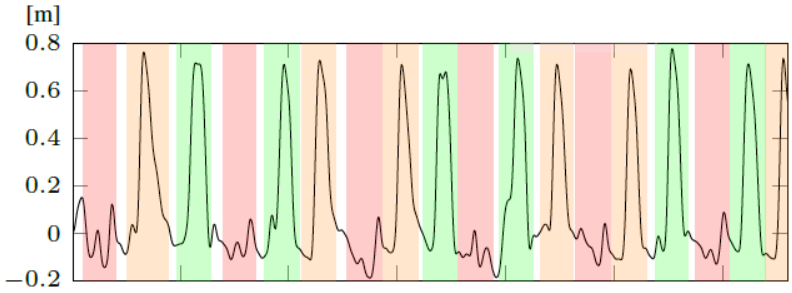
Decomposing signals generated by complex actions into sequences of elementary movements is fundamental both for studying the patient's behavioral patterns and for assessing the quality of rehabilitation exercises. In particular, for the extraction of key performance indicators — such as execution speed, range of motion, and movement trajectory — it is necessary to isolate and analyze the primitive actions individually to achieve a more precise and accurate evaluation.

In this study, segmentation was performed using additional signals to identify the start and end of each movement. Specifically, in the *Shelf Cans* activity, which requires the execution of three distinct movements combining flexion-extension and adduction-abduction of the upper limb, movements M_α, M_β, and M_γ were defined to correspond to placing the red, orange, and green cans onto their respective shelves.

A multilevel signal called *CanColor*, acquired during the game session, was employed to guide segmentation. This signal consists of four discrete levels indicating the state of interaction: 0 — no can held, 1 — green can, 2 — orange can, 3 — red can. By monitoring this signal, it was possible to isolate and separately analyze the three trajectories performed by the patient. Segmentation begins at the moment the patient picks up a can and ends when it is placed on a shelf. An example of this segmentation process is illustrated in Figure 4.7.



(a) X coordinate.



(b) Y coordinate.

Figure 4.7: Signals segmentation of X and Y coordinates of the right hand joint.

4.10 Features

The ReMoVES system captures data and signals from each game session, including meta-features like the game name, start and end times, final score, and additional game parameters to support efficient database querying. All this data are collected in JSON files.

Below the indicators and parameters extracted based on the application or type of exergame analyzed in this thesis will be described.

Beyond patient prescriptions and other meta-features [200], the primary features analyzed here include kinematic instant features and dynamic features. An automatic signal segmentation step is proposed, as described in the previous subsection, which automatically detects the start and end times of each repetition of movement. After this automatic localization of the instantaneous time referring to some specific positions of the subject, the precise measurement of each joint constitutes the instant features that allow one to understand the posture and compare it with the posture indicated by the medical staff through

the exergame. In particular, angles extracted from joint positions are of clinical interest.

Dynamic features describe the kinematic movement that the subjects performed and refer to the Range of Motion (ROM), the difference between the maximum and minimum angles in a given period of interest, the actual temporal displacement of the joints, or the change in angles during a specific movement.

The base features are the joint positions at an instant point, which are directly accessible by the signal punctual value. Even though coordinates in $3D$ space are available for each joint at a given time, their position on only one axis or anatomical plane is often of interest. Starting from base features, the most relevant derived feature extracted by processing and combining the acquired signals is the distance between joints and instantaneous angles.

As an example, the difference between the j and k joint positions on the i axis is defined as $\Delta i(t) = v_{ij}(t) - v_{ik}(t)$.

Angles are computed by applying the absolute or relative angle formula, depending on the specific analysis. An absolute angle of a body segment linking two joints j and k in a given body plane is measured with respect to one of the axes by using the *arctangent* formula:

$$\theta = arctan\left(\frac{||\vec{v} \times \vec{u}||}{\vec{v} \cdot \vec{u}}\right) \tag{4.27}$$

where $\vec{v}$ and $\vec{u}$ are generic vector, $||\vec{v} \times \vec{u}||$ indicates the magnitude of the cross product and $\vec{v} \cdot \vec{u}$ the dot product.

The coordinates of the Kinect joints, numbered as shown in Figure 4.1, will be used as a reference framework. In this system, the X-axis corresponds to the horizontal (mediolateral) direction of motion, the Y-axis is orthogonal to the X-axis and points vertically upwards, and the Z-axis is perpendicular to the X-Y plane, oriented in the outward (anteroposterior) direction.

Accordingly, the variables x, y, and z indicate the mediolateral, vertical, and anteroposterior directions, respectively, as illustrated in Figure 4.8 and defined in [201].

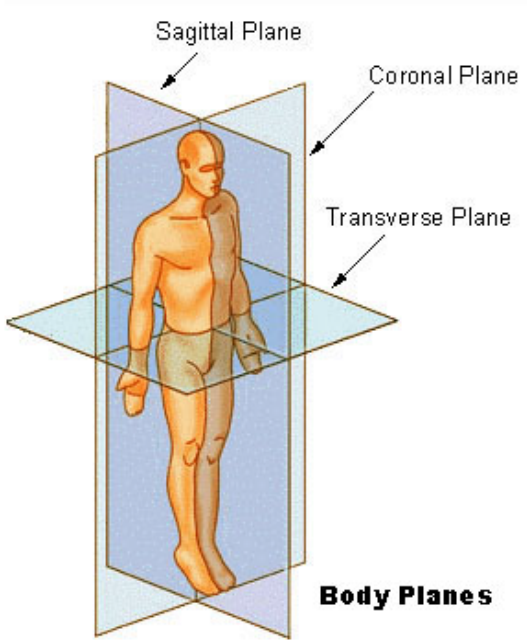


Figure 4.8: Reference anatomical planes [1].

4.10.1 Upper-limb movement

Patients data have been analyzed both with $3D$ coordinates of Kinect joints and through $2D$ data obtained by the exergame itself.

In trigonometry, in any triangle, the square of one side is equal to the sum of the squares of the other two sides decreased by the double product of these two sides multiplied by the cosine of the angle they form. Using notation as in Figure 4.9, the law of cosines states:

$$a^2 = b^2 + c^2 - 2bc\cos\alpha \tag{4.28}$$
$$b^2 = a^2 + c^2 - 2ac\cos\beta \tag{4.29}$$
$$c^2 = a^2 + b^2 - 2ab\cos\gamma \tag{4.30}$$

This theorem is used to find one of the angles of the triangle if the three sides are known:

$$\gamma = arccos\left(\frac{a^2 + b^2 - c^2}{2ab}\right). \tag{4.31}$$

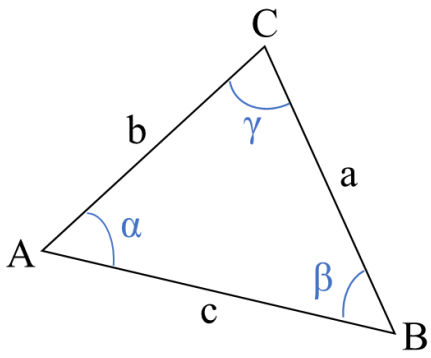


Figure 4.9: A triangle with a, b, and c sides and the corrisponding α, β and γ angles.

Referring to the upper-limb movement, consider a fixed time, the possible indicators are the ROM of the shoulder, and elbow in the coronal plane, defined as:

$$ROM(\theta_{shoulder}) = max\theta_{shoulder} - min\theta_{shoulder},$$
$$ROM(\theta_{elbow}) = max\theta_{elbow} - min\theta_{elbow}$$

where the angles are described by the following equations:

$$\theta_{shoulder} = arctan\frac{z_{elbow} - z_{shoulder}}{x_{elbow} - x_{shoulder}},$$

$$\theta_{elbow} = \theta_{shoulder} - arctan\frac{z_{elbow} - z_{wrist}}{x_{elbow} - x_{wrist}}.$$

Possible compensations with the trunk can be observed using the following formula, in the sagittal plane:

$$ROM(\theta_{trunk}) = max\theta_{trunk} - min\theta_{trunk},$$

where

$$\theta_{trunk} = arctan\frac{z_{spine\_shoulder} - z_{spine\_middle}}{y_{spine\_shoulder} - y_{spine\_middle}}.$$

The characteristic ROM of the elbow and shoulder joints in the three anatomical planes and in $3D$ space was calculated. Based on Equation 4.30 the elbow angle in the coronal plane is the following:

$$a = \sqrt{(x_5 - x_8)^2 + (y_5 - y_8)^2}$$
$$b = \sqrt{(x_{16} - x_8)^2 + (y_{16} - y_8)^2}$$
$$c = \sqrt{(x_5 - x_{16})^2 + (y_5 - y_{16})^2}$$

where the joints 5, 8, and 16 are respectively for right shoulder, right elbow, and right wrist. From these equations the θ_{elbow} is derived:

$$\theta_{elbow} = arccos\frac{a^2 + b^2 - c^2}{2ab}. \tag{4.32}$$

For many activity of daily living, the angles required reach at least 150°for the elbow flexion and 130°for shoulder flexion and abduction [202]. Other indicators are analyzed, such as

- Target: the number of object correctly taken.
- Angle between trajectories: the three specific tasks of Shelf Cans, referring to the differently colored cans, are split. Then, the straight line connecting starting position and targets is computed, and will be hereinafter referred to as the optimal trajectory. In addition, the so-called approximate trajectory performed by the patient during the game session is computed as the regression line of hand-game positions during the considered task. The lower the angle between the two lines, the better and controlled movement was done by the patient. Indeed, small angles shows that the approximate fitted path is similar to the optimal one. Conversely, large angles are typical of trajectories which are far from the optimal one. The linearity of scope-oriented movement is usually valued during physical therapy for patients affected by pathologies of motor learning such as cerebellar stroke. The importance of such an indirect analysis relies on the fact that it allows for quantifying the degree of improvement of the pathology which has caused the deficit of movement, and also to quantify the motor learning.
- Hand-Shoulder distance: the Euclidean distance between the hand and the shoulder joints while performing the movement is evaluated in order to assess whether the patient reaches the targets by performing a correct flexion–extension or abduction–adduction movements or makes a compensatory motion with the trunk instead, implying that in the latter case the Hand-Shoulder distance will be lower. The equation is the following, where joint 15 is for the right hand:

$$d = \sqrt{(x_5 - x_{15})^2 + (y_5 - y_{15})^2 + (z_5 - z_{15})^2} \tag{4.33}$$

- Time: the three specific tasks required by the exercise are split. The time taken to place the can on the shelf is calculated.

4.10.2 Cognitive assessment

The neuropsychological battery consists of tests that are commonly used for cognitive assessment in individuals with disabilities. Specifically, the Brief International Cognitive Assessment for MS (BICAMS) and the Paced Auditory

Serial Addition Task (PASAT) at 3 s and 2 s intervals were used. BICAMS includes the Symbol Digit Modalities Test (SDMT), the California Verbal Learning Test II edition (CVLT-II), and the Revised Brief Visuo-Spatial Memory Test (BVMT-R). Patients were recruited if they scored below the 5th percentile for normative data adjusted for age, sex, and education in at least two of the aforementioned tests. Written informed consent was obtained from all participants before the study began.

All cognitive measures were administered at baseline (i.e., time T0), at the end of 10 exergame sessions (T1), and one month after the end of treatment (T2).

In addition to these neuropsychological tests conducted at three fixed moments, daily assessments of motor function were available, thanks to the parameters observed remotely during the execution of the exergames. This continuous monitoring allowed a more detailed and dynamic understanding of the patient's motor progress. The patient's perception of effort and fatigue is indirectly assessed through daily observations of key parameters, which will be detailed in the results section. These parameters include the number of exergames played, the number of repeated movements along with their speed and trajectory, game scores, posture, angles, range of motion, etc. Comparing the exercises actually performed with those prescribed is crucial for understanding both the patient's level of engagement and the appropriateness of the exercise regimen. This approach provides valuable insights into how well the exercises are tailored to the patient's abilities and needs, helping to optimize the rehabilitation process.

Chapter 5

Application of diagnostic imaging techniques

This chapter discusses studies in the context of image processing. First, a reflection on the use of the SRAD as a initial processing of image. Second, a Focus-of-Attention Mechanism is introduced, with an application regarding the enhancement of Glisson's capsule in US images of patients with CLD. Issues related to improvement techniques are then addressed. The use of a binary mask, obtained through an adaptive method, is introduced and used to extract the Glissonian line to allow its feature analysis with SVM. Finally, a pipeline for evaluating the results obtained through CNN ranking is presented.

5.1 Available Dataset

This retrospective study, conducted at Ospedale Policlinico San Martino IRCCS in Genoa, involved 215 CLD patients (male and female) with an average age of 57.84 $\pm$ 13.39 years (range 20–92). Patients were diagnosed with conditions such as non-alcoholic fatty liver disease (NAFLD), hepatitis C (HCV), hepatitis B (HBV), autoimmune hepatitis (AIH), and primary biliary cholangitis (PBC). The dataset (Figure 5.1) included 392 liver ultrasound images acquired using the $MyLab^{TM}X9$ Esaote machine, by Esaote S.p.A. .

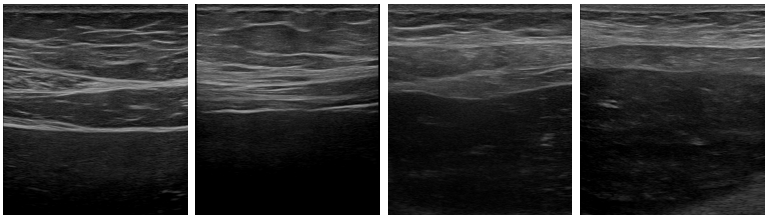


Figure 5.1: US original images from Esaote $MyLab^{TM}X9$ ultrasound scanner. From left to right, the stages correspond to fibrosis progression from F0–F1 to F4.

Despite convex probes being commonly used for liver diagnosis, scanning in this study was performed with a linear probe to focus on the liver's third segment and better highlight Glisson's capsule. All acquisitions were conducted by a single experienced operator. Annotation of fibrosis staging, according to

METAVIR score [203], was derived from elastographic exams using shear wave and FibroScan elastography. Only patients with consistent results from both methods were included. The population was divided into four groups based on fibrosis severity: 98 patients with F0-F1 (no/mild fibrosis), 79 patients with F2 (moderate fibrosis), 76 patients with F3 (severe fibrosis), and 104 patients with F4 (cirrhosis).

To enhance training data for a shallow convolutional neural network, the ultrasound images were augmented using image processing techniques, as described in the following subsection.

5.2 Data Preparing

Image acquisition by the Esaote ultrasound machine $MyLab^{TM}X9$ provided scans of different sizes (about 600×600) in a 3-channel, 40 kB average size and BMP format. Before constructing the database, all the images underwent an anonymization process. They were cropped, converted to grayscale, and resized (scaled to have equal height and width) to a resolution of 150×150 pixels. This action helps to reduce the number of pixels in an image (not to mention *noise* to deal with), leads to faster and more accurate image processing algorithms and can therefore reduce the training time of a neural network since the greater the number of pixels in an image the greater the number of input nodes which in turn increases the complexity of the model.

Bilinear interpolation was then used to perform image scaling. Through interpolation, new data points can be constructed within the range of a discrete set of known data points. In the bilinear interpolation technique, one does not simply take the value of the closest pixel as in nearest-neighbor interpolation, but considers the 4-connected pixels and uses them to calculate an interpolated value. Instead of simply replicating the value of the nearest-neighbor pixel, this method considers the values of the surrounding pixels and calculates a weighted average to determine the value of the new pixel in the resized image. This process produces more uniform and accurate results than nearest-neighbor interpolation and helps avoid the blocking or pixelization effect that can occur with other interpolation methods.

Since much of the informative content is concentrated on the Glisson line, whose morphological changes provide insight into the staging of fibrosis, the proposed image processing approach mainly focuses on morphological features. This contrasts with traditional ultrasound processing, which typically emphasizes textural features.

To this end, a preliminary step is performed to prevent the classification process from being influenced by irrelevant information, such as speckle noise. Consequently, the pre-processing step focuses on reducing the prominence of the lower parenchyma region, which is heavily affected by speckle, as well as the

upper portion of the image, which depicts various tissues unrelated to Glissonian line features that we seek to enhance. Based on the typical acquisition defined in the present approach, we exclude the lower and upper portions of the image from processing, thereby defining the *region of interest* (ROI). We then apply an adaptive image transformation aimed at suppressing textural information related to speckle noise and interfaces between other organs and tissues outside the liver, which could act as distractors in the classification step. This results in the creation of the so-called *region of contrast interest* (ROCI).

5.3 Working Method

The data has been pre-processed to reduce the noise information in the images, then, the size of the dataset is increased with data augmentation strategy, so that the use of machine learning techniques for classification is enabled. A general overview of the work can be presented as follows.

Ultrasound image It is well known that radiologists and sonographers use characteristic echo patterns and transitions at organ interfaces to make accurate diagnoses. Anechoic areas appear completely black because they do not reflect any ultrasound waves, typically indicating the absence of structures. Hypoechoic areas appear darker than the surrounding tissues but are not completely black, suggesting that the tissue or structure in that region reflects fewer ultrasound waves. A hyperechoic signal, on the other hand, is caused by structures such as bone, calcifications, fibrous tissues, or fat. Significant transitions between hypoechoic and hyperechoic signals (or vice versa) typically correspond to the interfaces between different organs and tissues.

Glisson capsule In ultrasound images, the Glisson capsule appears in the shape of a line which will be hereinafter referred to as the Glisson line. A bright and continuous line represents a smooth surface while a dashed and dark line means that the Glisson capsule is irregular on the surface.

SRAD filter As described in Paragraph 3.2.1, the quality of an ultrasound image is low because of the grain pattern noise that occurs in the ultrasound image acquisition process (speckle noise). When the SRAD filter is applied to the ultrasound image, the resulting filtered image is not exactly privatized from its noisy component. However, the richness of information in ultrasound images stems from the strong textural properties caused by speckle [204, 205], which is not considered noise but rather valuable information for characterizing tissue properties. This textural information forms the basis for most analyses. CNN kernels and their related features heavily depend on this type of textural information, often at the expense of other features, such as morphological ones, which are central to the present analysis. For this reason, in this thesis, a procedure has been developed in which noise filtering is not performed but rather a specific Focus-of-Attention Mechanism for the ultrasound image.

Focus-of-Attention Mechanism During the pre-processing phase, an image transformation technique was applied (Paragraph 3.2.2). Ultrasound images are typically affected by low-contrast levels, making it difficult to detect pathological signs. This technique aims to highlight the features of interest of the Glisson line contained in the ROI while reducing the relevance of elements that might introduce confusion, such as textural information. This results in a transformed image called ROCI. The piecewise-linear transformation shrinks dark and light intensities while stretching the intermediate values to increase the likelihood of a relevant region's correct analysis.

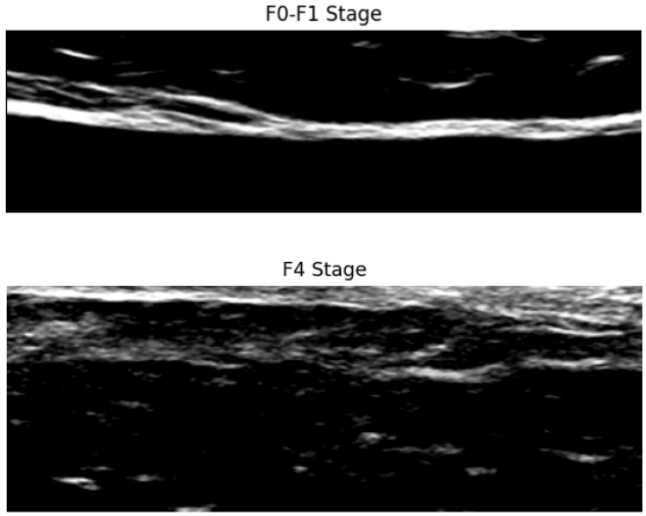


Figure 5.2: ROI of the ultrasound image after image processing. Top: mild stage F0-F1; bottom: cirrhotic stage F4.

To implement an adaptive transformation, a histogram-based transformation is proposed (Eq.3.1), similar to the one commonly used for contrast stretching, as described in Figure 3.1 and explained below, where U is the original gray level and V is the gray level in the transformed image.

The parameters a and b are adaptively set for each image to divide the dark (i.e., anechoic), medium (hypoechoic), and bright (hyperechoic) regions. The slopes α, β and γ are used for determining the relative transformation in these three regions.

From this modification of the histogram, we obtain a new image I_{out} : $(x, y) \rightarrow V$. Let a and b denote the lower and upper limits, respectively, where a represents the mean value of the parenchyma, while b represents the mean value of the other tissues. Both are calculated for each image and estimated based

on the mean value of the lower and upper portions of the ultrasound image divided vertically into three parts, respectively. This approach ensures that the technique remains adaptive. The corresponding transformed output values, V_a and V_b, are set to 50 and 200, respectively.

This method aims to *enhance* ultrasound images by expanding the intensity range of central pixels while shrinking the extremes. By applying contrast stretching transformation to the ROIs, we generate additional images alongside the original ultrasounds. These images, referred to as transformed images, emphasize the information content of the Glissonian membrane while attenuating—or even eliminating—details of other anatomical structures and potential artifacts, thus defining the ROCIs.

In doing so, an ad hoc data set is created, which can be provided as input to the neural network, focusing the classification on the key features of interest.

Data Augmentation Following this transformation, both the original and transformed images are used for network training. To further enhance the dataset, small rotations along the probe axis are simulated for both image types. This augmentation process results in an expanded dataset, which is then fed into the neural network for training.

Unlike other pathologies or conditions, such as cases involving tumors or lesions where abnormalities are clearly visible in ultrasound scans, liver staging relies on subtle contrast variations—particularly their presence, continuity, or absence.

While many studies employ traditional data augmentation techniques solely to increase the number of samples, leaving the task of feature extraction entirely to the neural network, our approach is different. Here, the application of contrast stretching not only augments the dataset but also enhances the quality of information within the image. This transformation helps guide the network's attention toward the key features of the Glissonian line, improving its ability to learn relevant patterns for liver staging.

In conclusion, from a numerical perspective, after applying the Focus-of-Attention mechanism and rotation, we obtain a dataset consisting of 830 elements (where each element refers to a combination of the original ultrasound image, its transformed version, and its rotated variant)divided into training and validation sets. Considering the images for testing as well, the final result is a dataset consisting of training, validation, and test sets, with proportions of 80%, 10%, and 10%, respectively. This setup enables 10-fold cross-validation and helps mitigate overfitting. N_t denotes the size of the test set.

Binary Mask and Line Extraction From the histogram analysis it can be seen that the four different stages have a pattern of occurrences that cannot be generalized into a single distribution model. If it were a bimodal histogram, finding the threshold value could be relatively simple. Consequently, it is necessary to proceed adaptively to enhance contrast features. In the binarization stage, the threshold is chosen adaptively to each individual image: each image

is divided into two parts (upper part, averagely whiter and lower part, averagely darker) and the average of the two parts is considered as thresholding. In this way a value is obtained that is always greater than the value observed in the lower part thus eliminating the textural liver and noise information and increasing the data present close to Glissonian membrane. With this procedure the binary mask was obtained thanks to which it will allow us to extract the so-called Glissonian line (Figure 5.3).

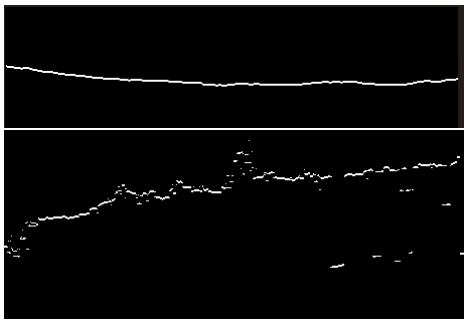


Figure 5.3: Binary mask extracted for F0-F1 staging (on top) and for F4 staging (on bottom)

Features The image features are extracted in the neighborhood of the detected line. By applying Sobel directional masks it was possible to obtain the gradient of the extracted Glissonian line and calculate its magnitude and phase. The idea is to verify if these features are capable of identifying different clusters, thus allowing anticipation of the classification (Figure 5.6, Figure 5.7, Figure 5.8).

From the extraction of lines, the reader can see that as the staging progresses, the continuity of the line is often broken, resulting in blocks of different lengths. Following this reasoning, it is possible to exploit the concept of connected-components that allows the identification of contiguous regions of pixels with the same characteristics, such as intensity or area, forming coherent objects or structures within the image. To find these regions it is necessary to label the components that are *connected*, which is determined by the application of graph theory. For the labeling procedure an array is used to store equivalence information between labels. This replaces the pointer-based rooted trees used to store the same equivalence information. It reduces the memory required and also produces consecutive final labels. Using an array instead of pointer-based rooted trees speeds up connected component labeling algorithms [67]. Considering 8-connected regions we identify the line pieces inside boxes and calculate their area. In the 2Dscatterplot (Figure 5.5) it's possible see the result of this operation.

Convolutional Neural Network A CNN with a simple and shallow ar-

chitecture is used here for classification, compared with more complex CNNs [56]. Experimental tests have shown that global features do not have any particular information relevance and, therefore, a very deep network is not needed to process the available data [54]. The network takes image spatial ROIs as input from the image dataset, performs a feature learning phase, characterized by two convolutional layers, ReLU, Max Pooling and Dropout, in which incoming image features are automatically learned and then classified. For dealing with imbalanced datasets a 10-fold Cross-Validation, was used to ensure fair and accurate model evaluation. In this way the model is trained and tested on a representative sample of each class, mitigating bias and improving overall performance. In this work, we chose to examine the behavior of the network for either low/moderate/circular. For classification, a sigmoid function was used in the former case, and a softmax in the latter.

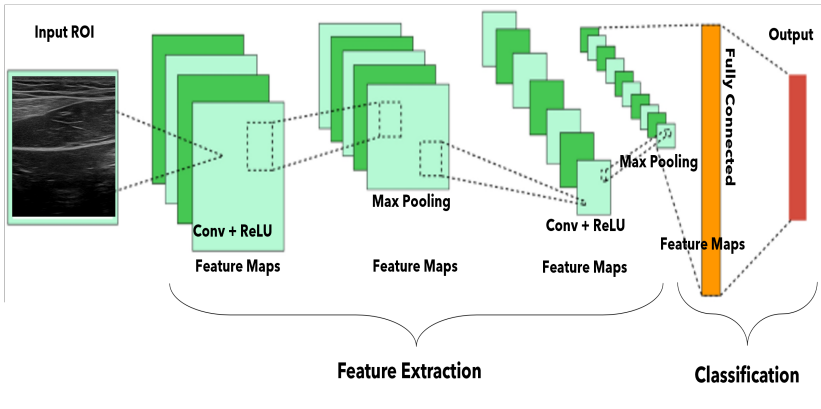


Figure 5.4: Architecture of proposed Convolutional Neural Network model.

The proposed model, depicted in Figure 5.4, was developed without employing any transfer learning processes or incorporating external information [51, 53]. During the training process, Cross Entropy was used as the loss function. The model was optimized using the Adaptive Moment Estimation (Adam) algorithm, with the learning rate set to 0.001. The computational environment consisted of Python 3.9, PyTorch 2.3.1, and CUDA 12.1. The hardware utilized for computation was an NVIDIA GeForce RTX 4060 GPU.

Principal Components Analysis Prior to classification using SVM, preprocessing of the data was performed with Principal Components Analysis (PCA) in order to extract the most informative features from large datasets while preserving the most relevant information from the initial dataset. PCA summarizes the information content of large data sets into a smaller set of uncorrelated variables, known as principal components. This reduces the com-

plexity of the model as the addition of each new feature negatively affects its performance, a phenomenon commonly known as the *curse of dimensionality*. By projecting a high-dimensionality dataset into a smaller feature space, PCA minimizes, or eliminates altogether, common problems such as multicollinearity and overfitting. Multicollinearity occurs when two or more independent variables are highly correlated, which can be problematic for causal modeling. Overfit models generalize poorly to new data, reducing their value.

Support Vector Machine (SVM) The characteristics of the images extracted near the line of interest aim to reliably quantify the diagnostic information relevant to the classification process. The fundamental assumption of this approach is that the extracted descriptors are able to capture significant structural variations and textures, potentially attributable to the different phases analyzed, thus allowing the identification of discriminating patterns useful for automatic classification.

Exploratory data analysis, supported by appropriate multidimensional visualizations, allowed us to examine the statistical distribution of features, identify any latent clusters, and assess the presence of correlations or dependencies between variables. In particular, the selected set of features, including the number of boxes identified, the magnitude of the gradient, and the associated phase, was organized into a structured dataset and stored in tabular format (Excel file), which was then used as input for training and validating a Support Vector Machine (SVM)-based classifier.

In order to evaluate the impact of the kernel function on the discriminatory power of the model, four different kernels were tested, allowing a comparative analysis of performance in terms of class separability (2, 3, and 4) and classifier generalization ability. For more information on the mathematical level, please refer to Paragraph 5.6.

5.4 Experimental results with CNN

5.4.1 CNN Without Initial Image Processing

In medical imaging, the Digital Imaging and Communications in Medicine (DICOM) protocol [206] is the standard for communication and management of medical imaging information; therefore, DICOM files are typically used. In this research, for convenience of image analysis, the original images were converted to 256-grayscale BMP files. First, the original images were marked by experienced clinicians and verified in clinical reality. Only selected ROIs around the Glissonian capsule line were considered.

Classification Results				
Class Type	Metrics	Training	Validation	Test
2	Loss	0.3254	0.3557	0.2901
	Accuracy	0.8950	0.9123	0.8602
	MAE	0.1203	0.1154	0.1956
3	Loss	0.4472	0.4677	0.3414
	Accuracy	0.8192	0.8333	0.5789
	MAE	0.1734	0.1644	1.1933

Table 5.1: Values of the CNN trained on the original ROI ultrasound scans for both binary and three-class classification scenarios: Loss, Overall Accuracy, Mean Absolute Error (MAE).

Experimental tests revealed that, without initial image processing, the network already yielded poor results for three-class classification and only moderate performance for two-class classification. In particular, for the three-class scenario, the test set metrics were notably low, suggesting an overfitting problem. Without the proposed image processing, the evaluation of the results is reported in Table 5.1. The CNN architecture used here exploits two convolutional layers (unlike the previous work [53] where 3 were used).

5.4.2 Focus-of-Attention Mechanism as a Guiding Tool

As summarized in Table 5.2, after the proposed histogram transformation the Overall Accuracy is at high levels for the validation set, and even better for the testing set.

Classification Results				
Class Type	Metrics	Training	Validation	Test
2	Loss	0.1254	0.1557	0.1901
	Accuracy	0.9214	0.9523	0.9402
	MAE	0.1003	0.1024	0.1736
3	MAE	0.1020	0.1181	0.2233
	Accuracy	0.9214	0.8873	0.701
	Loss	0.2166	0.2718	0.2232
4	MAE	0.1920	0.1611	0.3044
	Accuracy	0.7287	0.7857	0.4632
	Loss	0.6503	0.5557	1.2415

Table 5.2: Performance metrics of ROI-trained CNN transformations with Focus-of-Attention Mechanism for binary, three- and four-class classification scenarios: Loss, Overall Accuracy, Mean Absolute Error (MAE).

It can be seen from Table 5.3 that the convolutional network when aided by

image processing, can provide high performance in the area of liver classification. For comparison with the results in the literature, traditional ultrasound techniques (Section 2.1.3) for liver staging, despite their widespread clinical use, have significant limitations, particularly in the early stages of the disease, reducing their effectiveness in screening. Image analysis is heavily influenced by background noise, variable acquisition parameters, and operator dependence, resulting in low sensitivity and poor reproducibility. Only methods based on SWE report accuracy values ranging from 0.86 to 0.967. Moreover, most studies exploit a convex probe, whereas in this work, ultrasound scans were obtained by the linear probe.

For a comparison with the results in the literature, we refer to [207] for the classification of fibrosis stages $\geq$ F2, besides it can be seen that by locating an ROI within the medical image, without exploiting the deep networks [208, 209], nor of applying the transfer learning method [51, 210], the performance obtained in this work is comparable. Few works deal with image processing in support of neural networks [54, 53] yet even in this case the performance obtained can be considered more robust since the analysis did not perform any kind of point extraction of information.

Classification Results				
Class Type	**Metrics**	**Training**	**Validation**	**Test**
2-Class	Accuracy	0.9650	0.9523	0.9402
	MAE	0.1003	0.1024	0.1736
3-Class	Accuracy	0.9214	0.8873	0.7010
	MAE	0.1020	0.1181	0.2233

Table 5.3: Performance metrics of ROI-trained CNN transformations with Focus-of-Attention Mechanism for binary and three-class classification.

To evaluate the performance of the proposed method (Section 3.5) in comparison with deep models, we applied transfer learning to three alternative models, ResNet, VGG, and DenseNet, consisting of 50, 16, and 201 layers, respectively. As reported in Table 5.4, for binary classification, the proposed model achieved the highest recall (0.97), indicating superior sensitivity in identifying patients with the disease. Its overall accuracy (0.94) and precision (0.91) were very close to those of DenseNet, with the latter exhibiting a lower recall, thus reinforcing the reliability of our approach.

In the three-level classification of low, advanced, and cirrhotic stages, as reported in Table 5.5, the proposed model excelled again, achieving an overall accuracy of 0.701, a precision of 0.712, and a remarkable recall of 0.91, significantly outperforming the other models. On the other hand, VGG16 performed

Binary Classification (No Disease vs. Disease)			
Overall Accuracy	Precision (macro)	Recall (macro)	Model
0.80	0.75	0.78	ResNet
0.85	0.82	0.84	VGG16
0.99	**0.95**	0.92	DenseNet
0.94	0.91	**0.97**	Proposed Model

Table 5.4: Performance comparison of models for binary classification: values of overall accuracy, precision (macro), and recall (macro) obtained from transfer learning models classification with focus-of-attention mechanism for distinguishing between patients with and without disease. The best values are shown in bold.

slightly better in accuracy and precision. It exhibited a significantly lower recall, demonstrating that a higher number of layers is not always the optimal solution. This highlights the robustness of the proposed model in accurately classifying liver disease stages, especially in the case of small datasets, suggesting its potential as a valuable diagnostic support tool that merits further investigation.

Low/Advance/Cirrhotic Stage			
Overall Accuracy	Precision (macro)	Recall (macro)	Model
0.5894	0.6243	0.6408	ResNet
0.7263	**0.7238**	0.6663	VGG16
0.6526	0.6460	0.6818	DenseNet
0.701	0.712	**0.91**	Proposed Model

Table 5.5: Performance comparison of models: values of overall accuracy, precision (macro) and recall (macro) obtained from transfer learning models classification with focus-of-attention mechanism. The best values are shown in bold.

5.5 Graphical Representation of Features

As illustrated in Chapter 3, the use of these processing techniques aims to increase the information content of the scans and facilitate the detection of discriminative patterns.

The characteristics of the images were extracted in the vicinity of the line of interest, with the aim of quantifying diagnostic information useful for classification. In particular, by applying Sobel directional masks, it was possible to cal-

culate the gradient of the detected Glissonian line, obtaining both its amplitude and phase. Analysis of the extracted lines using the binary mask shows that, as the stage progresses, the continuity of the line tends to be progressively interrupted, generating fragments of different lengths. Considering the 8-connected regions, it was possible to identify the individual residual segments within the boxes and calculate their area (Figure 5.5). These measurements complete the set of available features and complement those already derived from the gradients. The hypothesis underlying this approach is that these parameters can reveal distinctive patterns between the different stages, thus allowing us to anticipate an automatic classification process (see Figure 5.6, Figure 5.7, Figure 5.8).

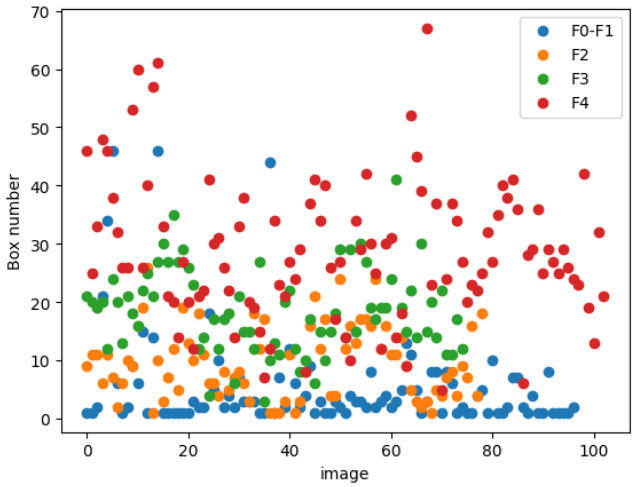


Figure 5.5: Two-dimensional representation obtained by calculating the area of the individual residual segments within the boxes for all stages.

The graphical representation of data, using tools such as scatter plots and box plots, is a fundamental step after feature extraction. These visualizations allow us to explore the distribution of features, identify any hidden patterns, and verify the presence of relationships between variables. Compared to numerical measures alone, visual representation allows for more immediate identification of outliers, natural groupings (clusters), and possible separations between classes. In addition, graphs provide essential interpretive support for evaluating the effectiveness of selected features for future automatic classification. In this context, various combinations of features were tested in order to identify those most effective in generating distinguishable clusters. The aim was to understand which variables, considered individually or in pairs, were best able to discriminate between the different stages. For example, by combining the

gradient phase with the box area, a two-dimensional scatter plot was obtained in Figure 5.6) showing how samples belonging to stage F0 tend to concentrate near the origin, reflecting reduced values for both the angle and the number of boxes. Stage F2, on the other hand, has a less defined distribution and is less distinguishable than the others. Stages F3 and F4, although both characterized by a high number of connected regions (boxes), differ in their correlation with the angle, showing two separate groupings. This allows them to be identified as two distinct clusters, suggesting the possibility of automatic classification based on these patterns.

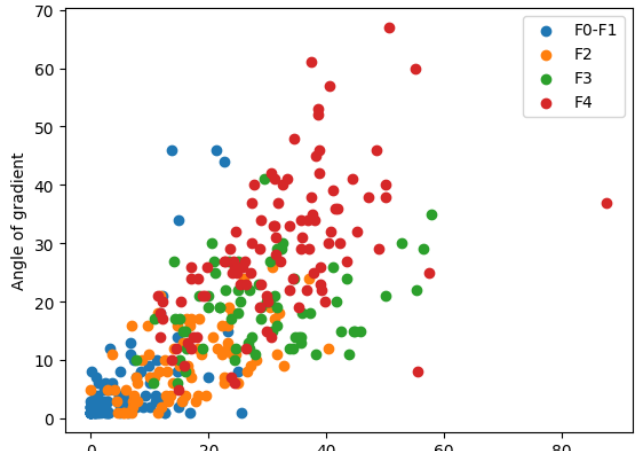


Figure 5.6: ıg both the anks to the

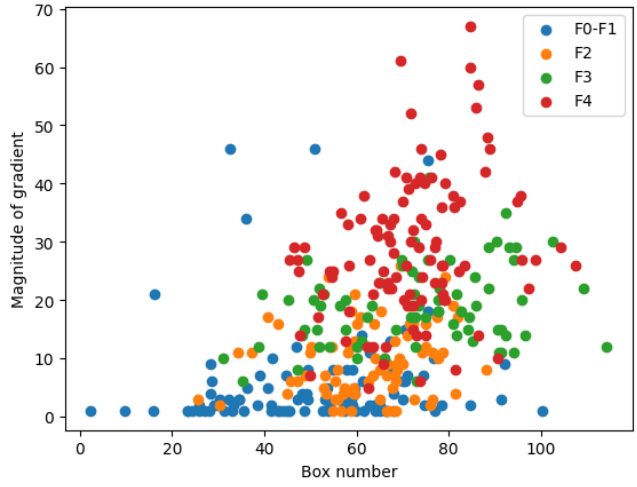


Figure 5.7: Two-dimensional representation obtained by considering both the magnitude of the Glissonian and the box area identified thanks to the connected components.

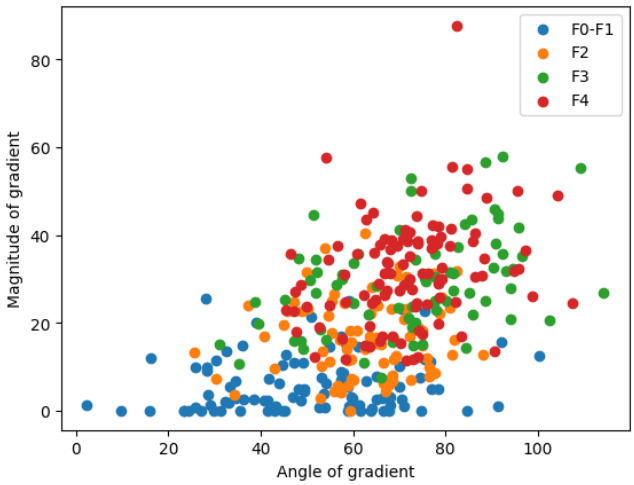


Figure 5.8: Two-dimensional representation obtained by considering both the phase and magnitude of the Glissonian line for the different stages of fibrosis.

Instead, observing the graphs obtained by combining amplitude and phase (Figure 5.8) and amplitude and box area (Figure 5.7), the distinctions between the different stages are less clear. For further clarification, a three-dimensional scatterplot is also shown in which all three variables are considered (Figure 5.9). From the various visualizations it is clear that the main problem in staging is in identifying the F2 staging, that is, the average case of the pathology.

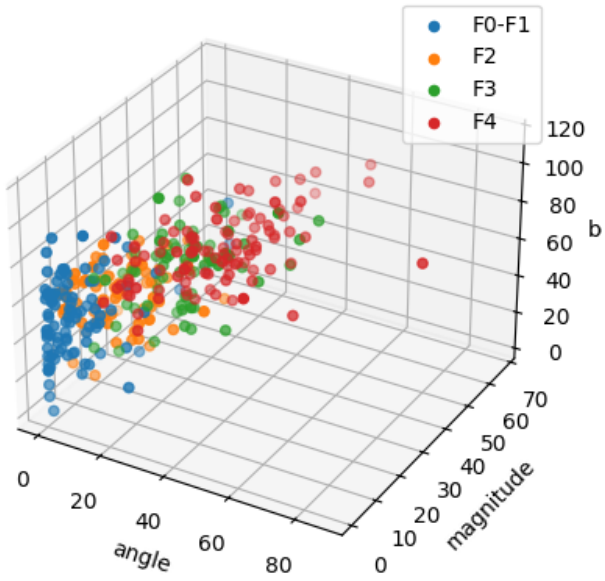


Figure 5.9: Three-dimensional representation obtained by considering the phase (x-axis), magnitude (y-axis) and the number of box (z-axis) of the Glissonian line for the different stages of fibrosis.

In order to study their trend and assess the critical values which define separation criteria among the various subgroups proposed, also the quartile method was used. The box plot in Figure 5.10 shows the trend of Magnitude, in the different stages of progression (F0–F1, F2, F3, F4). In the early stages (F0–F1), a very wide distribution is observed, with values ranging from close to zero to over 100.

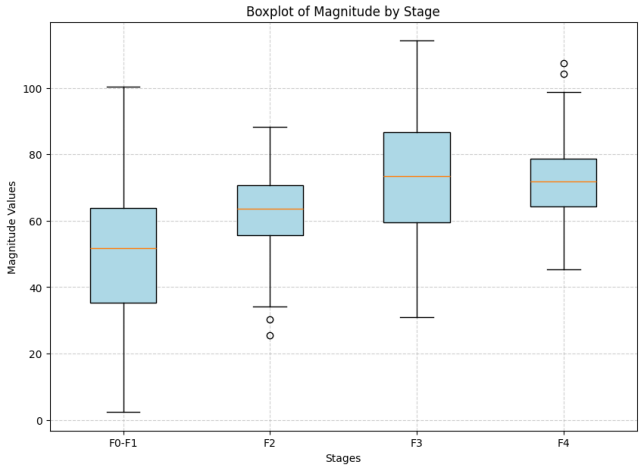


Figure 5.10: Box-plot of magnitude gradient calculated on the Glissonian line.

The median is relatively low, indicating that in the initial samples, Magnitude tends to take on low values, albeit with strong internal variability. In stage F2, the values are more concentrated: the median increases compared to F0–F1 and the dispersion decreases, although there are some outliers with particularly low values. With the transition to stage F3, there is a further increase: the median is among the highest and the distribution extends to the absolute maximum values, exceeding 110. This stage is also characterized by greater variability, a sign of marked differences between samples. Finally, in stage F4, the Magnitude remains at high values, with a more compact distribution than in F3 and few outliers, suggesting a tendency to stabilize.

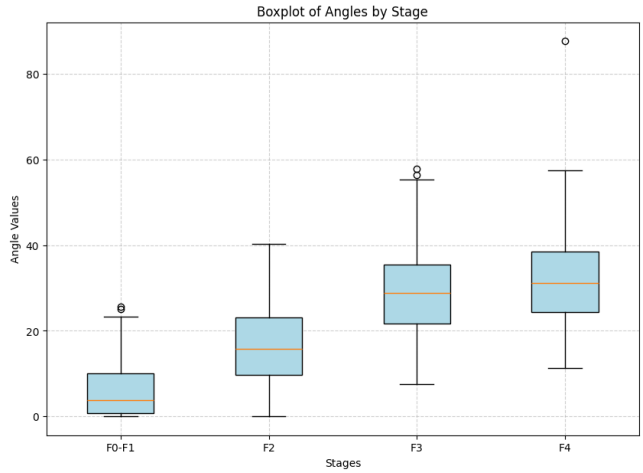


Figure 5.11: Box-plot of angle gradient calculated on the Glissonian line.

The box plot in the Figure 5.11 shows the distribution of Angle values in the different stages (F0–F1, F2, F3, F4). In the early stages (F0–F1), the values are very low and concentrated close to zero: the median is minimal and the dispersion extremely low, with few outliers just above 20 degrees. This indicates that, in the initial stages, the lines detected tend to have an almost zero slope, showing a certain regularity. In stage F2, there is an increase in both the median and the variability: the values begin to spread over a wider range, with a maximum exceeding 40 degrees. In F3, this trend is further accentuated: the median increases again, the interquartile range widens, and the dispersion becomes more evident, with values reaching close to 60 degrees. Finally, stage F4 shows the greatest overall variability: the median remains high, but the distribution extends to very high values (over 80 degrees), a sign that in the advanced stages the orientation of the structures becomes more irregular and uneven. Overall, the graph shows an increasing trend in angular values as the stages progress, with a transition from low and stable values in the early stages to high and variable values in the later stages. This trend suggests that the angle of the gradient may be a sensitive indicator of the loss of regularity and continuity of the line as the staging progresses.

Finally, we have the box plot of the areas (Figure 5.12). In the initial stages (F0–F1), the values are very low: the median is around $3 - 4$, and most of the observations are concentrated in the lower range, although there are numerous outliers that reach higher values, indicating occasional more marked interruptions in the line. With the transition to stage F2, the median increases significantly (to around 10) and variability increases: the values are distributed over a wider range, a sign that linear continuity is beginning to fragment more clearly. Stage F3 shows further growth, with a median close to 20 and a more compact

distribution than F2, but extending to values of around $35 - 40$, indicating a progressive increase in discontinuities. Finally, stage F4 shows the greatest overall variability: the median reaches the highest values (around $25 - 30$), while the dispersion increases significantly, with boxes reaching over 60. This trend reflects a very marked fragmentation of the lines, with segments of heterogeneous length and increasingly evident discontinuities. There is therefore a growing trend in the number and extent of areas of discontinuity (Box) as the stages progress. This behavior suggests that the Box measurement is a robust indicator of the loss of continuity of the Glissonian line, clearly distinguishing the advanced stages from the initial ones.

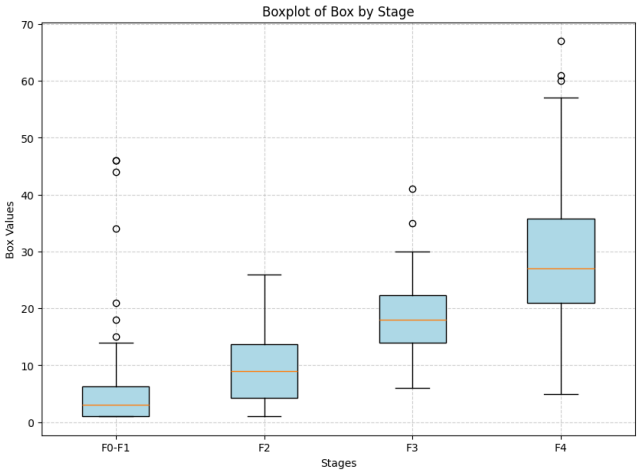


Figure 5.12: Box-plot of box area calculated on the Glissonian line.

5.6 Experimental results with SVM

Based on the statistical analysis of these parameters, application of a SVM classifier to divide the population into four (Figure 5.13), three (Figure 5.14) and two (Figure 5.15) classes.

The idea was to compare different SVM classifiers, trained on feature data extracted from an Excel file. In the first phase, the training data is loaded. Each sheet represents a type of feature—specifically angle, magnitude, and box—while the columns (A, B, C, D) correspond to four different classes of samples. Although there is no high dimension to improve the readability of the data, PCA was applied, which projects the original three-dimensional space (angle, magnitude, box) into two new principal components (PC1 and PC2). This approach does not select a single variable, but produces linear combinations of the original features, maximizing the explained variance. The combination

of features in two-dimensional representations using scatter plots allows us to verify the possible formation of distinct clusters, potentially associated with different pathological stages. The training phase involves the construction of four distinct SVM models, each characterized by a different kernel:

- SVC with linear kernel, which seeks a linear separation hyperplane between classes.
- LinearSVC, a variant optimized for large linear problems.
- SVC with RBF (Radial Basis Function) kernel, which allows modeling nonlinear relationships between features.
- SVC with Polynomial (Degree 3) kernel, capable of capturing complex structures through polynomial transformations of the feature space.

Finally, each model is evaluated and represented graphically. In particular, the decision boundaries learned by the classifiers are plotted in the two-dimensional space obtained with PCA. The results are organized in a grid of subgraphs 2x2, allowing a direct visual comparison between the different separation strategies during the training phase. This approach allows not only for quantitative performance evaluation (e.g., in terms of accuracy and confusion matrices), but also for qualitative observation of behavior in the reduced feature space, identifying which kernel best fits the data distribution.

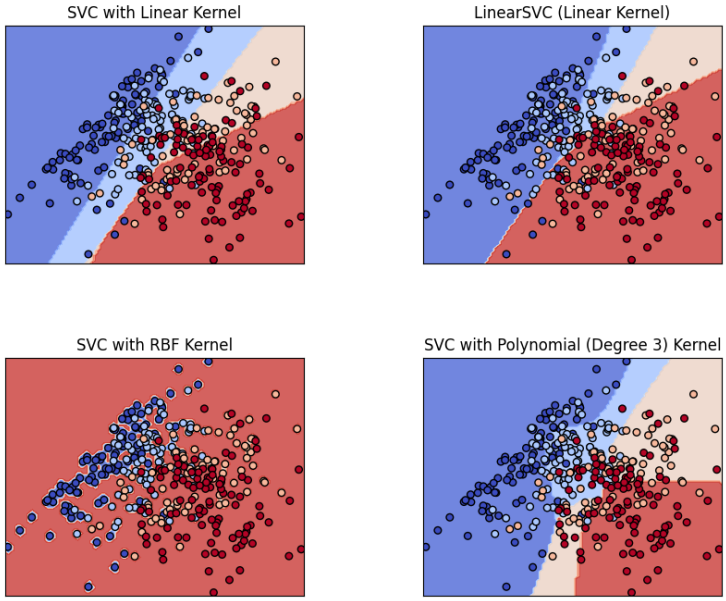


Figure 5.13: Training phase: four different kernel of SVM were compared to understand which one is the best for performing the classification of **four** different liver staging. The Gaussian kernel (RBF) provides the worst performance

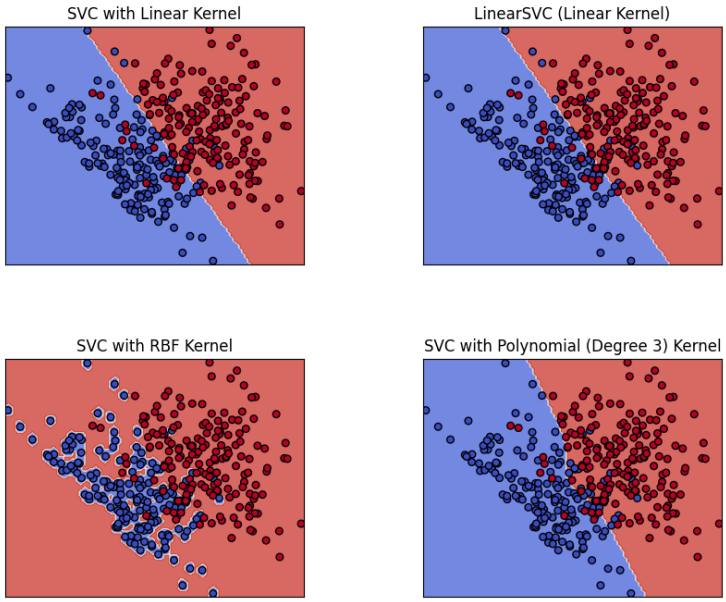


Figure 5.15: Training phase: four different kernel of SVM were compared to understand which one is the best for performing the classification of **two** different liver staging. The Gaussian kernel (RBF) provides the worst performance

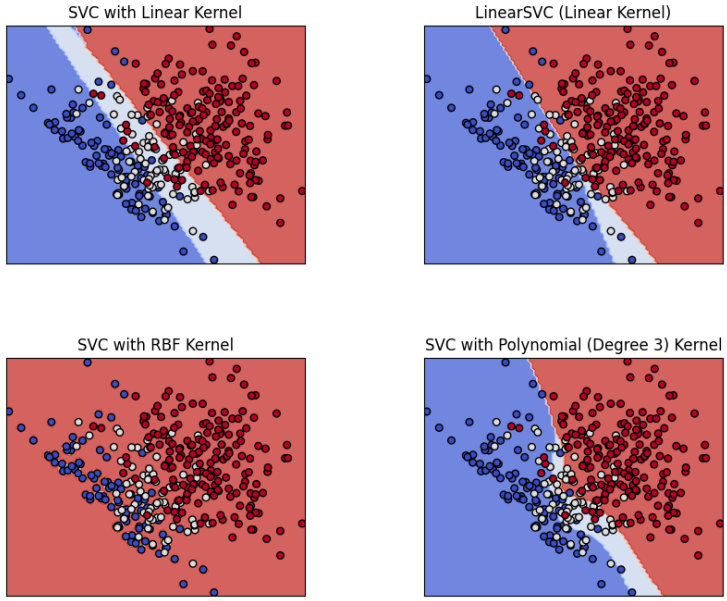


Figure 5.14: Training phase: four different kernel of SVM were compared to understand which one is the best for performing the classification of **three** different liver staging. The Gaussian kernel (RBF) provides the worst performance

The linear models *LinearSVC* and *SVC* (with *linear* kernel) yield slightly different decision boundaries. This can be a consequence of the following differences:

- *LinearSVC* minimizes the squared hinge loss while *SVC* minimizes the regular hinge loss;
- *LinearSVC* uses the One-vs-All (also known as One-vs-Rest) multiclass reduction while *SVC* uses the One-vs-One multiclass reduction.

Both linear models have linear decision boundaries (intersecting hyperplanes) while the non-linear kernel models (polynomial or Gaussian RBF) have more flexible non-linear decision boundaries with shapes that depend on the kind of kernel and its parameters.

The graphs presented previously, relating to the training phase of the four SVM models obtained through dimensional reduction with PCA, highlight an intrinsic limitation linked to representation: the two-dimensional projection, although useful for illustrative purposes, is not able to render faithfully the complexity of the decision boundaries in the feature space. In fact, while the

human eye struggles to interpret separations in three-dimensional spaces, model training is still conducted using the entire original feature set — angle, magnitude, and box area — so as to allow the SVM to learn a decision boundary in three-dimensional space. For this reason, the $3D$ graphic representations will be directly reported, relating to both the training and testing phases, considering both binary (Figure 5.20, 5.21) and multiclass (Figure 5.16, 5.18, 5.17, 5.19) classification scenarios, without further resorting to dimensional reduction via PCA. In these graphs, each sample is projected into a $3D$ scatter plot, whose coordinates correspond to the original features (Angle, Magnitude, BoxArea), while the color of the dots represents the class predicted by the model. The views are organized in a 2x2 grid, so as to allow a direct comparison between the four SVM approaches considered (linear, LinearSVC, RBF and polynomial).

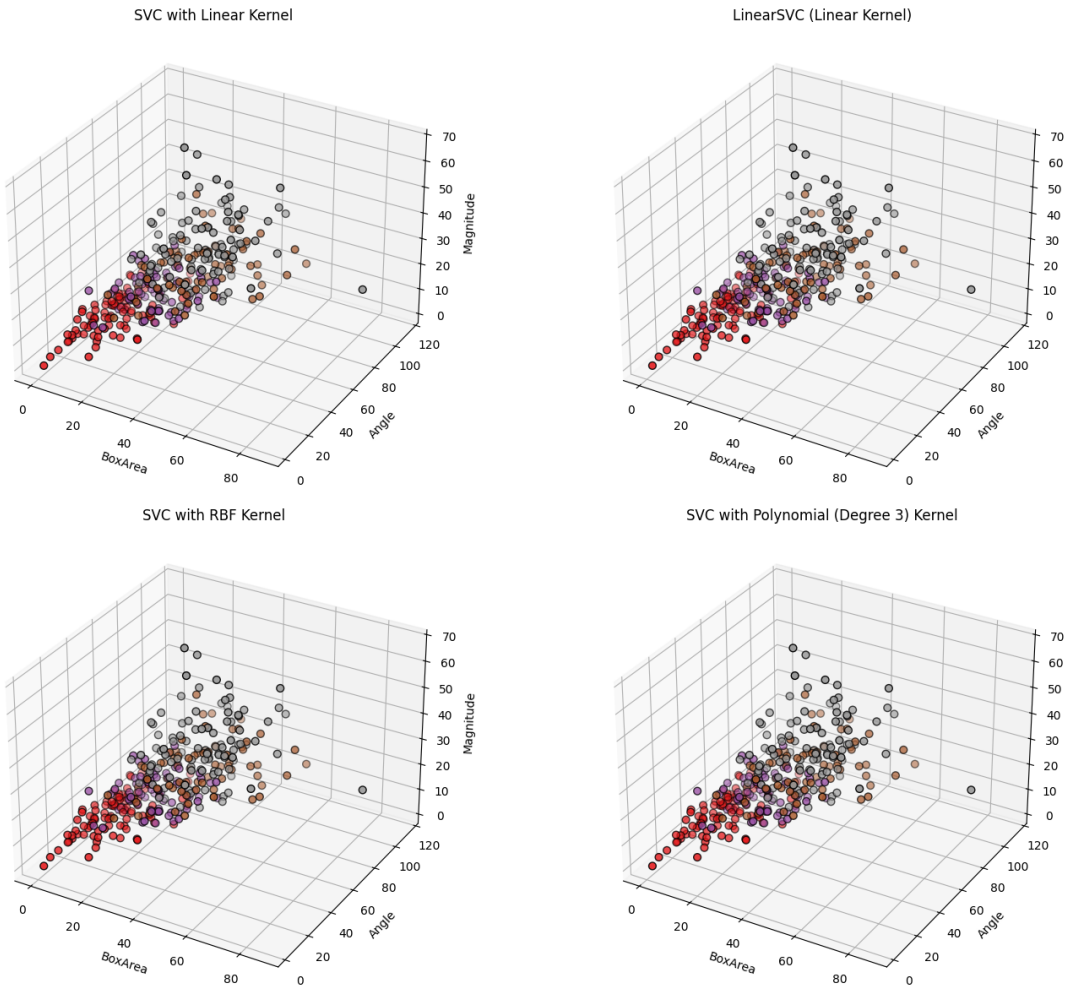


Figure 5.16: Training phase: Three-dimensional visualization of four SVM approches for 4-class classification without PCA approch.

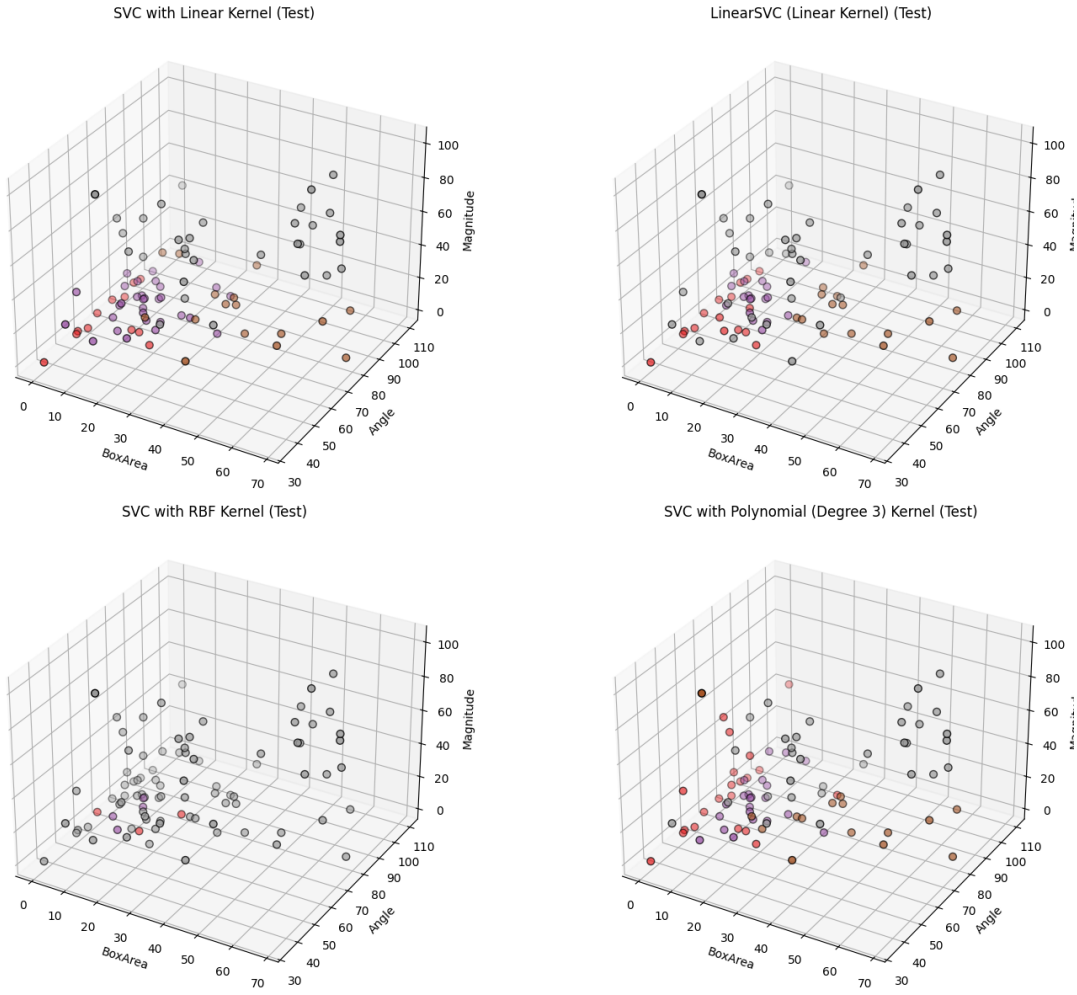


Figure 5.17: Testing phase: Three-dimensional visualization of four SVM approches for 4-class classification without PCA approch.

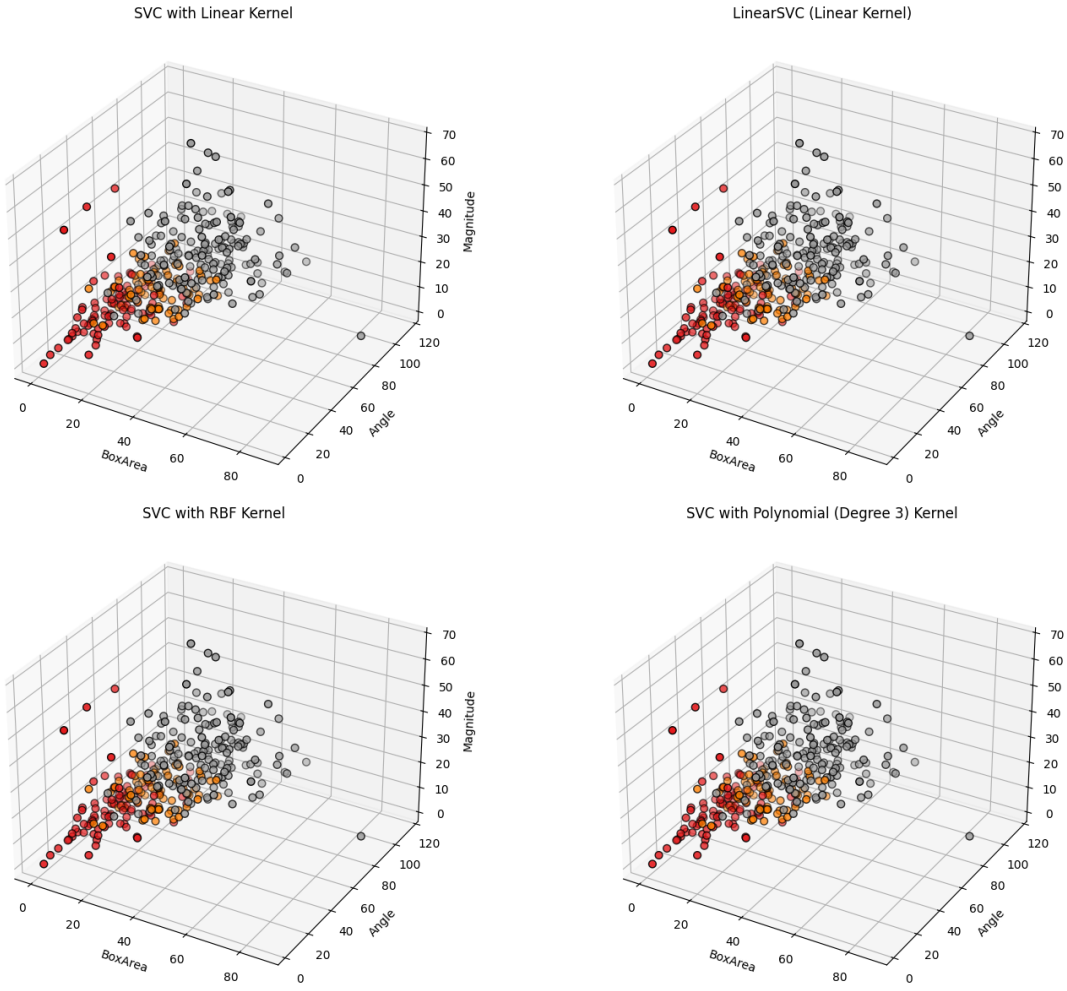


Figure 5.18: Training phase: Three-dimensional visualization of four SVM approches for 3-class classification without PCA approch.

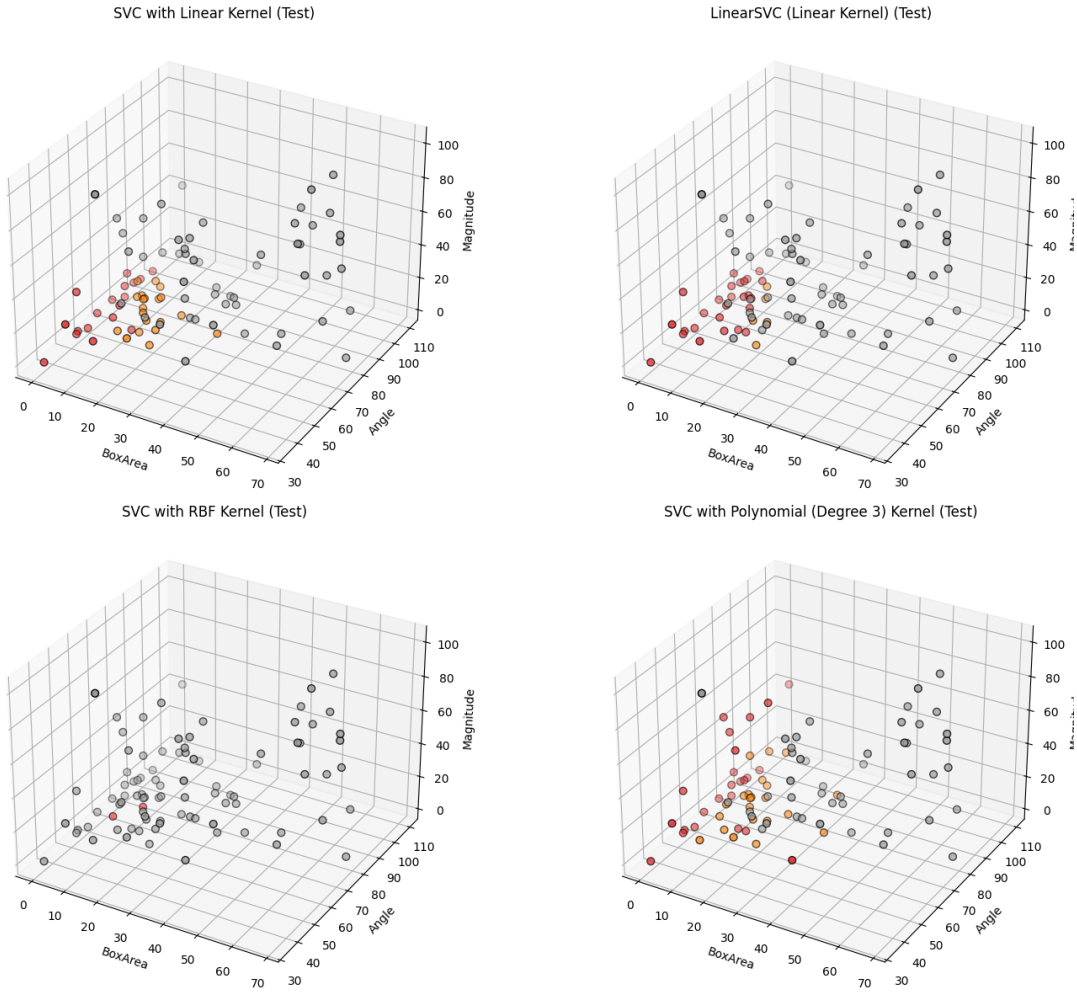


Figure 5.19: Testing phase: Three-dimensional visualization of four SVM approches for 3-class classification without PCA approch.

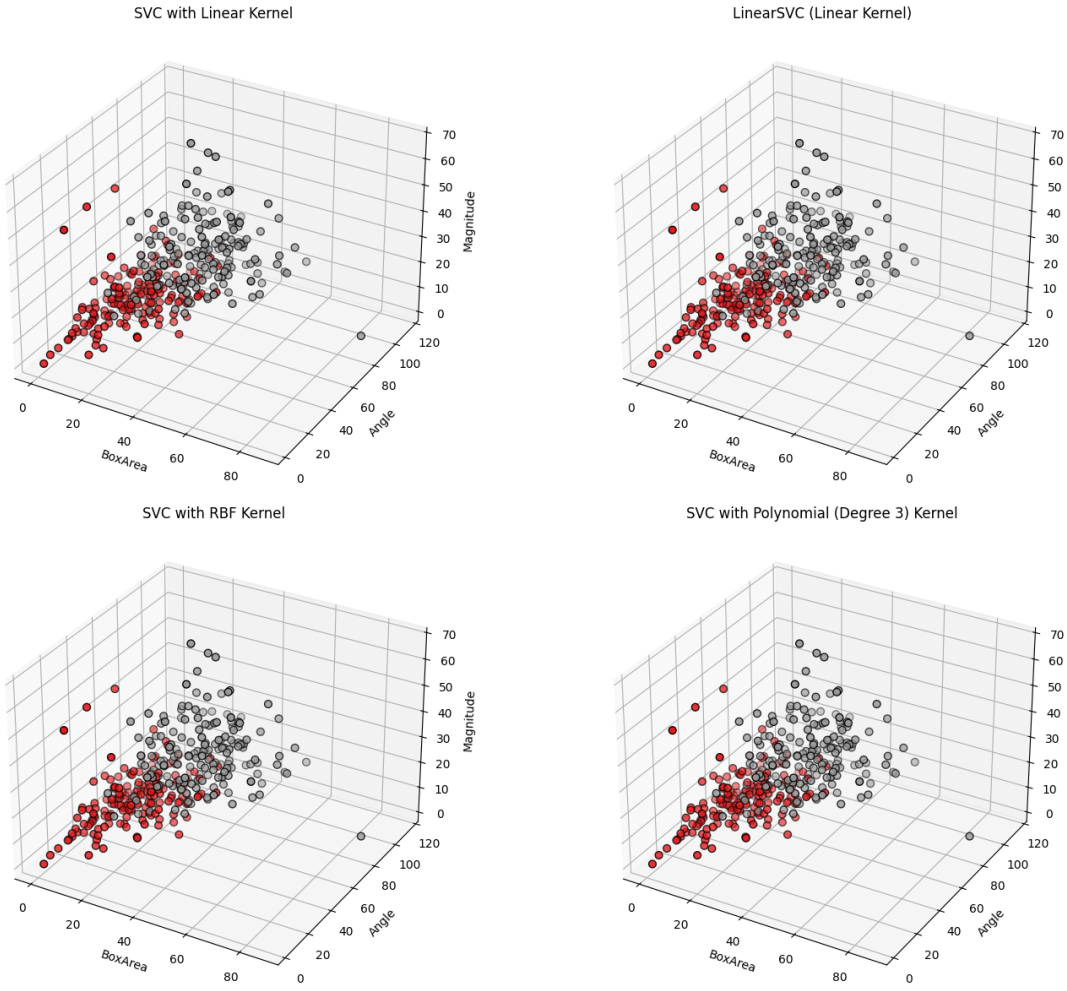


Figure 5.20: Training phase: Three-dimensional visualization of four SVM approches for binary classification without PCA approch.

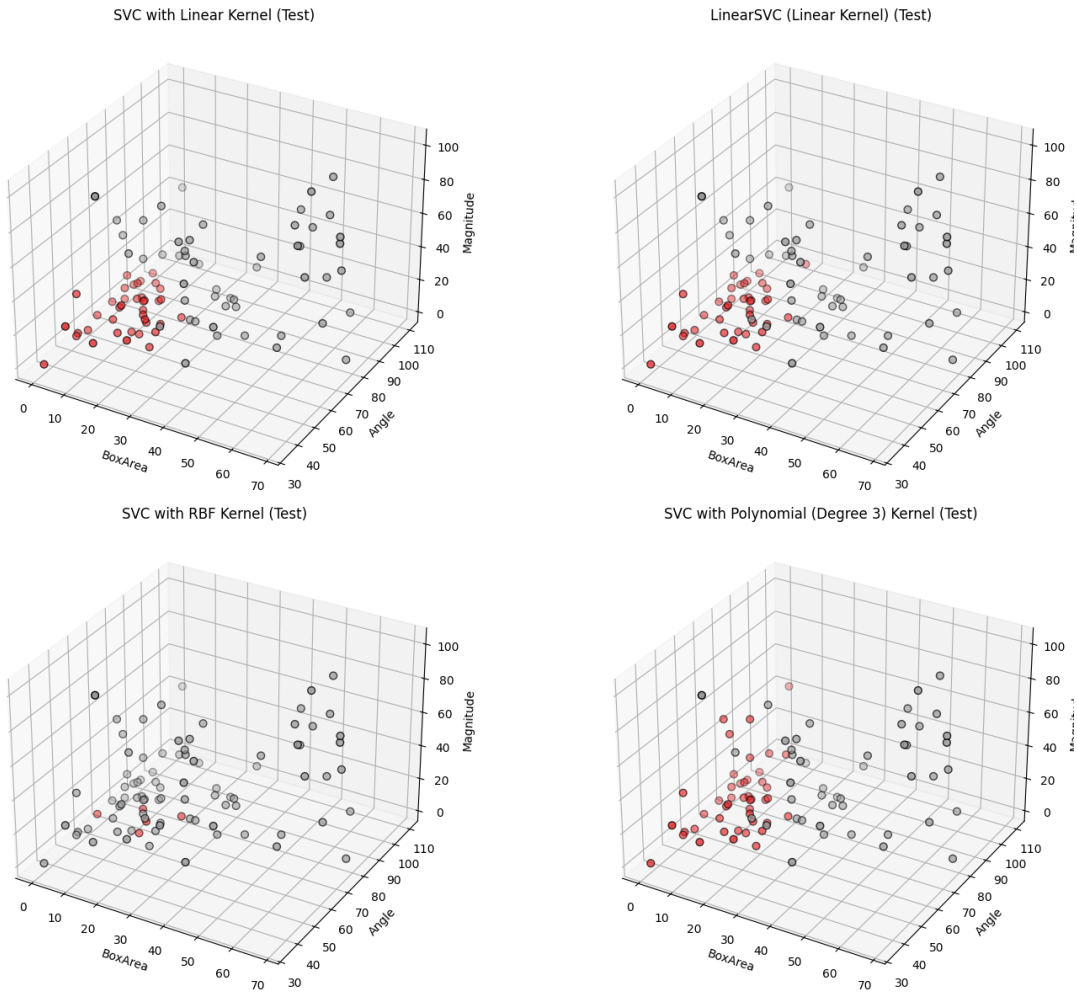


Figure 5.21: Testing phase: Three-dimensional visualization of four SVM approches for binary classification without PCA approch.

Even in the absence of size reduction via PCA, the procedure adopted for the training phase remains unchanged compared to what was previously described. Before introducing the analysis of evaluation metrics, it is useful to recall the main stages of the testing process. The testing phase aims to evaluate previously trained SVM models, using an independent dataset, also organized in Excel format. The testing phase is crucial in machine learning, as it allows us to estimate the actual predictive power of the models and identify which SVM kernel offers the best performance on previously unseen data.

For each classifier, the accuracy is calculated. From the comparison of the accuracies obtained in the training and testing phases for the four-class classification (Figure 5.22), it emerges that the SVC with linear kernel shows a training accuracy of 0.71 and a testing accuracy of 0.45, indicating moderate generalization ability and a small gap between the training and testing phases.

	4-classes	Training accuracy	Testing accuracy
Model	SVC with linear kernel	0.71	0.45
	LinearSVC	0.65	0.34
	SVC with RBF	0.99	0.21
	SVC with Polynomial (Degree 3) kernel	0.68	0.45

Figure 5.22: The table shows the classification performance obtained for a 4-class problem, comparing different SVM-based models with different kernels in terms of training accuracy and testing accuracy.

The LinearSVC model performs less well, with an accuracy of 0.65 in training and 0.34 in testing, suggesting less effectiveness in separating classes in the feature space considered. The SVC with RBF kernel achieves very high training accuracy (0.99), but with a drastic reduction in performance during testing (0.21). This behavior is indicative of marked overfitting, in which the model is strongly adapted to the training data but unable to generalize on unseen data. Finally, the SVC with a degree 3 polynomial kernel shows a training accuracy of 0.68 and a testing accuracy of 0.45, comparable to that of the linear kernel, but with a lower ability to adapt to the training data. Overall, the analysis shows that more complex models (e.g., the RBF kernel) do not necessarily guarantee better performance in terms of generalization, especially when there are a limited number of features or a well-defined nonlinear separability between classes.

The same was done for the 3-class classification (Figure 5.23) and binary (Figure 5.24) classification. A three-dimensional graphical representation for training and testing phase is visible in Figures 5.18 and 5.19 while the comparative accuracy tables are shown in Figure 5.23.

	3-classes	Training accuracy	Testing accuracy
Model	SVC with linear kernel	0.80	0.48
	LinearSVC	0.74	0.41
	SVC with RBF	0.99	0.31
	SVC with Polynomial (Degree 3) kernel	0.79	0.51

Figure 5.23: The table shows the classification performance obtained for a 3-class problem, comparing different SVM-based models with different kernels in terms of training accuracy and testing accuracy.

	2-classes	Training accuracy	Testing accuracy
Model	SVC with linear kernel	0.90	0.26
	LinearSVC	0.90	0.27
	SVC with RBF	1.00	0.38
	SVC with Polynomial (Degree 3) kernel	0.90	0.21

Figure 5.24: The table shows the classification performance obtained for a binary class problem, comparing different SVM-based models with different kernels in terms of training accuracy and testing accuracy.

In light of the obtained results, the linear-kernel SVM model outperforms the other considered variants, suggesting that, within the analyzed context, a lower-complexity model may be more effective. At first glance, this outcome may appear contradictory given the strong performance also achieved by the CNN, which represents a substantially more complex model. However, this apparent discrepancy can be explained by considering the fundamentally different ways in which SVMs and CNNs address the image classification problem.

SVMs operate on a set of previously extracted numerical features, and their discriminative capability largely depends on the quality and representativeness of these manually designed features. When such features are highly informative and well aligned with the classification task, a linear decision boundary may be sufficient to effectively separate the classes, while simultaneously reducing the

risk of overfitting typically associated with more complex models. Moreover, the lack of hierarchical feature learning makes SVMs particularly well suited to scenarios in which the dataset size or the number of available training samples is limited. Conversely, CNNs adopt an end-to-end learning paradigm, in which hierarchical feature representations are learned directly from the raw input data. This allows CNNs to model complex spatial patterns and high-level semantic structures, but typically requires a substantially larger amount of training data to fully exploit their representational capacity. In the presence of relatively small or moderately sized datasets, the high model complexity of CNNs may not yield a clear performance advantage and can, in some cases, be affected by overfitting or suboptimal generalization.

Therefore, the observed results are not contradictory but rather reflect the trade-off between model complexity, feature representation, and dataset size. In particular, while linear SVMs can effectively leverage well-engineered features in low- to medium-data regimes, CNNs tend to demonstrate increasing advantages as the size of the dataset and the intrinsic complexity of the classification task grow.

Chapter 6

Application of tele-rehabilitation techniques

Within the overall scope of the STORMS project, an Internet of Medical Things (IoMT) system based on the use of ReMoVES has been specifically designed for multiple sclerosis rehabilitation.

This section presents the preliminary analysis of the results from the STORMS project, conducted in collaboration with San Martino Hospital.

Two exergames, specifically developed to meet the needs identified during the requirements analysis phase of the STORMS project, are introduced. The section details the data collection process, providing definitions for the recorded fields within the log files and describing the structure of both games, with particular attention to their levels and functionalities. To minimize patient distraction and avoid interference from irrelevant background elements, both exergames were designed using simple $2D$ graphical components.

6.1 Activities for Occupational Rehabilitation

In the field of tele-rehabilitation, to better analyze a patient's performance during motor activity, obtained through biomedical instrumentation and/or digital technologies, it is necessary to process and evaluate the signals extracted from the various sensors employed.

Based on Microsoft Kinect technology, ReMoVES Patient Client performs calibration on the individual user in order to provide *tailored* results and allows the therapist to monitor tests and activities through a Therapist Client that can be consulted from different technologies (PCs, phones, tablets). On the server side, all activity-related data and signals are processed using adaptive approaches, enabling performance quality assessment, detection of the pattern of complex activities actually performed, and detection of any compensating movements. The extraction of significant data is performed only after suitable nonlinear filters and adaptive segmentation which allow to extract statistics and key indicators from signal fragments characterized by stationarity properties and related to homogeneous measurement states. According to the ISO/IEC/IEEE 12207 standard, the development of the ReMoVES IoT system followed the requirements engineering process consisting in defining the scope of the project, the IoT system and its requirements [211], based on the defined stakeholders and to the final destination user group.

This doctoral thesis presents ReMoVES, an IoMT system that provides telemedicine services in the context of multiple sclerosis rehabilitation, as part of the STORMS project. The STORMS initiative, as the genesis of digital tele-rehabilitation solutions, aims to enhance the quality of life for multiple sclerosis patients. Offers specific exergames that encourage high repetition, provide guidance to the patient, and can be adapted to the needs of the individual rehabilitation plan [212]. Through exergames, the automatic synchronization of the acquired signals is made possible while fully respecting the patient's time. Also, its interface is very simple and intuitive even for people with impairments, the delivery of the exercises is personalized and follows the individual activity plan defined by the therapist in agreement with a multidimensional team. The playful component of the exergame will help to increase motivation and adherence to rehabilitation. At the same time, the system will favor the practice of task-oriented exercises and muscle strengthening, favoring the improvement of the patient's functionality. To take into account the wide variability of patients' conditions and disease progression, the most appropriate exercises can be assigned and the complexity of the required task can be adapted by defining a range of difficulty levels. Thanks to the collection, integration and remote analysis of patient signals and data, the solution allows continuous monitoring of the activities on which the therapist constantly updates the personalized exercise plan.

6.2 Inclusion and Exclusion Criteria

Participant recruitment was based on the following inclusion criteria:

- Confirmed diagnosis of Multiple Sclerosis (MS).
- Age between 18 and 60 years.
- EDSS score ≤ 7.5.

Additionally, participants were required to obtain scores below normative thresholds (5th percentile, adjusted for age, sex, and education) in at least two neuropsychological tests.

Exclusion criteria included:

- Presence of severe mood disorders.
- Recent steroid therapy (within the 2 months preceding the assessment).
- Inability to maintain adequate visual fixation (e.g., due to nystagmus).
- Presence of post-chiasmatic perimetric visual field defects.
- Diagnosis of photosensitive epilepsy.

- Insufficient compliance or motivation to adhere to the treatment protocol.

The neuropsychological assessment battery included standardized tests frequently employed in cognitive evaluations of individuals with disabilities. Specifically, the Brief International Cognitive Assessment for MS (BICAMS) and the Paced Auditory Serial Addition Task (PASAT) with 3s and 2s intervals were administered. The BICAMS battery comprises the Symbol Digit Modalities Test (SDMT), the California Verbal Learning Test—Second Edition (CVLT-II), and the Revised Brief Visuospatial Memory Test (BVMT-R). Recruitment eligibility required participants to score below the $5th$ percentile in at least two of these assessments.

Written informed consent was obtained from all participants prior to study initiation.

Cognitive measures were collected at three time points: baseline ($T0$), after completing 10 exergame sessions ($T1$), and one month following the end of the intervention ($T2$).

In addition to these scheduled neuropsychological assessments, daily monitoring of motor performance was made possible through parameters automatically recorded during exergame sessions. This continuous data acquisition enabled a dynamic and detailed evaluation of each patient's motor progress over time. Moreover, indirect measures of effort and perceived fatigue were derived from daily tracking of relevant parameters, such as the number of games played, repetitions performed, movement speed and trajectory, game scores, postural control, joint angles, range of motion, and more. Comparing the exercises executed with those prescribed provided valuable insights into the patient's engagement levels and the appropriateness of the proposed exercise regimen. This approach allowed for a more individualized and adaptive rehabilitation pathway, ensuring that the program was effectively aligned with each patient's functional abilities and therapeutic needs.

6.3 Study Design

The present pilot study was designed as a small-scale investigation with the primary objective of evaluating the feasibility, operational logistics, and potential methodological challenges associated with the forthcoming clinical trial, rather than assessing the efficacy of the proposed intervention [213]. As shown in Figure 6.1, a cohort of individuals with Multiple Sclerosis (pwMS) was recruited at baseline ($T0$) at San Martino Hospital, yielding a final sample of twenty participants. This sample size was determined to satisfy the statistical requirement for estimating the recruitment rate with a 90% confidence level and a precision of $\pm 15\%$, assuming an anticipated recruitment rate of approximately 20%.

This step provides a mixed-gender group of adults who vary in age, educational background, and the length of time they have had the disease. We have

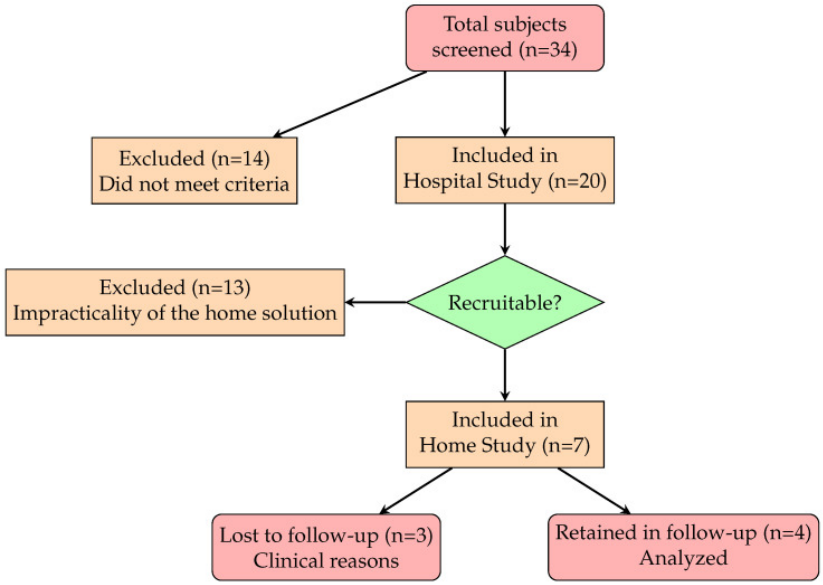


Figure 6.1: Flow diagram of the study.

20 adults (11 males and 9 females) with the following characteristics:

- Age range: $35-69$ years, mean: 53.8, standard deviation: 7.8,
- Education range: $8-18$ years, mean: 14.6, standard deviation: 3.1,
- Disease duration range: $8-20$ years, mean: 11.6, standard deviation: 3.4,
- Expanded Disability Status Scale (EDSS) range: $4.5-7$, mean: 5.8, standard deviation: 1.0.

Following a series of supervised training sessions conducted within neurorehabilitation clinics (Hospital Study), seven patients were selected to transition to a telerehabilitation program delivered at home under remote clinical supervision. Exclusion from the home-based intervention was applied in cases of limited technological literacy, despite the relatively young age of the cohort, or in the absence of a competent caregiver available to support the patient at home.

Of these, four patients successfully completed the home-based program for a minimum duration of one month, up to the $T2$ follow-up assessment. Three participants discontinued the intervention prematurely due to concomitant clinical complications, yielding a retention rate of 57%.

In the initial phase of the home study, no structured exercise prescription was provided. In the absence of specific recommendations, participants were

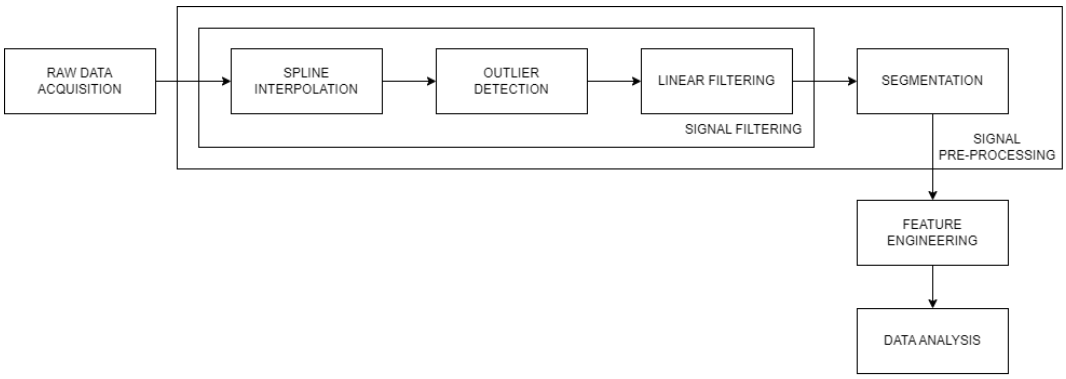


Figure 6.2: Flow chart of the proposed approach.

free to select activities based on personal preference. This approach resulted in a marked tendency to favor simpler and more entertaining exercises, often to the detriment of tasks with greater motor or cognitive demands. Consequently, patients exhibited proficiency in basic exergames while showing limited progression toward more challenging and functionally meaningful activities.

In light of these observations, the study protocol for subsequent home-based interventions was revised to incorporate individualized, structured treatment plans. A personalized exercise schedule was established for each patient, prescribing specific activities to be performed from Monday to Thursday, thereby promoting a balanced engagement with both basic and progressively demanding tasks.

6.4 Working Method

The signal and data processing workflow adopted in this thesis is outlined in Figure 6.2. Raw data is first acquired from the ReMoVES system, after which the signals undergo filtering through various techniques designed to address the potential issues typically associated with data captured by the Kinect sensor.

Subsequently, depending on the specific exergame under consideration, the signals are segmented into smaller sections corresponding to individual primitive movements. Following segmentation, a set of features is extracted from each segment, which are then analyzed individually.

In particular, for angular features, this work adopts an approach similar to that proposed by the authors in [214], considering the relative angles of body joints. These angles are computed using Equations 4.31 and 4.27, which provide the basis for analyzing joint kinematics within each movement segment.

6.5 Motor/Cognitive Exergames

As detailed in [114], a series of new cognitive exergames were developed within the framework of the STORMS project, with the objective of addressing several of the most prevalent symptoms associated with Multiple Sclerosis, including:

- Coordination impairments;
- Balance disturbances and dizziness;
- Visual disturbances, including potential deficits in color discrimination;
- Cognitive dysfunctions involving memory and learning difficulties;
- Reduced concentration capacity;
- Attention deficits and impaired computational abilities;
- Difficulties in executing tasks of higher cognitive or motor complexity;
- Impaired environmental perception.

All exergames are based on voluntary movements of the limbs. During gameplay, patients are required to coordinate their movements and execute corrective reactions as necessary. Adequate motor and visuomotor coordination is essential for the correct performance of the proposed tasks. Additionally, both direct and indirect observations of symptoms such as numbness in body regions and/or extremities, as well as spasticity affecting movement quality, can be obtained, particularly when the activities involve trunk and limb movements.

Patients interact with the system via body movements captured by the Microsoft Kinect sensor. Specifically, three types of movements are monitored: shoulder flexion–extension, shoulder abduction–adduction, and bilateral trunk inclinations.

An initial calibration procedure is performed for each participant to adapt the game parameters to individual motor abilities. Real-time adjustments are also implemented to optimize task difficulty and responsiveness during the sessions. Developers are granted access to positional and orientational data of body joints, which are directly integrated into the virtual environment within each game scenario. In both the ReMoVES and STORMS projects, all $2D$ and $3D$ graphical resources were sourced from various online repositories under Creative Commons licenses.

Performance data from each session are recorded. At the end of each gameplay session, relevant performance indicators are compiled into a JSON file for subsequent analysis. Furthermore, motion data are continuously logged, as Kinect captures and stores the positional data of body joints throughout the session. The resulting JSON file consists of an array of time-stamped events,

each represented as a set of key-value pairs. Certain keys, such as *Time*, *Score*, and *Kinect* (corresponding to the joint positions detected by the sensor), are common across all exergames. The key *Level* is included in games featuring multiple difficulty levels. Additionally, parameters such as *CalibSide*, *RealSide*, and *HandPositionX-Y* are recorded exclusively in games involving upper limb movements, representing respectively: the body side identified during the calibration phase (0 = none, 1 = right side, 2 = left side), the hand detected during active gameplay (0 = none, 1 = right hand, 2 = left hand), and the x and y coordinates of the hand within the game environment.

The principal cognitive functions targeted by the exergames include attention, memory, and executive functions. Within these domains, the activities are specifically designed to address the following cognitive components:

- **Working memory**: the capacity to control, regulate, and actively maintain task-relevant information in support of complex cognitive operations.
- **Inhibitory control**: an executive function enabling the suppression of automatic or impulsive responses in favor of goal-directed behaviors.
- **Selective attention**: the ability to focus cognitive resources on a specific object or stimulus in the environment for a defined period.
- **Task switching and cognitive shifting**: executive processes involving, respectively, the unconscious and conscious redirection of attention from one task to another.
- **Multitasking**: the capability to allocate attention to multiple simultaneous tasks or activities.
- **Sustained attention**: the ability to maintain cognitive focus on a task or stimulus over an extended duration.
- **Top-down attentional control**: the capacity to direct attention voluntarily toward task-relevant features, objects, or spatial regions, while suppressing irrelevant stimuli.

In some exergames, the patient is encouraged to reach consecutive on-screen targets with the arm motion (reaching task). The more targets are taken, the higher the in-game score. Such games aim at improving hand–eye coordination, and spatial awareness.

Indirect measurements of symptoms such as numbness of the body and/or extremities or spasticity that can complicate movements when playing with pelvic or limb movements can be obtained. Only the analysis of *Supermarkets* and *Shelf Cans* (Table 6.1) will be discussed in this thesis. The total game time is 60 seconds for both.

Table 6.1: Targeted cognitive domains of *Supermarket* and *Shelf Cans* exercise.

Exergames	**Targeted Exercise Gains**
Supermarket	Hand–eye coordination of voluntary arm movements Working memory Visual memory Inhibition and selective attention Processing speed
Shelf Cans	Hand–eye coordination of voluntary arm movements Reaction Time Inhibition and selective attention Processing speed

6.5.1 *Supermarket*

This exergame is set within a virtual supermarket environment, where the player is tasked with selecting specific items from store shelves. At the beginning of each session, a list of target objects to be collected is presented on screen, with a memorization period ranging from 8 to 25 seconds, depending on the selected difficulty level. Following this interval, the patient is required to identify and select the correct items by performing arm movements, detected via the Kinect sensor.

Real-time auditory feedback is provided throughout the task, delivering a distinct positive sound for correct selections and a negative sound in the event of incorrect choices. In levels one, three, and four, once the player successfully completes the initial list by collecting all required items, the same objects are repositioned randomly within the environment and displayed again, prompting the patient to locate and collect them a second time. This mechanic is intended to increase cognitive demand by requiring sustained attention, memory recall, and visuomotor coordination across successive trials.

- First level: a temporary list of three food names must be memorized in eight seconds. Then, the player has to collect the relative objects, moving them from the two lateral shelves into the shopping bag (placed in the middle of the screen). Some non-food distractors appear on the shelves. The semantic property (food or non-food) of the objects is the crucial correctness factor of the activity. The game score decreases if wrong objects are put into the bag. A visual feedback is also provided (initially yellow health bar turns red when mistakes are made).

- Second level: the player has to memorize four ordered objects in ten seconds. These objects must be collected among other distracting items sliding on the conveyor belt. The game score decreases if the wrong object is taken or if the order is not followed.

- Third level: the player must memorize and follow four sequenced instructions in twenty seconds. Each instruction refers to a different semantic characteristic (shape, color, or material) of the objects to be collected. Collected items will not re-appear.
- Fourth level: it is like the third level, but it differs for the higher number of the objects in the scene and the number of instructions (five instead of four, to remember in 25 s). In addition, the objects will reappear on the screen once collected.

In Figure 6.3, the screenshots of the four levels are provided, while Table 6.2 contains the list of features extracted. The aim of this game is to train memory (verbal, non-verbal, and visual), inhibition and selective attention, and hand–eye coordination.

(a) List of items in the Level 2.

(b) Level 1.

(c) Level 2.

(d) Level 3.

Figure 6.3: Some screenshots of the levels of Supermarket exergame.

Feature	**Description**
CorrectTarget	Number of targets correctly grabbed
WrongTarget	Number of targets erroneously grabbed
CorrectTargetNOList	Number of targets not in the list but semantically correct (only Level 1)
ObjOutOfSequence	Number of targets grabbed out of sequence (Levels 2, 3, 4 only)
PassedLevel	Number of completed rounds
TargetPositionX-Y	X and Y coordinates of the grabbed target
BagPositionX-Y	X and Y coordinates of the shopping bag (Level 1 only)
Hand-Object	Indicates whether the hand has grabbed an object (0 = no object, 1 = object grabbed)

Table 6.2: Features extracted from the *Supermarket* exergame.

6.5.2 *Shelf Cans*

In this exergame, the player is required to arrange colored soda cans onto a virtual shelving unit. Specifically, the task involves moving a can of a given color toward the corresponding shelf compartment. The shelving layout is organized such that the top-right, top-left, and bottom-left corners are designated for orange, green, and red cans, respectively.

The primary objective of *Shelf Cans* is to enhance patients' selective attention and rule-based decision-making, as participants must accurately match the color of each can to the appropriate shelf location. Errors are penalized through a score reduction and the delivery of a negative auditory cue, reinforcing correct execution and promoting attentional control.

A screenshot of the game is provided in Figure 6.4 and Table 6.3 contains the list of the extracted parameters. This game is aimed at stimulating hand-eye coordination of voluntary arm movements, reaction time, inhibition and selective attention, and processing speed.

Figure 6.4: Screenshot of Shelf Cans exergame.

Feature	**Description**
CorrectTarget	Number of cans correctly placed on the intended shelf
WrongTarget	Number of cans placed on the wrong shelf
originPositionX-Y	X and Y coordinates of the position where each can appears
targetPositionX-Y	X and Y coordinates of the target shelf position

Table 6.3: Features extracted from the *Shelf Cans* exergame.

6.6 Case Studies

In the following sections, a detailed analysis is presented for two patients who engaged in home-based telerehabilitation using the system over an extended period. The parameters monitored for these individuals, hereafter referred to as *Patient A* and *Patient B*, are comprehensively described.

Patient A is a 49-year-old male with an Expanded Disability Status Scale (EDSS) score of 7, who relies on a wheelchair for mobility. He used the ReMoVES system at home for approximately one month without a predefined therapeutic prescription. The principal findings emerging from his participation are reported in Section 6.6.1.

The patient discussed in Section 6.6.2, referred to as *Patient B*, is a 56-year-old female with an EDSS score of 6, requiring assisted ambulation. She utilized the system at home for a duration of four weeks. In her case, a structured weekly rehabilitation program was prescribed by the clinical team, specifying the exergames to be performed each week. At the conclusion of every week, a brief performance report was generated, enabling healthcare professionals to review outcomes and tailor subsequent therapy plans according to the patient's evolving needs. The weekly results will be examined in detail, comparing the prescribed activities with those actually completed by the patient.

Cognitive assessment scores at baseline ($T0$) for both *Patient A* and *Patient B* are reported in Table 6.4. The table includes scores from the Mini Mental State Examination (MMSE), Paced Auditory Serial Addition Test (PASAT, 3-second and 2-second intervals), Symbol Digit Modalities Test (SDMT), California Verbal Learning Test – Second Edition (CVLT-II), and Brief Visuospatial Memory Test – Revised (BVMT-R), along with the corresponding 5th percentile cutoff values. These cutoffs were adjusted to account for the patients' educational attainment, both having completed less than 12 years of formal education. Patient A is below the cutoff in PASAT, SDMT, and BVMT-R; Patient B is below the cutoff in SDMT and BVMT-R. PASAT tests failed (i.e., N).

	MMSE		PASAT-3″		PASAT-2″		SDMT		CVLT-II		BVMT-R
	Score	Cutoff	Score	Cutoff	Score	Cutoff	Score	Cutoff	Score	Cutoff	Score
Patient A	28.62	23.8	19	32	5	20	28.5	34.20	51.32	29.05	6/36
Patient B	29	23.8	N	32	N	20	26.2	34.20	52.05	29.05	2/36

Table 6.4: Cognitive assessment scores for Patient A and Patient B at $T0$. Below-grade values are highlighted in red. The 5th percentile cutoff values.

6.6.1 Patient A

Patient A used the ReMoVES system at home for almost a month without any prescription and his activity was observed throughout the period. Table 6.5 summarizes the number of sessions played for each exergame, specifying the chosen level.

Exergame	Level	N.Session
Shelf Cans	1	26
Supermarket	1	3

Table 6.5: Number of sessions performed by *Patient A* for each exergame.

As one can notice, the patient played mainly at the first level of each activity, focusing more on *Shelf Cans*.

Referring to the *Supermarket* exergame, the patient's cognitive failures concern two types of errors that may occur when taking various objects. A simple error, the so-called *semantic error* occurs when taking a wrong object belonging to the same category as the one described (food or no-food); a more serious error occurs when the taken object belongs to a different semantic class than the requested object. From Figure 6.5, it can be observed that the patient made some errors, even at the semantic level.

The errors committed in sessions 1, 2 and 3, respectively, are reported in Table 6.6 together with the objects in the to-do list. Such errors are most likely the reason why the patient played this game so little and only at the first level.

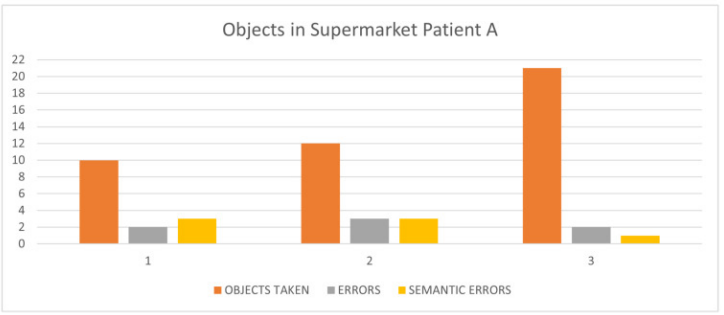


Figure 6.5: Objects taken, errors and semantic errors in supermarket sessions (Patient A). Semantic errors (in yellow) are considered less "serious" than errors (in grey).

Session	To-do-List	Wrong Objects
1	Banana–Pretzel–Watermelon	Book–Book
2	Chicken–Watermelon–Mushrooms	Glasses–Cap–Book
3	Pretzel–Mushrooms–Banana	Book–Book

Table 6.6: Error objects correspond to those highlighted in grey in Figure 6.5

The learning curve for the *Shelf Cans* activity reported in Figure 6.6 depicts the progression of gaming performance with increased experience measured by the number of sessions. Gaming performance is quantified as a percentage increase relative to the first session. As expected, one can observe a low score in the initial sessions and a subsequent steady improvement during the later sessions.

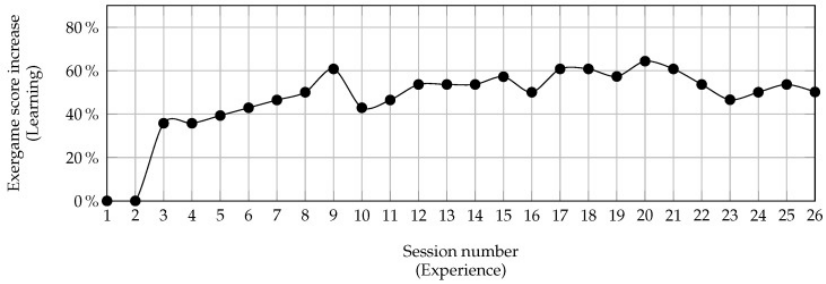


Figure 6.6: The learning curve for the *Shelf Cans* over 26 sessions, Patient A.

To analyze patient performance in executing the *Shelf Cans* exergame, various parameters are observed and plotted. Figure 6.7 shows a motor-cognitive analysis when the patient was playing *Shelf Cans* activity. Despite the trajectories seeming more controlled and precise, the patient hesitates and fails to lead the can of soda towards the shelf of the corresponding color, indicating a cognitive deficit.

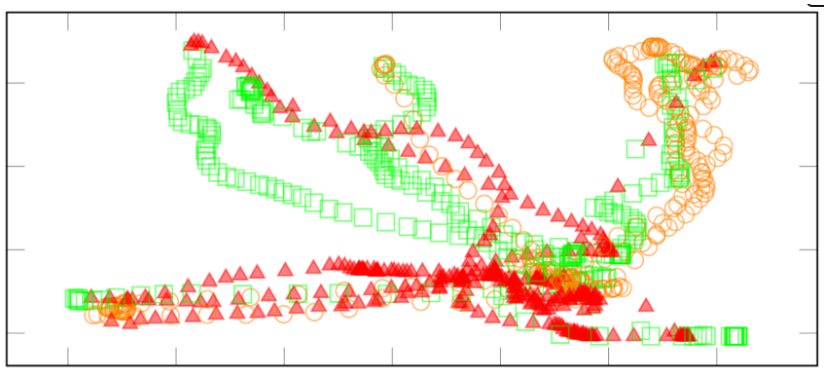


Figure 6.7: Each point corresponds to the temporal position of the tin can on the screen. Orange circles, red triangles, and green squares refer to orange cans, red cans, and green cans, respectively. In the session, the subject has a controlled movement but sometimes delivers the tin can with hesitation in the wrong shelf

Figure 6.8 provides an overview of the results, illustrating the angles between the optimal trajectories and the trajectories performed by patients, along with the time elapsed in moving the colored cans to their corresponding shelves. The graphs on the left depict the angles between the optimal trajectory and the trajectory executed by the patient. Meanwhile, the graphs on the right illustrate the time taken to complete the required movement. The results for the red, orange, and green trajectories are displayed from top to bottom, respectively. In addition, a correlation between the angles (indicating precision of movement) and the times (reflecting speed of execution) can be observed. For the red can, there exists a negative correlation ($\rho = -0.48$), suggesting that as speed increased, accuracy decreased. In contrast, for the orange can, there is a low correlation ($\rho = -0.10$). Notably, for the green can, a positive correlation is evident ($\rho = 0.53$), indicating the individual's ability to maintain both speed and precision in their movements.

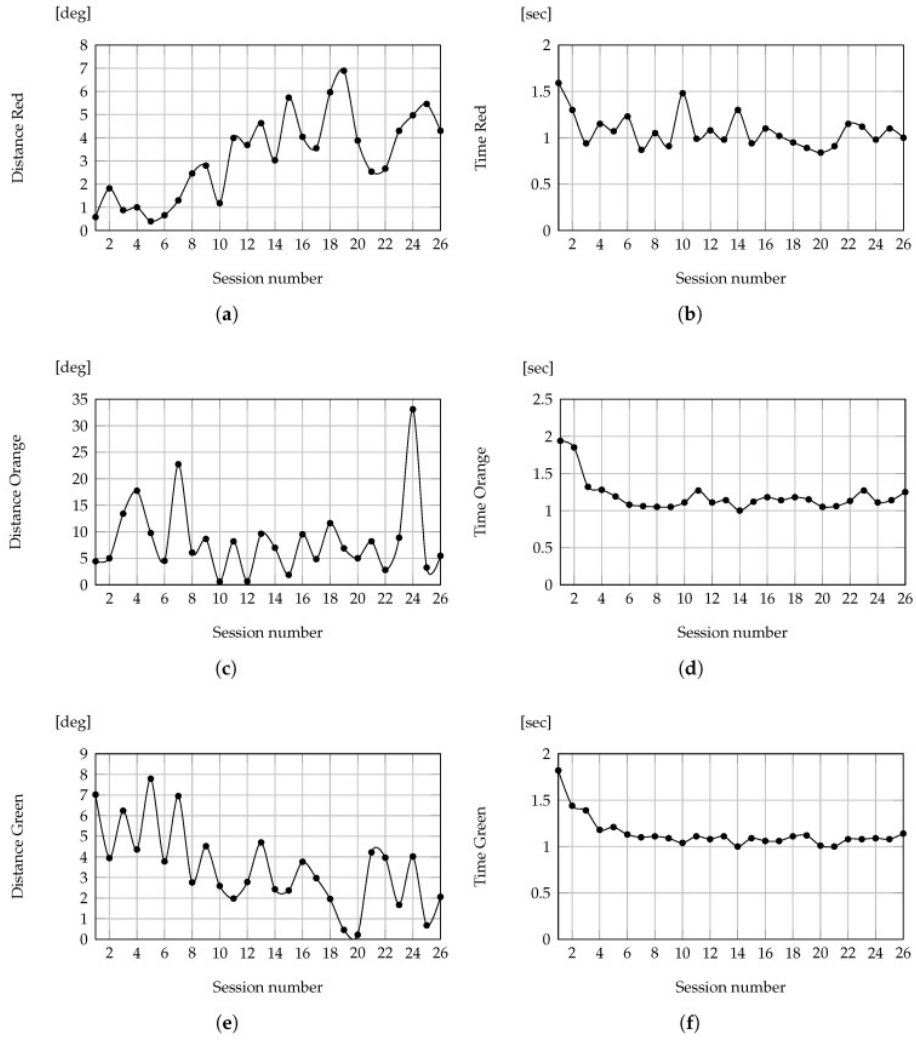


Figure 6.8: Angles between the optimal trajectory and the one performed by the patient (left). Execution times to perform the required movement (right). Subfigures (a,b) refer to the red cans, (c,d) to the orange cans, and (e,f) to the green cans.

The Figure 6.9 illustrates the temporal variations of the functional shoulder joint angles evaluated in three planes: frontal, sagittal and transverse. Each point represents a measure at a specific point in the follow-up. It is observed that the sagittal plane consistently shows the highest values, approaching $170-180^{\circ}$, indicating a wide freedom of movement forwards and backwards. This

suggests that this component of shoulder ROM is the least limited in the patient considered. In the frontal plane, however, the movement starts from lower values in the first measurements, but gradually tends to increase, stabilizing at values around $110-130°$. This may reflect progressive recovery or decreased restrictions in lateral movements. The transverse angle has the greatest limitations: the values are generally lower than in the other planes and show a lower tendency to improve over time. This suggests that the movements associated with the transverse plane (rotations, horizontal abductions or complex movements) are those most hindered in joint functioning. These data indicate, after the first sessions, significant recovery in sagittal and frontal movements, while restrictions persist in the transverse plane. Furthermore, the stability observed in each plane after the first increments suggests that a functional plateau was reached, useful as a reference for future evaluations.

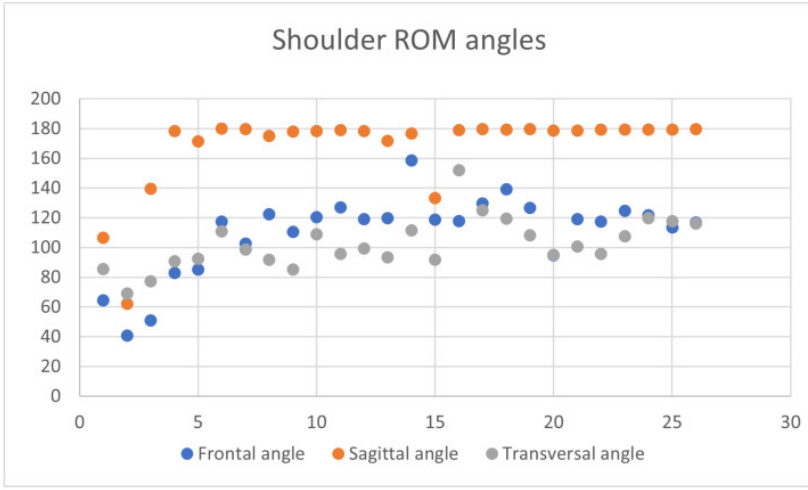
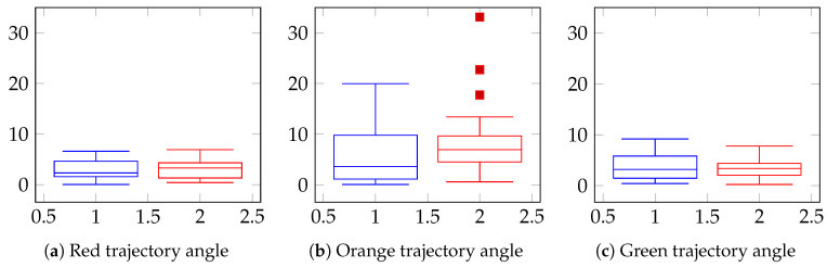


Figure 6.9: Shoulder angle range of motion: *Patient A*.

Finally, a comparison between *Patient A* and a group of healthy subjects is depicted in the box plots illustrating the angle between trajectories (see Figure 6.10). For the red and green trajectories, both the patient's (in red) and the healthy group's (in blue) box plots exhibit striking similarity, displaying a certain symmetry in the data distribution. In contrast, for the orange trajectory, the patient's box plots include some outliers, representing the mistakes made in placing some cans on incorrect shelves.



(a) Red trajectory angle (b) Orange trajectory angle (c) Green trajectory angle

Figure 6.10: Box plot depicting the values of trajectory angles of Healthy Subjects (HS, blue) and *Patient A* (patA, red).

6.6.2 Patient B

The second study incorporates the prescription provided to the patient undertaking rehabilitation exercises at home. *Patient B* used the ReMoVES system at home for a 4-week period. Table 6.7 lists the exergames recommended for each week, along with the level and the minimum number of sessions suggested.

Week	Exergame	Level	Prescription
1	Shelf Cans	1	4
	Supermarket	1	4
2	Shelf Cans	-	-
	Supermarket	1	5
3	Shelf Cans	-	-
	Supermarket	1	5
4	Supermarket	1	5
	Supermarket	2	5

Table 6.7: Plan of care of the activities prescribed for 4 weeks.

Remote observation of the patient's activity includes an assessment of adherence to the prescription and provides insights into the patient's condition, including levels of difficulty in task execution, fatigue, and stress. Although the patient did not strictly adhere to the prescription during the first week, she achieved commendable results in the assigned activities. Furthermore, there was a notable decrease in the number of sessions conducted during the final week.

A comparison between actual and prescribed gaming sessions highlighted a significant discrepancy. Various factors likely contributed to this gap, affecting the patient's participation across different activities and prescriptions. The necessity to juggle multiple activities may have impacted her motivation and willingness to consistently engage in rehabilitation sessions. To enhance future interventions, it may be beneficial to adopt a more comprehensive and balanced

approach tailored to individual needs and prescribed activities, which can benefit from the available observations. Addressing factors influencing participation can help ensure greater consistency in exercise adherence and facilitate the achievement of desired outcomes.

6.6.3 Cognitive Assessment

Table 6.8 presents the cognitive assessment scores for *Patient A* and *Patient B* at $T1$, at the end of 10 sessions. One can notice a general improvement, which is particularly significant for PASAT-3″ and SDMT, which is now above the cutoff for both patients. Finally, one month after the end of treatment (i.e, $T2$), Table 6.9 shows that all indexes have been preserved or improved. BVMT-R scores show a significant increase from $T0$ and after $T1$ even though they are still below the cutoff but not too far from reaching it.

	MMSE		PASAT-3″		PASAT-2″		SDMT		CVLT-II		BVMT-R
	Score	Cutoff	Score	Cutoff	Score	Cutoff	Score	Cutoff	Score	Cutoff	Score
Patient A	30	23.8	27	32	6	20	54	34.20	55	29.05	23/36
Patient B	29	23.8	15	32	3	20	36	34.20	56	29.05	20/36

Table 6.8: Cognitive assessment scores for *Patient A* and *Patient B* at $T1$. Values below the cutoff are highlighted in red. Values in blue are below the cutoff but not far from reaching it.

	MMSE		PASAT-3″		PASAT-2″		SDMT		CVLT-II		BVMT-R
	Score	Cutoff	Score	Cutoff	Score	Cutoff	Score	Cutoff	Score	Cutoff	Score
Patient A	30	23.8	27	32	8	20	60	34.20	52	29.05	25/36
Patient B	30	23.8	16	32	5	20	36	34.20	54	29.05	27/36

Table 6.9: Cognitive assessment scores for *Patient A* and *Patient B* at $T2$. Values below the cutoff are highlighted in red. Values in blue are below the cutoff but not far from reaching it.

Finally, Figure 6.11 and Figure 6.12 compare all cognitive measures at baseline ($T0$), at the end of the 10 sessions ($T1$), and one month after the end of treatment ($T2$).

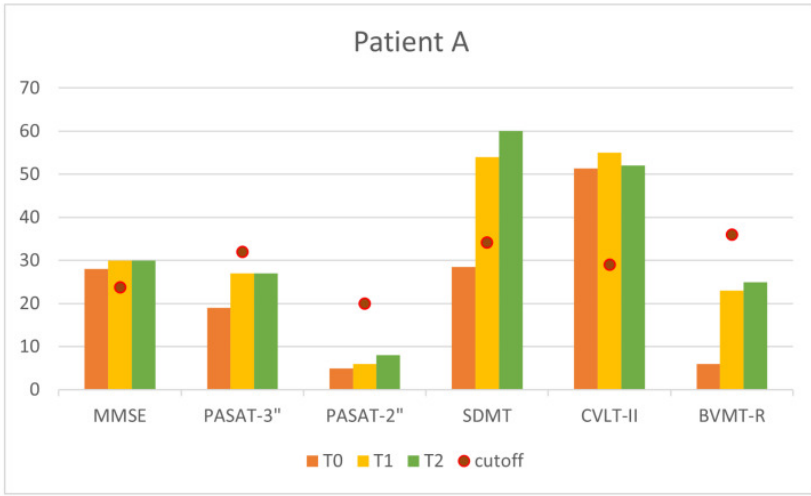


Figure 6.11: Cognitive measures at $T0$,$T1$, and $T2$ for Patient A.

The first one shows the trend of *Patient A* cognitive performance evaluated at three different time points (T0=orange, T1=yellow, T2=green) through a battery of standardized neuropsychological tests. The red dots indicate the reference cut-off values, i.e. the clinical threshold below which performance is considered deficient. From the graph we observe that the MMSE score remains stable and above the clinical threshold in all sessions. In the PASAT-3″ and PASAT-2″ tests, the initial scores are below the cut-off, but show progressive improvement, particularly for PASAT-2″. In the processing speed test (SDMT), the patient presents a significant increase from T0 to T2, significantly exceeding the cut-off. Verbal memory (CVLT-II) remains relatively stable, always at values above the threshold. Visuospatial memory (BVMT-R) shows improvement from T0 to T2, with scores only passing the cut-off in subsequent assessments. The general improvement of all the indexes can be observed at once.

The second one represent B patient's performance for the same different time points. Also in this case, the red dots indicate the clinical reference cut-off values, used to discriminate normal performance from deficient performance.

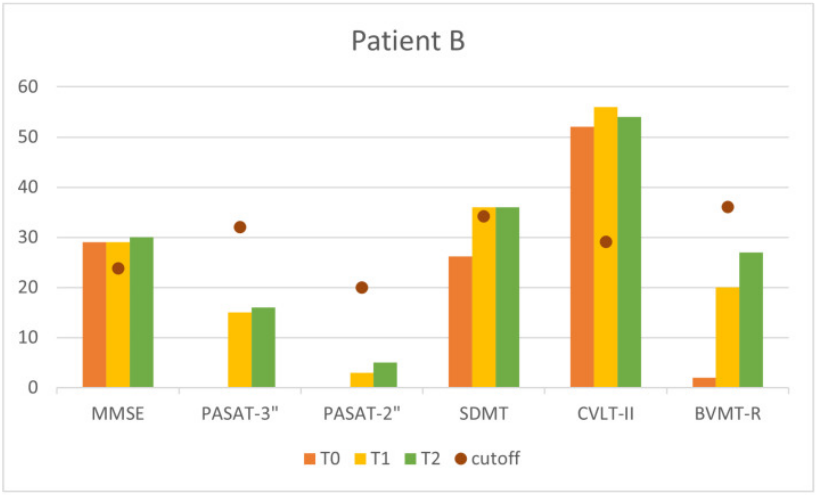


Figure 6.12: Cognitive measures at $T0$,$T1$, and $T2$ for Patient B.

The analysis shows that the MMSE score remains constant and stably above the clinical threshold, indicating adequate global cognitive function. In divided attention tests (PASAT-3″ and PASAT-2″), the values are low and consistently below the cut-off, suggesting persistent impairment in this cognitive domain, with no obvious improvement over time. In the processing speed test (SDMT), the patient shows a significant increase from T0 to T2, reaching values close to or above the clinical threshold, indicating a partial recovery of attentional skills and rapid processing. Verbal memory (CVLT-II) remains elevated and consistently above the cut-off in all assessments, with a slight improvement over time. Visuospatial memory (BVMT-R) shows initially deficient values (T0), but with a clear improvement at T2, in which the patient exceeds the reference threshold. Overall, the *Patient B* profile highlights stability of general cognitive functions, a persistent deficit in PASAT tasks, but also a progressive improvement in fast processing skills and visuospatial memory, which become clinically relevant in final assessments.

6.7 Case Study Results

In the framework of the STORMS project, the utilization of ReMoVES has been positively embraced by all patients, regardless of age, including elderly individuals typically less inclined towards emerging technologies. This has resulted in increased adherence to rehabilitation protocols, also enhancing the duration of treatment per session. An informed consensus was obtained from the participants before the beginning of the activity and they consented to the processing

of the data obtained for scientific research purposes; in addition, the principles of the Declaration of Helsinki were followed.

The cognitive assessments administered at T0, T1, and T2 indicate an overall improvement across all evaluated indices for Patients A and B. In particular, marked gains are observed in the SDMT, BVMT-R, and PASAT scores. Nevertheless, PASAT-2″ and PASAT-3″ values remain below the established cutoff thresholds. With respect to motor performance, continuous monitoring was shown to provide a comprehensive and dynamic characterization of both patient status and functional progression. The collected metrics include, but are not limited to, the number of exergames performed, repetition frequency and execution speed, game scores, postural parameters, joint angles, and range of motion. A comparison between the prescribed exercises and those actually executed proved essential for assessing patient adherence and the appropriateness of the rehabilitation protocol. This analysis enables the identification of mismatches between task demands and patient capabilities, thereby supporting the optimization and personalization of the rehabilitation process. Furthermore, the graphs and figures reported in Section 6.6 for both patients demonstrate a rapid learning curve, accompanied by consistent improvements in movement accuracy, control, execution precision, and task comprehension.

Current literature highlights a limited number of studies addressing home-based telerehabilitation, particularly in the context of multiple sclerosis. This paucity can be attributed to several factors, including technological constraints, as well as challenges related to system usability, user acceptance, and the frequent need for caregiver involvement. The majority of existing works have focused on rehabilitation technologies deployed in controlled clinical or laboratory environments, with comparatively less attention devoted to their feasibility, effectiveness, and long-term adoption in home-based settings.

Chapter 7

Conclusion

The aim of this thesis was to exploit the potential and cross-cutting nature of digital signal processing to introduce innovative methodologies for biomedical image processing and data analysis using markerless systems, with the overall goal of contributing to the advancement of digital and enabling technologies in the medical field. The work primarily focused on two main areas: diagnostic imaging and telerehabilitation. Both domains have gained increasing importance in recent years, reflecting the growing demand for flexible and efficient medical technologies.

7.1 Discussion

Diagnostic imaging and telerehabilitation are closely interconnected, both from a clinical perspective and from a technological standpoint. Clinically, rehabilitation plays a crucial role in the management of numerous pathological conditions, often initially identified through diagnostic imaging. From a technological perspective, tools developed for diagnostic purposes provide essential information that supports the design of effective rehabilitation strategies. Furthermore, the emergence of web-based technologies has enabled improved access to diagnostic resources and disease-management information, underscoring their growing significance in contemporary healthcare. The proposed approaches aim to integrate signal and image processing methodologies, based on mathematical modeling and machine learning algorithm, for the extraction, analysis, and interpretation of signals and imaging data. The overall objective is to ensure robust, adaptable and clinically meaningful data analysis in heterogeneous but complementary medical applications.

With regard to diagnostic imaging (Chapters 3 and 5), the main contribution of this work lies in the integration of traditional image processing techniques within a preprocessing stage, combined with machine learning algorithms for classification. Specifically, the preprocessing phase introduces a Focus-of-Attention Mechanism designed to enhance the gray levels of the Glisson line while attenuating the intensity values associated with the hepatic parenchyma and surrounding tissues. This procedure yields a Region of Contrast Interest (ROCI), in which potential sources of classification bias are minimized. For the classification task, a convolutional neural network (CNN)-based model is employed.

With regard to telerehabilitation (Chapters 4 and 6), and within framework

of the STORMS project, an Internet of Medical Things (IoMT) system leveraging ReMoVES technology was specifically developed for the rehabilitation of patients with multiple sclerosis. The data collected from such patients were analyzed using State-of-The-Art methodologies, with the aim of supporting control tasks and identifying suitable indicators to effectively summarize game-based rehabilitation sessions.

7.2 Future Perspectives

The integration of IoT, DSP, and AI technologies is enabling novel and increasingly effective approaches for the analysis and interpretation of complex biomedical data.

Conducting research within clinical and industrial environments represents a valuable opportunity, as it allows researchers to directly address one of the most critical challenges in translational research: the fact that a growing volume of scientific output in this field does not necessarily lead to tangible clinical advancements. In many instances, research efforts are primarily driven by academic incentives rather than by the concrete needs of clinicians and patients. In this context, the deployment and evaluation of the proposed research framework within real clinical workflows proved to be essential for systematically assessing both its potential clinical benefits and its inherent limitations.

Specifically, the proposed pipeline—comprising ultrasound image acquisition, application of a Focus-of-Attention Mechanism, and subsequent classification via a neural network—may serve as a valuable diagnostic support tool by significantly reducing operator-dependent variability. Nevertheless, for such a system to be clinically applicable, a minimum accuracy threshold of approximately 80% is generally required. While the proposed approach satisfies this requirement in the binary classification setting, the performance achieved in the multiclass scenario remains insufficient. Addressing this limitation necessitates the availability of larger and more diverse datasets, both within the proposed framework and across alternative deep learning-based approaches. A promising direction for future work involves validating the proposed model on datasets acquired using heterogeneous imaging devices, in order to assess the robustness and generalization capability of the Focus-of-Attention Mechanism under real-world variability.

In the field of telerehabilitation, IoT-based technologies offer significant potential benefits for patient care, including improved continuity of treatment, increased personalization of rehabilitation protocols, and enhanced patient engagement. Despite this potential, the existing literature reveals a limited number of contributions specifically addressing home-based telerehabilitation for patients affected by multiple sclerosis. This shortage can be attributed to multiple factors, including technological constraints, as well as challenges related to system usability, user acceptance, and the frequent requirement for caregiver as-

sistance. Consequently, most studies have focused on rehabilitation technologies deployed in controlled clinical or experimental environments, with comparatively limited attention devoted to their real-world applicability and effectiveness in home-based scenarios. A promising direction for future research can be lies in the integration of machine learning–based biofeedback systems within telerehabilitation platforms, capable of reliably discriminating between correctly and incorrectly executed exercises. Such systems could provide real-time feedback to patients, reduce the risk of improper execution and associated injuries, and ultimately contribute to improved rehabilitation outcomes.

The world of IoT, includes the development of architectures and platforms, security requirements, policies and regulations that should be taken into account in future research especially if other technologies, such as big data and cognitive systems, could fit into this context. At the same time, artificial intelligence presents both opportunities and challenges. While neural networks and advanced learning models have shown remarkable potential for classification and prediction tasks, their adoption is limited by the scarcity of annotated data, variability across clinical populations, and the well-known *black-box* problem. Ethical, regulatory, and computational constraints further highlight the need for caution and rigorous validation. Future prospects are moving towards the adoption of approaches such as federated learning, which allows for the distributed training of models on clinical data belonging to multiple institutions without violating privacy, and the development of explainable AI techniques, aimed at making the decision-making processes of models more transparent and interpretable.

Essential will be the synergy between DSP and AI. Signal processing will play a key role in the extraction of meaningful features, thereby improving the reliability and interpretability of machine learning models applied in clinical practice. These advances will also support the development of personalized therapeutic strategies, where adaptive signal analysis allows treatments to be tailored dynamically to the individual patient's condition. Taken together, these perspectives highlight how DSP will continue to play a central role in the evolution of digital medicine, supporting both diagnostic innovation and the delivery of more effective, personalized, and secure healthcare solutions.

Chapter 8

PhD activities

8.1 Publications record

International Journals

- M. Trombini, F. Ferraro, G. Iaconi, L. Vestito, F. Bandini, L. Mori, C. Trompetto, S. Dellepiane, "A study protocol for occupational rehabilitation in multiple sclerosis," Sensors, vol. 21, no. 24, p. 8436, 2021. DOI: 10.3390/s21248436
- Vestito, L., Ferraro, F., Iaconi, G., Genesio, G., Bandini, F., Mori, L., Trompetto, C., Dellepiane, S. STORMS: A Pilot Feasibility Study for Occupational TeleRehabilitation in Multiple Sclerosis. Sensors 2024, 24, 6470.

  DOI: https://doi.org/10.3390/s24196470
- Dellepiane, S., Vestito, L., Mori, L., Trompetto, C., Bandini, F., Iaconi, G., Ferraro, F. and Genesio, G., 2024. STORMS: Occupational Telerehabilitation in Multiple Sclerosis.
- Iaconi, G., Wehbe, A., Borro, P., Macciò, M. and Dellepiane, S., 2025. A Supervised System Integrating Image Processing and Machine Learning for the Staging of Chronic Hepatic Diseases. Electronics, 14(8), p.1534.

International Conferences

- Conference SysInt: Iaconi, G. et al. (2022). Graph-Based Segmentation and Markov Random Field for Covid-19 Infection in Lung CT Volumes. In:, et al. Advances in System-Integrated Intelligence. SYSINT 2022. Lecture Notes in Networks and Systems, vol 546. Springer, Cham. https://doi.org/10.1007/978-3-031-16281-7_5
- Conference Healthcom: F. Ferraro, G. Iaconi, M. Simonini, S. Dellepiane (2022, October). "Signal processing for remote monitoring of home-based rehabilitation support activities", In 2022 IEEE International Conference on E-health Networking, Application & Services (HealthCom) (pp. 192-198). IEEE, DOI: 10.1109/HealthCom54947.2022.9982749, ISBN: 978-1-6654- 8016-1

Book Chapters

- F. Ferraro, G. Iaconi, G. Genesio, R. Truffelli, R. Amella, M. Simonini, S. Dellepiane. (2023, September). "Spatial Exploration Indicators in the Remote Assessment of Visual Neglect". In International Conference on Image Analysis and Processing (pp. 552-563). Cham: Springer Nature Switzerland.DOI: 10.1007/978-3-031-43153-1_46,ISBN:978-3- 031-43153-1

- G. Iaconi, F. Ferraro, M. Balletto, D. Solarna, M. Trombini, G. Moser, S. Dellepiane, "Graph-based segmentation and markov random field for covid-19 infection in lung ct volumes," in Advances in System-Integrated Intelligence (M. Valle, D. Lehmhus, C. Gianoglio, E. Ragusa, L. Seminara, S. Bosse, A. Ibrahim, and K.-D. Thoben, eds.), (Cham), pp. 43–52, Springer International Publishing, 2022. DOI: 10.1007/978-3-031- 16281-7_5, ISBN: 978-3-031-16281-7

8.2 Scientific Society

- IEEE EMBS - Engineering in Medicine and Biology Society - graduate student member.

- IEEE GRSS - Geoscience and Remote Sensing Society - graduate student member.

- IEEE SPS - Signal Processing Society - graduate student member.

- IEEE BISP Technical Committee - Bio Imaging and Signal Processing - affiliate member.

8.3 Conferences and Workshops

- International Conference on Image Analysis and Processing (ICIAP 2023), 11-15 September 2023, Udine. Poster presentation of "Spatial Exploration Indicators in the Remote Assessment of Visual Neglect".

8.4 Ph.D Summer/Winter schools

- Machine Learning: A Computational Intelligence Approach 2022, held by Prof. Francesco Masulli, Prof. Stefano Rovetta, Unige.

- Deep Learning and Computer Vision School (DLCV) 2023, MaLGa Center, Genoa from the 5th to the 9th of June 2023.

8.5 Seminars

- "La medicina digitale: stato dell'arte e prospettive", Liguria Digitale, Genova.
- "La plasticità del cervello in età adolescenziale", IANUA seminar, Genova.
- Pillole Di Telemedicina (IV) - Telemonitoraggio e teleriabilitazione con focus sulle tecnologie, Ordine degli Ingegneri di Genova, Genova.

8.6 Teaching activities

- Assistant and Subject Expert for the course of Elaborazione digitale delle immagini storico-artistiche, Master of Storia dell'Arte e Valorizzazione del Patrimonio Artistico, University of Genoa, Italy.
- Teaching support activities (laboratories) for the course of Digital Image Processing, Master of science course in Internet and Multimedia Engineering, Master of science course in Electronic Engineering, University of Genoa, Italy.
- Seminar on Convolution Neural Network (CNN) for the course of Digital Image Processing, Master of science course in Internet and Multimedia Engineering, Master of science course in Electronic Engineering, University of Genoa, Italy.
- Co-supervisor of B.Sc. Theses in Biomedical Engineering.

8.7 Experience developed outside the University of Genoa during PhD

- Activity: STORMS project (Solution Towards Occupational Rehabilitation for Multiple Sclerosis) aimed at the design and development of serious games for cognitive assessment and rehabilitation for patients with multiple sclerosis.

  Period of time: 6 month (2022)

  Hosting institution / industry: Ospedale Policlinico San Martino

- Activity: GLQ-O1 project to create a classification system for liver fibrosis, using a set of B-mode ultrasound images, as an alternative to currently proposed methods that base staging level identification on elastometry and biopsy.

Period of time: 1 years (2023-2024)

Hosting institution/industry: Ospedale Policlinico San Martino ed Esaote S.p.A.

8.8 Other Activities

- Partecipation in the Rotary Project "Medicina Digitale per la Prevenzione e la Cura", January- May 2023, Genova.
- Project organization and coordination in the Erasmus+ KA 107 (Italia-Africa), 2021-2023.
- Collaboration with ASL3 for tele-rehabilitation and telemedicine project.

Acronyms

AI Artificial Intelligence.

ANN Artificial Neural Network.

CAD Computer-aided detection.

CI Cognitive impairment.

CLD chronic liver diseases.

CNN Convolutional Neural Network.

CNS central nervous system.

CT X-ray Computed Tomography.

DSP digital signal processing.

ICT Information and Communication Technologies.

IID Independent and Identically Distributed.

IoMT Internet of Medical Things.

IoT Internet of Things.

M2M machine-to-machine.

ML Machine Learning.

MRI Magnetic Resonance Imaging.

MS Multiple Sclerosis.

ODV Outlier Detection Values.

PCA Principal Components Analysis.

PET Positron Emission Tomography.

ReMoVES Remote Monitoring Validation Engineering System.

RMI Measurement Reliability Index.

ROCI Regions of Contrast Interest.

ROI Region of Interest.

ROM Range-of-Movement.

SNR Signal-to-Noise Ratio (SNR).

SoA State of the Art.

SRAD Speckle Reducing Anisotropic Diffusion.

STORMS Solution Towards Occupational Rehabilitation in Multiple Sclerosis.

SVM Support Vector Machine.

SWE Shear Wave Elastography.

ToF Time-of-Flight.

US Ultrasound Imaging.

VR virtual reality.

References

[1] N. C. Institute, "Category: Anatomical planes on human," *National Cancer Institute*, accessed on June 2025. [Online]. Available: https://commons.wikimedia.org/wiki/Category:Anatomical_planes_on_human

[2] P. Prandoni and M. Vetterli, *Signal processing for communications*. EPFL press, 2008.

[3] S. G. Dellepiane, F. Ferraro, C. Baffigo, and M. Simonini, "Signal processing and feature extraction in markerless telerehabilitation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 33, pp. 911–924, 2025.

[4] G. Kurillo, E. Hemingway, M.-L. Cheng, and L. Cheng, "Evaluating the accuracy of the azure kinect and kinect v2," *Sensors*, vol. 22, no. 7, p. 2469, 2022.

[5] H. Azhari, J. A. Kennedy, N. Weiss, and L. Volokh, *From Signals to Image*. Springer, 2020.

[6] J. Thirumaran and S. Shylaja, "Medical image processing–an introduction," *International Journal of Science and Research (IJSR)*, vol. 4, no. 11, pp. 1197–1199, 2015.

[7] T. H. Marwick, "Future of echocardiography in australia," *Heart, Lung and Circulation*, vol. 28, no. 9, pp. 1307–1309, 2019.

[8] A. S. Klauser and P. Peetrons, "Developments in musculoskeletal ultrasound and clinical applications," *Skeletal radiology*, vol. 39, pp. 1061–1071, 2010.

[9] A. Berzigotti and L. Castera, "Update on ultrasound imaging of liver fibrosis," *Journal of hepatology*, vol. 59, no. 1, pp. 180–182, 2013.

[10] H. Routh, "Doppler ultrasound," *IEEE Engineering in Medicine and Biology Magazine*, vol. 15, no. 6, pp. 31–40, 1996.

[11] X.-W. Cui, K.-N. Li, A.-J. Yi, B. Wang, Q. Wei, G.-G. Wu, and C. F. Dietrich, "Ultrasound elastography," *Endoscopic ultrasound*, vol. 11, no. 4, pp. 252–274, 2022.

[12] S. L. Mulvagh, A. N. DeMaria, S. B. Feinstein, P. N. Burns, S. Kaul, J. G. Miller, M. Monaghan, T. R. Porter, L. J. Shaw, F. S. Villanueva *et al.*, "Contrast echocardiography: current and future applications," *Journal of the American Society of Echocardiography*, vol. 13, no. 4, pp. 331–342, 2000.

[13] T. A. Michailovich OV, "Despeckling of medical ultrasound images," *IEEE Trans Ultrason Ferroelectr Freq Control*, 2006.

[14] S. K. Jespersen, J. E. Wilhjelm, and H. Sillesen, "Multi-angle compound imaging," *Ultrasonic imaging*, vol. 20, no. 2, pp. 81–102, 1998.

[15] B. Halalli and A. Makandar, "Computer aided diagnosis-medical image analysis techniques," *Breast imaging*, vol. 85, pp. 85–109, 2018.

[16] D. J. Winkel, T. Heye, T. J. Weikert, D. T. Boll, and B. Stieltjes, "Evaluation of an ai-based detection software for acute findings in abdominal computed tomography scans: toward an automated work list prioritization of routine ct examinations," *Investigative radiology*, vol. 54, no. 1, pp. 55–59, 2019.

[17] M. Wu, Q. Sun, and J. Wang, "Medical image retrieval based on combination of visual semantic and local features," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 5, no. 4, pp. 43–55, 2012.

[18] L. Rundo, A. Tangherloni, M. S. Nobile, C. Militello, D. Besozzi, G. Mauri, and P. Cazzaniga, "Medga: a novel evolutionary method for image enhancement in medical imaging systems," *Expert Systems with Applications*, vol. 119, pp. 387–399, 2019.

[19] P. Mittal, R. Saini, and N. K. Jain, "Image enhancement using fuzzy logic techniques," in *Soft Computing: Theories and Applications: Proceedings of SoCTA 2017*. Springer, 2019, pp. 537–546.

[20] L. Fang, S. Li, D. Cunefare, and S. Farsiu, "Segmentation based sparse reconstruction of optical coherence tomography images," *IEEE transactions on medical imaging*, vol. 36, no. 2, pp. 407–421, 2016.

[21] N. Kanopoulos, N. Vasanthavada, and R. L. Baker, "Design of an image edge detection filter using the sobel operator," *IEEE Journal of solid-state circuits*, vol. 23, no. 2, pp. 358–367, 1988.

[22] H. S. Islam, Shaikh Mahmudul e Mondal, "Analisi delle immagini mediche basata sul miglioramento dell'immagine," in *2019 10a Conferenza Internazionale sulle Tecnologie di Informatica, Comunicazione e Networking (ICCCNT)*, 2019.

[23] R. Kaur and S. Kaur, "Comparison of contrast enhancement techniques for medical image," in *2016 conference on emerging devices and smart systems (ICEDSS)*. IEEE, 2016, pp. 155–159.

[24] G. Singh, A. Mittal *et al.*, "Various image enhancement techniques-a critical review," *International Journal of Innovation and Scientific Research*, vol. 10, no. 2, pp. 267–274, 2014.

[25] B. Ganesan, G. Yamuna, and S. K. Suman, "Hybrid contrast enhancement approach for medical image," in *Proc. Nat. Conf. Emerg. Trends Inf., Commun. Technol., Int. J. Comput. Appl.* Citeseer, 2013, pp. 1–12.

[26] S. M. Hama and M. S. Al-Ani, "Medical image enhancement based on an efficient approach for adaptive anisotropic diffusion," *International Journal of Advances in Engineering and Technology*, vol. 6, no. 3, p. 1424, 2013.

[27] S. Bhattacharya, S. Gupta, and V. K. Subramanian, "Localized image enhancement," in *2014 Twentieth National Conference on Communications (NCC)*. IEEE, 2014, pp. 1–6.

[28] J. Gerstenmaier and R. Gibson, "Ultrasound in chronic liver disease," *Insights into imaging*, vol. 5, no. 4, pp. 441–455, 2014.

[29] H. Toyoda, T. Kumada, N. Kamiyama, K. Shiraki, K. Takase, T. Yamaguchi, and H. Hachiya, "B-mode ultrasound with algorithm based on statistical analysis of signals: evaluation of liver fibrosis in patients with chronic hepatitis c," *American Journal of Roentgenology*, vol. 193, no. 4, pp. 1037–1043, 2009.

[30] C.-C. Choong, S. K. Venkatesh, and E. P. Siew, "Accuracy of routine clinical ultrasound for staging of liver fibrosis," *Journal of clinical imaging science*, vol. 2, p. 58, 2012.

[31] T. Nishiura, H. Watanabe, M. Ito, Y. Matsuoka, K. Yano, M. Daikoku, H. Yatsuhashi, K. Dohmen, and H. Ishibashi, "Ultrasound evaluation of the fibrosis stage in chronic liver disease by the simultaneous use of low and high frequency probes," *The British journal of radiology*, vol. 78, no. 927, pp. 189–197, 2005.

[32] R. N. Gibson, "Ultrasound evaluation of hepatic vein morphology: a promising simple diagnostic tool?" *Imaging in Medicine*, vol. 3, no. 2, p. 145, 2011.

[33] R. A. Filly, S. G. Reddy, A. B. Nalbandian, Y. Lu, and P. W. Callen, "Sonographic evaluation of liver nodularity: Inspection of deep versus superficial surfaces of the liver," *Journal of clinical ultrasound*, vol. 30, no. 7, pp. 399–407, 2002.

[34] S. Vessal, S. Naidoo, J. Hodson, D. L. Stella, and R. N. Gibson, "Hepatic vein morphology: a new sonographic diagnostic parameter in the investigation of cirrhosis?" *Journal of Ultrasound in Medicine*, vol. 28, no. 9, pp. 1219–1227, 2009.

[35] A. K. Lim, N. Patel, R. J. Eckersley, Y.-T. Kuo, R. D. Goldin, H. C. Thomas, D. O. Cosgrove, S. D. Taylor-Robinson, and M. J. Blomley, "Can doppler sonography grade the severity of hepatitis c-related liver disease?" *American Journal of Roentgenology*, vol. 184, no. 6, pp. 1848–1853, 2005.

[36] F. Piscaglia, L. Bolondi *et al.*, "The safety of sonovue® in abdominal applications: Retrospective analysis of 23188 investigations," *Ultrasound in medicine & biology*, vol. 32, no. 9, pp. 1369–1375, 2006.

[37] N. Frulio and H. Trillaud, "Ultrasound elastography in liver," *Diagnostic and interventional imaging*, vol. 94, no. 5, pp. 515–534, 2013.

[38] T. Shiina, "Jsum ultrasound elastography practice guidelines: basics and terminology," *Journal of Medical Ultrasonics*, vol. 40, no. 4, pp. 309–323, 2013.

[39] A. Guibal, G. Renosi, A. Rode, J. Scoazec, O. Guillaud, L. Chardon, M. Munteanu, J. Dumortier, F. Collin, and T. Lefort, "Shear wave elastography: an accurate technique to stage liver fibrosis in chronic liver diseases," *Diagnostic and interventional imaging*, vol. 97, no. 1, pp. 91–99, 2016.

[40] X.-q. Zhang, R.-q. Zheng, J.-y. Jin, J.-f. Wang, T. Zhang, and J. Zeng, "Us shear-wave elastography dispersion for characterization of chronic liver disease," *Radiology*, vol. 305, no. 3, pp. 597–605, 2022.

[41] F. Destrempes, M. Gesnik, B. Chayer, M.-H. Roy-Cardinal, D. Olivié, J.-M. Giard, G. Sebastiani, B. N. Nguyen, G. Cloutier, and A. Tang, "Quantitative ultrasound, elastography, and machine learning for assessment of steatosis, inflammation, and fibrosis in chronic liver disease," *PLoS One*, vol. 17, no. 1, p. e0262291, 2022.

[42] I. Gatos, S. Tsantis, S. Spiliopoulos, D. Karnabatidis, I. Theotokas, P. Zoumpoulis, T. Loupas, J. D. Hazle, and G. C. Kagadis, "A machine-learning algorithm toward color analysis for chronic liver disease classification, employing ultrasound shear wave elastography," *Ultrasound in medicine & biology*, vol. 43, no. 9, pp. 1797–1810, 2017.

[43] G. Iaconi, A. Wehbe, P. Borro, M. Macciò, and S. Dellepiane, "A supervised system integrating image processing and machine learning for the staging of chronic hepatic diseases," *Electronics*, vol. 14, no. 8, p. 1534, 2025.

[44] F. Destrempes, M. Gesnik, B. Chayer, M.-H. R. Cardinal, D. Olivié, J.-M. Giard, G. Sebastiani, B. N. Nguyen, G. Cloutier, and A. Tang, "Machine learning based on quantitative ultrasound for assessment of chronic liver disease," in *2020 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2020, pp. 1–3.

[45] M. M. Adnan, M. S. M. Rahim, A. R. Khan, A. Alkhayyat, F. S. Alamri, T. Saba, and S. A. Bahaj, "Automated image annotation with novel features based on deep resnet50-slt," *IEEE Access*, vol. 11, pp. 40 258–40 277, 2023.

[46] N. Dong, Q. Feng, M. Zhai, J. Chang, and X. Mai, "A novel feature fusion based deep learning framework for white blood cell classification," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 8, pp. 9839–9851, 2023.

[47] M. Micucci and A. Iula, "Recent advances in machine learning applied to ultrasound imaging," *Electronics*, vol. 11, no. 11, p. 1800, 2022.

[48] S. S. Yadav and S. M. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *Journal of Big data*, vol. 6, no. 1, pp. 1–18, 2019.

[49] T. Mahmood, A. Rehman, T. Saba, L. Nadeem, and S. A. O. Bahaj, "Recent advancements and future prospects in active deep learning for medical image segmentation and classification," *IEEE Access*, vol. 11, pp. 113 623–113 652, 2023.

[50] S. Yadav and S. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *Journal of Big Data*, vol. 6, 12 2019.

[51] Y. Zhang, Y. Zhang, Y. Zhang, D. Wang, F. Peng, S. Cui, and Z. Yang, "Ultrasonic image fibrosis staging based on machine learning for chronic liver disease," in *2021 IEEE International Conference on Medical Imaging Physics and Engineering (ICMIPE)*, 2021, pp. 1–5.

[52] M. Biswas, V. Kuppili, D. R. Edla, H. S. Suri, L. Saba, R. T. Marinhoe, J. M. Sanches, and J. S. Suri, "Symtosis: A liver ultrasound tissue characterization and risk stratification in optimized deep learning paradigm," *Computer methods and programs in biomedicine*, vol. 155, pp. 165–177, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:3450471

[53] M. Trombini, P. Borro, S. Ziola, and S. Dellepiane, "A digital image processing approach for hepatic diseases staging based on the glisson's capsule," in *2020 2nd International Conference on Electrical, Control and Instrumentation Engineering (ICECIE)*. IEEE, 2020, pp. 1–6.

[54] P. Borro, S. Dellepiane, R. Pellicano, L. Gemme, S. Fagoonee, and G. Testino, "Quantification of ultrasound imaging in the staging of hepatic fibrosis," *Panminerva medica*, vol. 60, pp. 44–51, 02 2018.

[55] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

[56] A. Khan, A. Sohail, and U. e. a. Zahoora, "A survey of the recent architectures of deep convolutional neural networks," *Springer*, vol. 53, p. 5455–5516, 2020.

[57] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, 2016.

[58] K. Fukushima, "Neocognitron," *Scholarpedia*, vol. 2, no. 1, p. 1717, 2007.

[59] G. Hinton, A. Waibel *et al.*, "Phoneme recognition using time-delay neural network," *IEEE transactions on acoustics, speech, and signal processing*, vol. 37, pp. 328–339, 1989.

[60] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.

[61] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[62] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[63] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[64] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[65] S. H. Gheshlaghi, A. Ranjbar, A. A. Suratgar, M. B. Menhaj, and F. Faraji, "A superpixel segmentation based technique for multiple sclerosis lesion detection," *arXiv preprint arXiv:1907.03109*, 2019.

[66] L. Fang, X. Wang, and M. Wang, "Superpixel/voxel medical image segmentation algorithm based on the regional interlinked value," *Pattern Analysis and Applications*, vol. 24, no. 4, pp. 1685–1698, 2021.

[67] K. Wu, E. Otoo, and A. Shoshani, "Optimizing connected component labeling algorithms," in *Medical Imaging 2005: Image Processing*, vol. 5747. SPIE, 2005, pp. 1965–1976.

[68] K.-S. Fu and J. Mui, "A survey on image segmentation," *Pattern recognition*, vol. 13, no. 1, pp. 3–16, 1981.

[69] D. C. Ciresan, U. Meier, J. Masci, L. Maria Gambardella, and J. Schmidhuber, "Flexible, high performance convolutional neural networks for image classification," in *IJCAI proceedings-international joint conference on artificial intelligence*, vol. 22, no. 1. Barcelona, Spain:, 2011, p. 1237.

[70] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3642–3649.

[71] A. Menegola, M. Fornaciali, R. Pires, S. Avila, and E. Valle, "Towards automated melanoma screening: Exploring transfer learning schemes," 2016.

[72] O. Hadad, R. Bakalo, R. Ben-Ari, S. Hashoul, and G. Amit, "Classification of breast lesions using cross-modal deep learning," in *2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)*. IEEE, 2017, pp. 109–112.

[73] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[74] H. Chen, D. Ni, J. Qin, S. Li, X. Yang, T. Wang, and P. Heng, "Standard plane localization in fetal ultrasound via domain transferred deep neural networks," *IEEE journal of biomedical and health informatics*, vol. 19, 04 2015.

[75] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies protocols and applications," *IEEE Commun. Surveys Tuts., vol. 17, pp. 2347-2376, 4th Quart.*, 2015.

[76] A. I. L. Atzori and G. Morabito, "The internet of things: A survey," *Comput. Netw., vol. 54, no. 15, pp. 2787-2805, Oct.*, 2010.

[77] Q. H. X. Li and D. Wu, "Distributed large-scale co-simulation for iot-aided smart grid control," *IEEE Access, vol. 5, pp. 19951-19960*, 2017.

[78] I. E. Etim and J. Lota, "Power control in cognitive radios internet-of things (iot) for factories and industrial automation," *Proc. Annu. Conf. IEEE Ind. Electron. Soc., pp. 4701-4705*, 2016.

[79] P. S. S. Feng and S. Haykin, "Smart home: Cognitive interactive people-centric internet of things," *IEEE Commun. Mag., vol. 55, no. 2, pp. 34-39, Feb.*, 2017.

[80] R. K. S. H. Sutar and R. Suryavanshi, "Integration of smart phone and iot for development of smart public transportation system," *Proc. Int. Conf. Internet Things Appl., pp. 73-78, Jan.*, 2016.

[81] S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K.-S. Kwak, "The internet of things for health care: A comprehensive survey," *IEEE Access, vol. 3, pp. 678-708*, 2015.

[82] A. M. Alberti, G. D. Scarpioni, V. J. Magalhaes, A. Cerqueira, J. J. Rodrigues, and R. da Rosa Righi, "Advancing novagenesis architecture towards future internet of things," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 215–229, 2017.

[83] A. M. Elmisery, S. Rho, and D. Botvich, "A fog based middleware for automated compliance with oecd privacy principles in internet of healthcare things," *IEEE access*, vol. 4, pp. 8418–8441, 2017.

[84] Y. Oh and S. Yang, "Defining exergames and exergaming," *Proc. Meaningful Play, pp. 21-23*, 2010.

[85] D. H. et al., "Canadian stroke best practice recommendations: Stroke rehabilitation practice guidelines update 2015," *Int. J. Stroke, vol. 11, no. 4, pp. 459-484*, 2016.

[86] D. Johnson, S. Deterding, K.-A. Kuhn, A. Staneva, S. Stoyanov, and L. Hides, "Gamification for health and wellbeing: A systematic review of the literature," *Internet interventions*, vol. 6, pp. 89–106, 2016.

[87] J. Maule and L. Chestnutt, "Telemedicine in the 21st century: Opportunities for citizens, society and industry," *An International Space University Workshop, November 4–5, 1999, Hilton Hotel, Strasbourg, France*, 1999.

[88] T. G. Russell, "Physical rehabilitation using telemedicine," *Journal of telemedicine and telecare*, vol. 13, no. 5, pp. 217–220, 2007.

[89] B. Parmanto and A. Saptono, "Telerehabilitation: state-of-the-art from an informatics perspective," *International journal of telerehabilitation*, vol. 1, no. 1, p. 73, 2009.

[90] J. M. Winters, "Telerehabilitation research: emerging opportunities," *Annual Review of Biomedical Engineering*, vol. 4, no. 1, pp. 287–320, 2002.

[91] E. Ferrara, S. Nardotto, S. Ponte, and S. G. . Dellepiane, "Infrastructure for data management and user centered rehabilitation in rehab@home project," *In Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments (Vol. 21). ACM*, 2014.

[92] S. Mani, S. Sharma, B. Omar, A. Paungmali, and L. Joseph, "Validity and reliability of internet-based physiotherapy assessment for musculoskeletal disorders: a systematic review," *Journal of telemedicine and telecare*, vol. 23, no. 3, pp. 379–391, 2017.

[93] N. Gal, D. Andrei, D. I. Nemeş, E. Nădăşan, and V. Stoicu-Tivadar, "A kinect based intelligent e-rehabilitation system in physical therapy," in *Digital Healthcare Empowering Europeans*. IOS Press, 2015, pp. 489–493.

[94] H. Jagos, V. David, M. Haller, S. Kotzian, M. Hofmann, S. Schlossarek, K. Eichholzer, M. Winkler, M. Frohner, M. Reichel *et al.*, "A framework for (tele-) monitoring of the rehabilitation progress in stroke patients," *Applied Clinical Informatics*, vol. 6, no. 04, pp. 757–768, 2015.

[95] E. B. Larson, M. Feigon, P. Gagliardo, and A. Y. Dvorkin, "Virtual reality and cognitive rehabilitation: a review of current outcome research," *NeuroRehabilitation*, vol. 34, no. 4, pp. 759–772, 2014.

[96] E. A. Keshner and P. Weiss, "Introduction to the special issue from the proceedings of the 2006 international workshop on virtual reality in rehabilitation," *Journal of NeuroEngineering and Rehabilitation*, vol. 4, no. 1, p. 18, 2007.

[97] G. C. Burdea, D. Cioi, A. Kale, W. E. Janes, S. A. Ross, and J. R. Engsberg, "Robotics and gaming to improve ankle strength, motor control, and function in children with cerebral palsy—a case study series," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 2, pp. 165–173, 2012.

[98] C. Busch, C. Baumbach, D. Willemsen, O. Nee, T. Gorath, A. Hein, and T. Scheffold, "Supervised training with wireless monitoring of ecg, blood pressure and oxygen-saturation in cardiac patients," *Journal of telemedicine and telecare*, vol. 15, no. 3, pp. 112–114, 2009.

[99] B. Dinesen, L. K. Haesum, N. Soerensen, C. Nielsen, O. Grann, O. Hejlesen, E. Toft, and L. Ehlers, "Using preventive home monitoring to reduce hospital admission rates and reduce costs: a case study of telehealth among chronic obstructive pulmonary disease patients," *Journal of telemedicine and telecare*, vol. 18, no. 4, pp. 221–225, 2012.

[100] M. Trombini, F. Ferraro, M. Morando, G. Regesta, and S. Dellepiane, "A solution for the remote care of frail elderly individuals via exergames," *Sensors*, vol. 21, no. 8, p. 2719, 2021.

[101] M. Trombini, L. Vestito, M. Morando, L. Mori, C. Trompetto, F. Bandini, and S. Dellepiane, "Unilateral spatial neglect rehabilitation supported

by a digital solution: Two case-studies," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2020, pp. 3670–3675.

[102] F. Ferraro, M. Trombini, R. Truffelli, M. Simonini, and S. Dellepiane, "On the assessment of unilateral spatial neglect via digital tests," in *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2021, pp. 802–806.

[103] M. Morando, E. B. Bonotti, G. Giannarelli, S. Olivieri, S. Dellepiane, and F. Cecchi, "Monitoring home-based activity of stroke patients: A digital solution for visuo-spatial neglect evaluation," in *Converging Clinical and Engineering Research on Neurorehabilitation III: Proceedings of the 4th International Conference on NeuroRehabilitation (ICNR2018), October 16-20, 2018, Pisa, Italy 5*. Springer, 2019, pp. 696–701.

[104] F. Ferraro, M. Trombini, M. Morando, M. Doveri, G. Bianchi, and S. Dellepiane, "Exergames for systemic sclerosis rehabilitation: A pilot study," in *Advances in Computer Vision and Computational Biology: Proceedings from IPCV'20, HIMS'20, BIOCOMP'20, and BIOENG'20*. Springer, 2021, pp. 281–291.

[105] A. Rintala, S. Hakala, J. Paltamaa, A. Heinonen, J. Karvanen, and T. S. and, "Effectiveness of technology-based distance physical rehabilitation interventions on physical activity and walking in multiple sclerosis: a systematic review and meta-analysis of randomized controlled trials," *Disability and Rehabilitation*, vol. 40, no. 4, pp. 373–387, 2018.

[106] A. Rajavenkatanarayanan, V. Kanal, K. Tsiakas, D. Calderon, M. Papakostas, M. Abujelala, M. Galib, J. C. Ford, G. Wylie, and F. Makedon, "A survey of assistive technologies for assessment and rehabilitation of motor impairments in multiple sclerosis," *Multimodal Technologies and Interaction*, vol. 3, no. 1, p. 6, 2019.

[107] L. Lavorgna, F. Brigo, M. Moccia, L. Leocani, R. Lanzillo, M. Clerico, G. Abbadessa, K. Schmierer, C. Solaro, L. Prosperini *et al.*, "e-health and multiple sclerosis: An update," *Multiple Sclerosis Journal*, vol. 24, no. 13, pp. 1657–1664, 2018.

[108] M. J. Taylor and M. Griffin, "The use of gaming technology for rehabilitation in people with multiple sclerosis," *Multiple Sclerosis Journal*, vol. 21, no. 4, pp. 355–371, 2015.

[109] R. Negaresh, R. W. Motl, P. Zimmer, M. Mokhtarzade, and J. S. Baker, "Effects of exercise training on multiple sclerosis biomarkers of central nervous system and disease status: a systematic review of

intervention studies," *European Journal of Neurology*, 2019. [Online]. Available: https://doi.org/10.1111/ene.13929

[110] A. Manuli, M. G. Maggio, D. Tripoli, M. Gullì, A. Cannavò, G. La Rosa, F. Sciarrone, G. Avena, and R. S. Calabrò, "Patients' perspective and usability of innovation technology in a new rehabilitation pathway: An exploratory study in patients with multiple sclerosis," *Multiple Sclerosis and Related Disorders*, vol. 44, 2020. [Online]. Available: https://doi.org/10.1016/J.MSARD.2020.102312

[111] A. S. Nascimento, C. V. Fagundes, F. A. d. S. Mendes, and J. C. Leal, "Effectiveness of virtual reality rehabilitation in persons with multiple sclerosis: A systematic review and meta-analysis of randomized controlled trials," *Multiple Sclerosis and Related Disorders*, vol. 54, 2021. [Online]. Available: https://doi.org/10.1016/J.MSARD.2021.103128

[112] J. Wood and J. Finkelstein, "Telerehabilitation system to support multi-pronged exercise in patients with multiple sclerosis," in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2017, pp. 880–885.

[113] A. Achiron, R. Alonix, B. Croisile, M. Gurevich, G. Doniger, and F. Tarpin-Bernard, "Design and evaluation of computerized cognitive games to improve cognitive performance in multiple sclerosis: A correlation and validation study," in IEEE 7th International Conference on Serious Games and Applications for Health (SeGAH)*, IEEE 1-–7.*, 2019.

[114] M. Trombini, F. Ferraro, G. Iaconi, L. Vestito, F. Bandini, L. Mori, C. Trompetto, and S. Dellepiane, "A study protocol for occupational rehabilitation in multiple sclerosis," *Sensors*, vol. 21(24), 2021. [Online]. Available: https://doi.org/10.3390/S21248436

[115] F. Gholami, D. Trojan, J. Kövecses, W. Haddad, and B. Gholami, "A microsoft kinect-based point-of-care gait assessment framework for multiple sclerosis patients," *IEEE journal of biomedical and health informatics*, 2016.

[116] R. S. Calabrò, M. Russo, A. Naro, R. De Luca, A. Leo, P. Tomasello, F. Molonia, V. Dattola, A. Bramanti, and P. Bramanti, "Robotic gait training in multiple sclerosis rehabilitation: Can virtual reality make the difference? findings from a randomized controlled trial," *Journal of the neurological sciences*, vol. 377, pp. 25–30, 2017.

[117] M. G. Maggio, M. Russo, M. F. Cuzzola, M. Destro, G. La Rosa, F. Molonia, P. Bramanti, G. Lombardo, R. De Luca, and R. S. Calabrò, "Virtual

reality in multiple sclerosis rehabilitation: A review on cognitive and motor outcomes," *Journal of Clinical Neuroscience*, vol. 65, pp. 106–111, 2019.

[118] N. Grigoriadis, C. Bakirtzis, C. Politis, K. Danas, and C. Thuemmler, "Health 4.0: The case of multiple sclerosis," in *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*. IEEE, 2016, pp. 1–5.

[119] J. H. Burridge and A.-M. Hughes, "Potential for new technologies in clinical practice," *Current opinion in neurology*, vol. 23, no. 6, pp. 671–677, 2010.

[120] Y. Goverover, B. Z. Stern, A. Hurst, and J. DeLuca, "Internet-based technology in multiple sclerosis: Exploring perceived use and skills and actual performance." *Neuropsychology*, vol. 35, no. 1, p. 69, 2021.

[121] S. K. Asrani, H. Devarbhavi, J. Eaton, and P. S. Kamath, "Burden of liver diseases in the world," *Journal of Hepatology*, vol. 70, no. 1, pp. 151–171, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168827818323882

[122] A. M. Moon, A. G. Singal, and E. B. Tapper, "Contemporary epidemiology of chronic liver disease and cirrhosis," *Clinical Gastroenterology and Hepatology*, vol. 18, no. 12, pp. 2650–2666, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1542356519308493

[123] F. M. C. S. Group and P. Bedossa, "Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis c," *Hepatology*, vol. 20, no. 1, pp. 15–20, 1994.

[124] V. de Lédinghen and J. Vergniol, "Transient elastography (fibroscan)," *Gastroentérologie Clinique et Biologique*, vol. 32, no. 6, Supplement 1, pp. 58–67, 2008, diagnosis of liver fibrosis in 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0399832008739940

[125] A. P. Sarvazyan, O. V. Rudenko, S. D. Swanson, J. Fowlkes, and S. Y. Emelianov, "Shear wave elasticity imaging: a new ultrasonic technology of medical diagnostics," *Ultrasound in Medicine and Biology*, vol. 24, no. 9, pp. 1419–1435, 1998. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301562998001100

[126] L. Castera, X. Forns, and A. Alberti, "Non-invasive evaluation of liver fibrosis using transient elastography," *Journal of Hepatology*, vol. 48, no. 5, pp. 835–847, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168827808001232

[127] X.-W. Cui, M. Friedrich-Rust, C. De Molo, A. Ignee, D. Schreiber-Dietrich, and C. F. Dietrich, "Liver elastography, comments on efsumb elastography guidelines 2013," *World journal of gastroenterology*, vol. 19, no. 38, p. 6329—6347, October 2013. [Online]. Available: https://europepmc.org/articles/PMC3801303

[128] D. Han, Q. Liu, and W. Fan, "A new image classification method using cnn transfer learning and web data augmentation," *Expert Systems with Applications*, vol. 95, pp. 43–56, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417417307844

[129] P. M. Keune, S. Hansen, E. Weber, F. Zapf, J. Habich, J. Muenssinger, S. Wolf, M. Schönenberg, and P. Oschmann, "Exploring resting-state eeg brain oscillatory activity in relation to cognitive functioning in multiple sclerosis," *Clinical Neurophysiology*, vol. 128, no. 9, pp. 1746–1754, 2017.

[130] R. H. Benedict, M. P. Amato, J. DeLuca, and J. J. Geurts, "Cognitive impairment in multiple sclerosis: clinical management, mri, and therapeutic avenues," *The Lancet Neurology*, vol. 19, no. 10, pp. 860–871, 2020.

[131] N. D. Chiaravalloti and J. DeLuca, "Cognitive impairment in multiple sclerosis," *The Lancet Neurology*, vol. 7, no. 12, pp. 1139–1151, 2008.

[132] M. Moccia, R. Lanzillo, R. Palladino, K. C.-M. Chang, T. Costabile, C. Russo, A. De Rosa, A. Carotenuto, F. Saccà, G. T. Maniscalco *et al.*, "Cognitive impairment at diagnosis predicts 10-year multiple sclerosis progression," *Multiple Sclerosis Journal*, vol. 22, no. 5, pp. 659–667, 2016.

[133] S.-M. Yap, L. O'Donnell, Z. Togher, M. Dillon, N. McNicholas, N. Tubridy, M. Hutchinson, and C. McGuigan, "Safety monitoring of alemtuzumab therapy in active relapsing multiple sclerosis: necessary, manageable but resource-intensive," *Neurol*, vol. 88, no. 16 Supplement, pp. P5–401, 2017.

[134] G. W. Rebok, K. Ball, L. T. Guey, R. N. Jones, H.-Y. Kim, J. W. King, M. Marsiske, J. N. Morris, S. L. Tennstedt, F. W. Unverzagt *et al.*, "Ten-year effects of the advanced cognitive training for independent and vital elderly cognitive training trial on cognition and everyday functioning in older adults," *Journal of the American Geriatrics Society*, vol. 62, no. 1, pp. 16–24, 2014.

[135] S. I. Gallant *et al.*, "Perceptron-based learning algorithms," *IEEE Transactions on neural networks*, vol. 1, no. 2, pp. 179–191, 1990.

[136] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.

[137] A. Paszke, "Pytorch: An imperative style, high-performance deep learning library," *arXiv preprint arXiv:1912.01703*, 2019.

[138] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[139] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 675–678.

[140] F. K. Chollet, "https://github.com/fchollet/keras," (accessed on 20 May 2025).

[141] X. Meng, J. Bradley, B. Yavuz, E. Sparks, S. Venkataraman, D. Liu, J. Freeman, D. Tsai, M. Amde, S. Owen *et al.*, "Mllib: machine learning in apache spark (2015)," *arXiv preprint arXiv:1505.06807*, 2016.

[142] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[143] E. Strickland, "Ibm watson, heal thyself: How ibm overpromised and underdelivered on ai health care," *IEEE Spectrum*, vol. 56, no. 4, pp. 24–31, 2019.

[144] T. A. Stephenson, "An introduction to bayesian network theory and usage," *IDIAP-RR 00-03*, pp. 78–83, 2000.

[145] J.-G. Lee, S. Jun, Y.-W. Cho, H. Lee, G. B. Kim, J. B. Seo, and N. Kim, "Deep learning in medical imaging: general overview," *Korean journal of radiology*, vol. 18, no. 4, pp. 570–584, 2017.

[146] S. J. Oh, B. Schiele, and M. Fritz, "Towards reverse-engineering black-box neural networks," *Explainable AI: interpreting, explaining and visualizing deep learning*, pp. 121–144, 2019.

[147] V. Buhrmester, D. Münch, and M. Arens, "Analysis of explainers of black box deep neural networks for computer vision: A survey," *Machine Learning and Knowledge Extraction*, vol. 3, no. 4, pp. 966–989, 2021.

[148] S. Gambhir, S. K. Malik, and Y. Kumar, "Role of soft computing approaches in healthcare domain: a mini review," *Journal of medical systems*, vol. 40, pp. 1–20, 2016.

[149] A. Maseleno, N. Sabani, M. Huda, R. Ahmad, K. A. Jasmi, and B. Basiron, "Demystifying learning analytics in personalised learning," *International Journal of Engineering and Technology*, vol. 7, no. 3, pp. 1124–1129, 2018.

[150] S. S. Kute, A. Shreyas Madhav, S. Kumari, and S. Aswathy, "Machine learning–based disease diagnosis and prediction for e-healthcare system," *Advanced analytics and deep learning models*, pp. 127–147, 2022.

[151] R. R. Chandan, J. Singh, V. Ravi, B. D. Shivahare, T. J. Alahmadi, P. Singh, and M. Diwakar, "Reviewing the impact of machine learning on disease diagnosis and prognosis: A comprehensive analysis," *The Open Pain Journal*, vol. 17, no. 1, 2024.

[152] Y. Wang, C. Liu, W. Hu, L. Luo, D. Shi, J. Zhang *et al.*, "Economic evaluation for medical artificial intelligence: accuracy vs. cost-effectiveness in a diabetic retinopathy screening case. npj digital medicine. 2024; 7: 43."

[153] A. A. Lamir, S. Razzagzadeh, and Z. Rezaei, "A comprehensive machine learning framework for heart disease prediction: Performance evaluation and future perspectives," *arXiv preprint arXiv:2505.09969*, 2025.

[154] X. Liu, W. Zhang, Q. Zhang, L. Chen, T. Zeng, J. Zhang, J. Min, S. Tian, H. Zhang, H. Huang *et al.*, "Development and validation of a machine learning-augmented algorithm for diabetes screening in community and primary care settings: A population-based study," *Frontiers in Endocrinology*, vol. 13, p. 1043919, 2022.

[155] A. Vaid, K. Johnson, M. Badgeley, S. Somani, M. Bicak, I. Landi, A. Russak, S. Zhao, M. Levin, R. Freeman *et al.*, "Using deep-learning algorithms to simultaneously identify right and left ventricular dysfunction from the electrocardiogram. jacc cardiovasc imaging. 2022; 15: 395–410."

[156] A. C. J. Janssens, "Roc curves for clinical prediction models part 2. the roc plot: the picture that could be worth a 1000 words," *Journal of Clinical Epidemiology*, vol. 126, pp. 217–219, 2020.

[157] J. Q. Sheng, P. J.-H. Hu, X. Liu, T.-S. Huang, and Y. H. Chen, "Predictive analytics for care and management of patients with acute diseases: Deep learning–based method to predict crucial complication phenotypes," *Journal of medical Internet research*, vol. 23, no. 2, p. e18372, 2021.

[158] G. O. Oluwafemi, R. Faith, J. Badmus, and H. Luz, "Hybrid models combining machine learning and traditional epidemiological models," *International Journal of Circumpolar Health*, 2024.

[159] M. K. Raj, J. P. Malardhas, and I. Devapriya, "Machine learning approach to predict multiple diseases based on symptoms," in *2024 10th International Conference on Communication and Signal Processing (ICCSP)*. IEEE, 2024, pp. 1195–1199.

[160] A. Mohamed, M. Abdelrehim, and R. Al-Barazie, "Context matters in machine learning based disease prediction with insights from diverse clinical and symptom data," *Scientific Reports*, vol. 15, no. 1, p. 42669, 2025.

[161] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[162] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[163] S. K. Zhou, H. Greenspan, C. Davatzikos, J. S. Duncan, B. Van Ginneken, A. Madabhushi, J. L. Prince, D. Rueckert, and R. M. Summers, "A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 820–838, 2021.

[164] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nature medicine*, vol. 25, no. 1, pp. 44–56, 2019.

[165] M. P. Sendak, J. D'Arcy, S. Kashyap, M. Gao, M. Nichols, K. Corey, W. Ratliff, and S. Balu, "A path for translation of machine learning products into healthcare delivery," *EMJ Innov*, vol. 10, pp. 19–00 172, 2020.

[166] H.-P. Chan, L. M. Hadjiiski, and R. K. Samala, "Computer-aided diagnosis in the era of deep learning," *Medical physics*, vol. 47, no. 5, pp. e218–e227, 2020.

[167] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering*, vol. 19, no. 1, pp. 221–248, 2017.

[168] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 2010, pp. 249–256.

[169] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *International conference on machine learning*. PMLR, 2013, pp. 1139–1147.

[170] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 2011, pp. 315–323.

[171] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, and L. Bottou, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion." *Journal of machine learning research*, vol. 11, no. 12, 2010.

[172] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[173] L. Wan, M. Zeiler, S. Zhang, Y. Le Cun, and R. Fergus, "Regularization of neural networks using dropconnect," in *International conference on machine learning*. PMLR, 2013, pp. 1058–1066.

[174] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. pmlr, 2015, pp. 448–456.

[175] F. Kanavati and M. Tsuneki, "Partial transfusion: on the expressive influence of trainable batch norm parameters for transfer learning," in *Medical Imaging with Deep Learning*. PMLR, 2021, pp. 338–353.

[176] C. Khosla and B. S. Saini, "Enhancing performance of deep learning models with different data augmentation techniques: A survey," in *2020 International Conference on Intelligent Engineering and Management (ICIEM)*, 2020, pp. 79–85.

[177] S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman *et al.*, "The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans," *Medical physics*, vol. 38, no. 2, pp. 915–931, 2011.

[178] H. J. Aerts, E. R. Velazquez, R. T. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature communications*, vol. 5, no. 1, p. 4006, 2014.

[179] M. Roberts, D. Driggs, M. Thorpe, J. Gilbey, M. Yeung, S. Ursprung, A. I. Aviles-Rivero, C. Etmann, C. McCague, L. Beer *et al.*, "Common pitfalls and recommendations for using machine learning to detect and

prognosticate for covid-19 using chest radiographs and ct scans," *Nature Machine Intelligence*, vol. 3, no. 3, pp. 199–217, 2021.

[180] G. Varoquaux and V. Cheplygina, "Machine learning for medical imaging: methodological failures and recommendations for the future," *NPJ digital medicine*, vol. 5, no. 1, p. 48, 2022.

[181] Y. Yu and S. Acton, "Speckle reducing anisotropic diffusion," *IEEE Transactions on Image Processing*, vol. 11, no. 11, pp. 1260–1270, 2002.

[182] H. Choi and J. Jeong, "Speckle noise reduction in ultrasound images using srad and guided filter," in *2018 International Workshop on Advanced Image Technology (IWAIT)*, 2018, pp. 1–4.

[183] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[184] J. Weston and C. Watkins, "Multi-class support vector machines," Citeseer, Tech. Rep., 1998.

[185] S. Salcedo-Sanz, J. L. Rojo-Álvarez, M. Martínez-Ramón, and G. Camps-Valls, "Support vector machines in engineering: an overview," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 4, no. 3, pp. 234–267, 2014.

[186] E. Lachat, H. Macher, T. Landes, and P. G. Assessment, "Calibration of a rgb-d camera (kinect v2 sensor) towards a potential use for close-range 3d modeling., 2015, 7," *DOI: https://doi. org/10.3390/rs71013070*, pp. 13 070–13 097, 2015.

[187] D. J. Geerse, B. H. Coolen, and M. Roerdink, "Kinematic validation of a multi-kinect v2 instrumented 10-meter walkway for quantitative gait assessments," *PloS one*, vol. 10, no. 10, p. e0139913, 2015.

[188] S. T. Pöhlmann, E. F. Harkness, C. J. Taylor, and S. M. Astley, "Evaluation of kinect 3d sensor for healthcare imaging," *Journal of medical and biological engineering*, vol. 36, pp. 857–870, 2016.

[189] K. Otte, B. Kayser, S. Mansow-Model, J. Verrel, F. Paul, A. U. Brandt, and T. Schmitz-Hübsch, "Accuracy and reliability of the kinect version 2 for clinical measurement of motor function," *PloS one*, vol. 11, no. 11, p. e0166532, 2016.

[190] G. M. Moreira, L. H. Giovanini, M. P. de Castro, G. N. Nogueira, T. C. Boumer, and E. F. Manffra, "Filtering motion signals from microsoft kinect® in the context of stroke rehabilitation," *Research on Biomedical Engineering*, vol. 35, pp. 265–270, 2019.

[191] Z. Niu, K. Lu, J. Xue, X. Qin, J. Wang, and L. Shao, "From method to application: A review of deep 3d human motion capture," *IEEE Transactions on Circuits and Systems for Video Technology*, 2024.

[192] J. F. Gauss, C. Brandin, A. Heberle, and W. Löwe, "Smoothing skeleton avatar visualizations using signal processing technology," *SN Computer Science*, vol. 2, no. 6, p. 429, 2021.

[193] D. A. Winter, *Biomechanics and motor control of human movement*. John wiley and sons, 2009.

[194] V. K. Yalavarthi, J. Burchert, and L. Schmidt-Thieme, "Tripletformer for probabilistic interpolation of asynchronous time series," *CoRR*, 2022.

[195] A. Papoulis, *Random variables and stochastic processes*. McGraw Hill, 1965.

[196] S. Dowdy, S. Wearden, and D. Chilko, *Statistics for research*. John Wiley and Sons, 2011.

[197] B. G. Amidan, T. A. Ferryman, and S. K. Cooley, "Data outlier detection using the chebyshev theorem," in *2005 IEEE Aerospace Conference*. IEEE, 2005, pp. 3814–3819.

[198] S. Butterworth *et al.*, "On the theory of filter amplifiers," *Wireless Engineer*, vol. 7, no. 6, pp. 536–541, 1930.

[199] Alejo2083, "Butterworth filter," *https://en.wikipedia.org/wiki/Butterworth filter#/media/File: Butterworth filter bode plot.svg [208] D. A. Neumann, Kinesiology of the musculoskeletal*, 2006.

[200] L. Vestito, F. Ferraro, G. Iaconi, G. Genesio, F. Bandini, L. Mori, C. Trompetto, and S. Dellepiane, "Storms: A pilot feasibility study for occupational telerehabilitation in multiple sclerosis," *Sensors (Basel, Switzerland)*, vol. 24, no. 19, p. 6470, 2024.

[201] D. A. Neumann, *Kinesiology of the Musculoskeletal System-E-Book*. Elsevier Health Sciences, 2016.

[202] A. M. Oosterwijk, M. K. Nieuwenhuis, C. P. van der Schans, and L. J. Mouton, "Shoulder and elbow range of motion for the performance of activities of daily living: A systematic review," *Physiotherapy theory and practice*, vol. 34, no. 7, pp. 505–528, 2018.

[203] P. Bedossa and T. Poynard, "An algorithm for the grading of activity in chronic hepatitis c," *Hepatology*, vol. 24, 1996. [Online]. Available: https://api.semanticscholar.org/CorpusID:44385981

[204] V. Dutt and J. Greenleaf, "Adaptive speckle reduction filter for log-compressed b-scan images," *IEEE Transactions on Medical Imaging*, vol. 15, no. 6, pp. 802–813, 1996.

[205] K. Krissian, R. Kikinis, C.-F. Westin, and K. Vosburgh, "Speckle-constrained filtering of ultrasound images," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, 2005, pp. 547–552 vol. 2.

[206] A. Gotra, L. Sivakumaran, and G. e. a. Chartrand, "Liver segmentation: indications, techniques and future directions." *Insights Imaging 8, 377–392 (2017)*, 2017. [Online]. Available: https://doi.org/10.1007/s13244-017-0558-1

[207] N. S. Punn, B. Patel, and I. Banerjee, "Liver fibrosis classification from ultrasound using machine learning: a systematic literature review," *Abdominal Radiology*, vol. 49, no. 1, pp. 69–80, 2024.

[208] D. Meng, L. Zhang, G. Cao, W. Cao, G. Zhang, and B. Hu, "Liver fibrosis classification based on transfer learning and fcnet for ultrasound images," *IEEE Access*, vol. 5, pp. 5804–5810, 2017.

[209] Z. Liu, B. Huang, H. Wen, Z. Lu, Q. Huang, M. Jiang, C. Dong, Y. Liu, X. Chen, and H. Lin, "Automatic diagnosis of significant liver fibrosis from ultrasound b-mode images using a handcrafted-feature-assisted deep convolutional neural network," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 10, pp. 4938–4949, 2023.

[210] J. Lee, I. Joo, and T. e. a. Kang, "Deep learning with ultrasonography: automated classification of liver fibrosis using a deep convolutional neural network." *Springer*, vol. 30, p. 1264–1273, 2020.

[211] D. Silva, T. G. Gonçalves, and A. R. C. da Rocha, "A requirements engineering process for iot systems," in *Proceedings of the XVIII Brazilian symposium on software quality*, 2019, pp. 204–209.

[212] S. Dellepiane, L. Vestito, L. Mori, C. Trompetto, F. Bandini, G. Iaconi, F. Ferraro, and G. Genesio, "Storms: Occupational telerehabilitation in multiple sclerosis," 2024.

[213] L. Thabane, J. Ma, R. Chu, J. Cheng, A. Ismaila, L. P. Rios, R. Robson, M. Thabane, L. Giangregorio, and C. H. Goldsmith, "A tutorial on pilot studies: the what, why and how," *BMC medical research methodology*, vol. 10, pp. 1–10, 2010.

[214] D. I. Stoia, C. Vigaru, and L. Rusu, "Relative and absolute angles computed from pathologic gait data," in *Acoustics and Vibration of Mechanical Structures—AVMS-2017: Proceedings of the 14th AVMS Conference, Timisoara, Romania, May 25–26, 2017.* Springer, 2017, pp. 201–206.