Fuzzy Sets and Systems 523 (2026) 109638

Contents lists available at ScienceDirect

Fuzzy Sets and Systems

journal homepage: www.elsevier.com/locate/fss

Modeling uncertainty with interval-valued type-2 fuzzy sets: Application to anomalous sound event detection

Zied Mnasri [a,c,*], Stefano Rovetta [b], Francesco Masulli [b]

[a] Faculty of Engineering and Digital Technology, University of Bradford, Richmond Road, Bradford, BD7 1DP, UK
[b] DIBRIS, University of Genoa, Via Dodecaneso 35, Genova, 16145, Italy
[c] Ecole Nationale d'Ingénieurs de Tunis, University Tunis El Manar, BP 37, Le Belvedere, 1002, Tunis, Tunisia

ARTICLE INFO

Keywords:
Sound event detection
Anomaly detection
Modeling uncertainty
Variational autoencoder
Interval-valued type-2 fuzzy sets
Interval comparison
Membership function

ABSTRACT

Since a few years, audio signal processing has been focused on detecting audio events in general, and defining anomalous/outlier sounds in particular. The application of such an anomaly detection problem to audio surveillance systems is made possible thanks to the advances of anomaly detection techniques, in particular for highly unbalanced data. However, outdoor audio signals are characterized by a high degree of uncertainty, since there is no way to model each category of sound, whether normal or anomalous, in presence of background noise. Thus, this paper proposes a rare/anomalous sound event detection method for road traffic surveillance, which aims at detecting hazardous events, such as car accidents, in presence of traffic noise. To model uncertainty for anomaly detection, the suggested method combines deep reconstruction techniques, interval-valued type-2 fuzzy sets and interval comparison methods. First, the reconstruction error of the input audio segment is yielded by a deep variational autoencoder which is trained on normal data only. Based on this reconstruction error, a fuzzy membership function with pessimistic/lower and optimistic/upper components is calculated. Next, a probabilistic interval comparison method is used to compute the membership score, and thus to evaluate the interval-valued fuzzy sets. Finally, defuzzification is used to classify events as normal or anomalous. During this process, several types of linear or nonlinear membership functions are utilized to model uncertainty with respect to the input, *i.e.*, the VAE reconstruction error, or to the output, *i.e.*, the value of the primary membership. According to the results obtained, the proposed method outperforms the state-of-the-art one-class SVM for anomaly detection and the baseline VAE error thresholding method, when specific parameters are carefully set, such as the weights of the anomalous/normal subsets and the lower/upper bounds of the membership function's components. Furthermore, the proposed linear and nonlinear membership functions succeed to improve modeling uncertainty in audio signals by interval-valued type-2 fuzzy sets, with regard to: a) the input, *i.e.*, the VAE reconstruction error, and b) the primary membership value, respectively.

* Corresponding author at: Faculty of Engineering and Digital Technology, University of Bradford, BP7 1DP, Bradford, UK.
E-mail addresses: z.mnasri@bradford.ac.uk (Z. Mnasri), stefano.rovetta@unige.it (S. Rovetta), francesco.masulli@unige.it (F. Masulli).

https://doi.org/10.1016/j.fss.2025.109638
Received 25 November 2023; Received in revised form 18 August 2025; Accepted 10 October 2025

Available online 15 October 2025

1. Introduction

1.1. General context: anomaly detection in audio signals

As advanced in [76], anomaly/outlierness/novelty can be defined in a variety of ways by virtue of: a) its scarcity, as anomalous/novel/outlier events occur less frequently than normal events; b) its characteristics, as anomalous/novel/outlier events should have different characteristics than normal events; and c) its meaning, as such events should carry a specific and distinct meaning from normal events.

According to the user's objective, this problem can be formalized in two ways: as a classification task for all perceived events, or as detection of only anomalous/outlier/novel events. However, it should be noted that the main challenge in both tasks consists in coping with the rarity of anomalous events, which causes a strong imbalance between classes. As a result, generative modeling, such as HMM-GMM, has traditionally been used to approach such an anomaly detection problem [56]. For instance, in the context of road audio surveillance, *anomalous* events are mainly car accidents (collision, glass breaking, hard braking, etc.), whereas the class *normal* encompasses all other events that may occur on the road, such as car noise, pedestrians voices, traffic sound and any other non-hazardous event such as tire skidding. Nonetheless, the task of anomalous sound event detection (SED) for road audio surveillance has to cope with two major challenges: The first is background noise, which completely or partially obscures all events; the second is the rarity of *interesting* events, such as car accidents, which are far less common than normal events.

1.2. Motivation: the uncertainty problem

The goal of this research is to detect unusual/rare, or in terms of computational intelligence, anomalous/outlier/novel events in audio signals recorded by roadside microphones. Thus, anomaly/outlierness/novelty consists primarily of dangerous events such as an automobile collision, tire skidding, forceful braking, and so on. It is also evident that the sounds associated with such events are insignificant in comparison to those associated with routine activities, such as traffic, people, or simply background street noise. Such rarity necessitates specialized anomaly detection processing, as a standard classifier based on generative or discriminative approaches may not uncover enough anomalous data to train on.

The topic of uncertainty in road traffic audio data, which has been recently investigated in [51,67], demonstrates that training a reconstruction neural network to model audio data is not enough to provide an efficient tool for anomaly identification. Therefore, in this work, fuzzy sets are assumed to be capable of modeling aberrant audio data, and in particular, an approach based on interval-valued fuzzy sets is proposed to provide a solution to the discussed problem.

As for the methods selected for this particular application, the choice of interval-valued type-2 fuzzy sets for this task is motivated by the following considerations in terms of: a) *Solution*, as we tried to select a method suitable to the addressed problem, *i.e.*, modeling uncertainty in audio data; b) *Feasibility*, since interval-valued type-2 FS are very appropriate to the context of the studied task (weakly/semi/un-supervised anomaly detection); c) *Novelty*: To our knowledge, interval-valued type-2 FS has not been used in other research works in audio processing, particularly in Anomalous Sound Detection —except the preliminary work of the authors [52]—; d) Accountability: The results confirm the improvement brought by using interval-valued type-2 FS, as will be shown hereafter.

1.3. Scope, methodology and contribution

The complexity of such ill-defined classes, especially in the presence of background noise, demonstrates that membership in any class is influenced by a degree of uncertainty, as illustrated in Fig. 1. This figure shows that the cluster of normal data (Class 0) contains several samples of the anomalous class (Class 1). This means that the input data share common characteristics, even though they belong to different clusters. This is particularly curious, as the analyzed features, *i.e.*, MFCC, are known to be highly discriminating in the domain of audio signals and have therefore been widely used to solve audio and speech classification problems since decades [16,63,78]. However, in our case, they fail to provide distinct clusters, which means that audio signals from different clusters contain common features. This can be explained by the presence of similar context (background noise) in all audio clips, whether they contain normal or anomalous events.

To deal with such uncertainty, which is primarily due to the lack of clear modeling of audio signals, interval-valued type-2 fuzzy sets [47] offer an alternative to crisp clustering or even to type-1 fuzzy sets. In this work, interval-valued fuzzy memberships are used to model classes for two main reasons: a) the inherent simplicity and popularity of fuzzy sets as a clustering tool, and b) the ability to reduce the need for arbitrary modeling decisions about the membership itself with respect to general type-2 fuzzy sets, which require to use some specified membership function as opposed to just two interval extremes.

The final decision is made by comparing the interval-valued memberships to the different classes using a conventional interval comparison metric known as degree of preference [73]. This procedure allows for the determination of the final class, *i.e.*, *normal* or *anomalous*, without discarding the information about uncertainty expressed by the 2-component fuzzy membership.

Thus, we use the workflow shown in Fig. 2 to address the issue of uncertainty modeling in audio data for anomalous event detection in this work.

Firstly, the input audio chunk is processed to extract the spectrograms from a sequence of overlapping frames. The input feature vectors, which are Mel-Frequency Cepstral Coefficients (MFCC) with their first and second derivatives (Δ-MFCC and Δ-Δ-MFCC), are provided by the extracted spectrograms.

Afterwards, using the input feature vectors, three different approaches are explored in order to identify anomalous events:

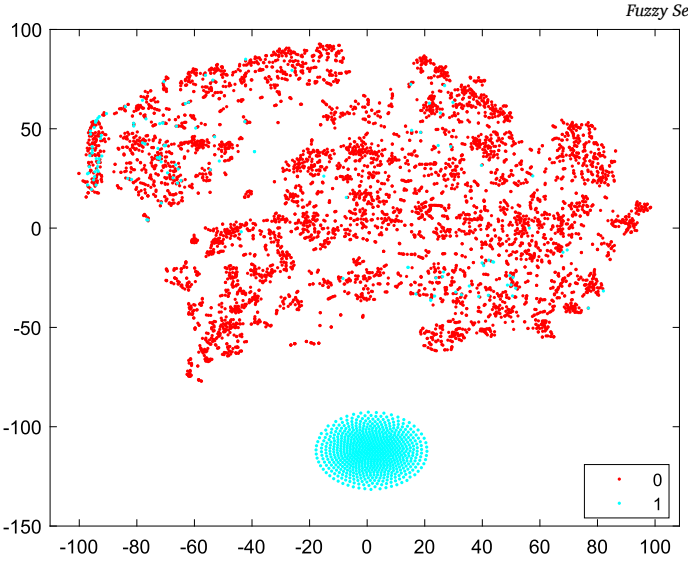


Fig. 1. Illustration of the anomalous sound event detection problem with t-SNE distribution of dimension-reduced input features (MFCC, Δ-MFCC and Δ-Δ-MFCC) for audio event classes, *i.e.*, *normal* (0) and *anomalous* (1): Despite the presence of two separate clusters for both classes, several *anomalous* samples are mapped in the cluster of *normal* ones.

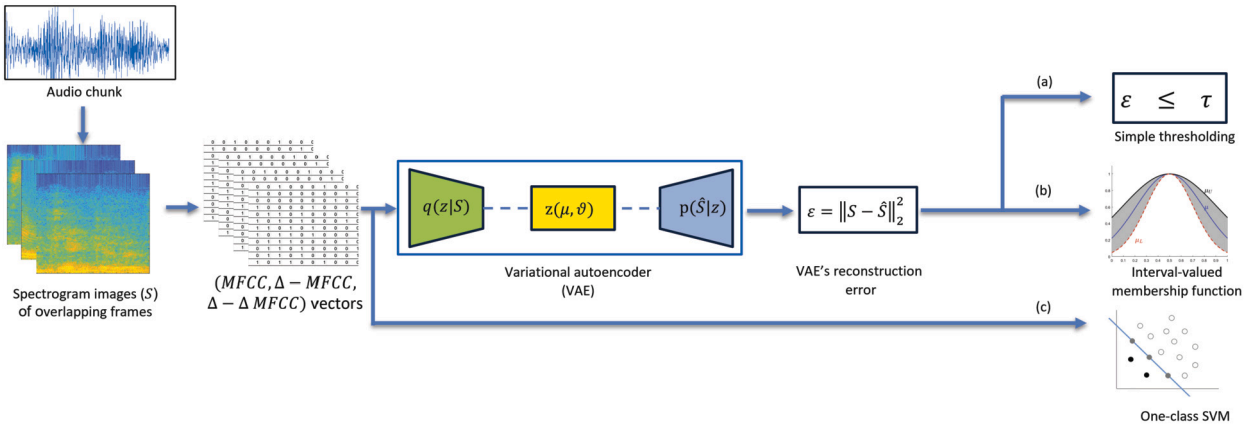


Fig. 2. Methodology used for anomalous sound event detection from audio chucks: (a) Baseline method, using simple thresholding of the VAE's reconstruction error (ϵ), (b) Proposed method, based on computing the interval-valued membership function of (ϵ), (c) Benchmarking method (for evaluation purposes) based on one-class support vector machines (OC-SVM).

a A baseline approach using a variational autoencoder (VAE) to reconstruct the input feature vector. The decision about *anomaly vs. normality* is subsequently made by comparing the reconstruction error to a predetermined threshold;
b A proposed approach that computes an interval-valued fuzzy membership function by using the VAE's reconstruction error. This approach involves testing a variety of membership functions as well as proposing a number of configurations to represent the uncertainty of the suggested membership function. This approach is the paper's primary contribution;
c A benchmarking approach that classifies input feature vectors into two categories, *i.e.*, *normal* and *anomalous*, using one-class support vector machines (OC-SVM). In this respect, OC-SVM has been chosen as the benchmark, because it is the state-of-the-art anomaly detection technique based on one-class classification, *i.e.*, using only *normal* data for training, as in our study.

The remainder of this paper is organized as follows: Section 2 reviews the related work, including the main anomalous SED state-of-the-art methods; Section 3 presents the proposed approach as well as the benchmarking methods; and Section 4 details the experimental protocol and analyzes the obtained results. Finally, the work is summarized and commented in Section 6.

2. Related work

Anomalous SED, as an anomaly detection problem, has been approached with different methods. This may seem heterogeneous, but the diversity of methods used is mainly due to the adaptation of the anomalous SED problem to the evolution of computational modeling of anomaly detection. Thus, like most signal recognition/identification problems, SED has traditionally been performed

Table 1
Methods and models for anomalous Sound Event Detection.

Background	Method	Model
Generative methods	GMM	GMM-based probabilistic novelty detection [56] Context-dependent GMM [24]
	HMM	HMM-based probabilistic novelty detection [56]
Discriminative methods	OC-SVM	OC-SVM for anomalous SED [4] Ensemble-based OC-SVM for anomalous SED [15] Support vector data description and plane-based OC-SVM [13] Ensemble-based MLP-OC-SVM outlier detection model [66] One-class condition random fields for sequential anomalous SED [77]
Supervised learning	RNN and CNN	RNN-BLSTM classifier based on standard features [38] MLP-CNN hierarchic multi-scaled model [81] 1D-ConvNets model [39] Spectrogram-image-based CNN classifier [12] Region-based convolutional RNN (R-CRNN) model [29] Dilated-gated R-CRNN with a discriminative penalty loss function [58] RNN-LSTM classifier with different pooling functions [30]
	Weighted and/or multitask learning	Acoustic event type and frame position detection using multitask learning [57] DNN-CNN model with a weighted and multitask loss function [58] Multitask multilabel modelbased on CRNN [59] Multitask learning for joint SED and ASC [27] Sample weight initialization for ensemble MLP-OC-SVM outlier detection [49]
Un/Semi/ Weakly-supervised learning	GAN learning	Gaussian-mixture GAN model [10]
	Few-shot learning	Few-shot learning model with ensured true positive rate [34] Few-shot model with metric learning [74]) Attention network for one-shot learning [35]
	Autoregressive neural networks	Complementary set variational autoencoder [33] WaveNet-based anomalous SED [68] Group-masked autoencoder [18] Denoising autoencoder [44] Deep autoencoder based on GMM with hyperparameters [61]

using generative approaches such as Gaussian mixture models (GMM) and Hidden Markov models (HMM), which take into account contextual and temporal cues of the signal. For the past few years, however, the processing of anomalous SEDs has mainly made use of discriminative methods, such as one-class support vector machines (OC-SVMs) and deep neural networks (DNNs). Alternatively, unsupervised clustering has also been explored to provide models for large sets of unlabeled and/or weakly labeled audio signals. Table 1 presents a taxonomy of state-of-the-art methods developed so far for anomalous SED. The main models are described in detail below.

2.1. Generative models

Ntalampiras et al. [56] used three generative approaches for probabilistic novelty detection in real-world acoustic surveillance, namely a universal GMM model, a universal HMM model, and a GMM clustering model. In addition, they used a maximum a posteriori adaptation model (MAP), originally proposed by [65], for updating the parameters of the Gaussian components. In the same line, Heittola et al. suggested a context-dependent SED in [24]. This method is divided into two stages: i) automated context recognition, where GMM is used to model contexts, and ii) sound event detection, using a 3-state left-to-right HMM to describe sound events.

2.2. Discriminative models

Discriminative approaches, primarily based on support vector machines (SVM) and neural networks (NN), have also been used to perform SED, either for sound classification or anomalous SED. Aurino et al. [4] built an OC-SVM model to identify burst-like anomalous sound events such as gunshots, shattered glasses, and screaming voices. The features are taken from the audio signal's time and frequency representations and supplied into the OC-SVM classifier.

One-Class Support Vector Machines (OC-SVM). As reported by [55], OC-SVM has been successfully applied as anomaly/outlier/novelty detection method for different data types, such as time series [42] and data streams [64]. Anomaly detection using OC-SVM was utilized in several applications, such as healthcare, e.g. seizure prediction in patients based on change detection in EEG time series [17], predictive maintenance, e.g. vibration spectra monitoring of jet engines [22], and network security, e.g. intrusion detection into Microsoft Windows operating systems [25], as reported in [60].

Ensemble OC-SVM. OC-SVM are also used in ensemble-architecture to model anomalous SED. For instance, Foggia et al. [15] proposed a two-layer approach based on low-level audio feature extraction, and high level bag-of-words approach to classify events into short and sustained ones. Then, an ensemble SVM is used for event classification. Other OC-SVM based models for anomalous SED are mentioned in Table 1. Recently, an ensemble model composed of an OC-SVM parallel to a multilayer perceptron (MLP) has been used by the authors of this paper to calculate the resulting anomaly score for audio events [66]. In this approach, the OC-SVM yields a primary anomaly score, whereas the MLP probability output indicates the event class score. Finally, the product of both scores is thresholded to indicate whether the event classified by the MLP is really an outlier.

Isolation forest. Isolation forest has been qualified as an efficient method for anomaly detection [31]. It is based on the binary search trees that find out the partition of multidimensional dataset containing a particular sample and estimate its anomaly score. The isolation forest method developed in [40] comprises two main stages, (a) *training*, where binary search trees are built on a basis of samples of the dataset and (b) *scoring*, computed by searching these trees using all the records of the dataset as arguments.

2.3. Supervised learning-based models

If the training set is fully labeled, anomalous SED may be handled as a classification problem. Thus, several approaches and models have been created employing labeled datasets. Deep learning approaches, such as recurrent and convolutional neural networks, and multitask learning, have received a particular attention. Several supervised learning approaches and models have been developed for anomalous SED, particularly in the following Detection and Classification of Acoustic Scene and Events (DCASE) challenges: DCASE'2016-Task3 (real-life SED) [82], DCASE'2017-task2 (rare SED) [83] and DCASE'2019-Task4 (SED in home contexts) [43]. In those challenges (and also in other events and publications), several innovative methods based on convolutional and recurrent neural networks were proposed, such as region-based convolutional recurrent neural network (R-CRNN) technique [29], dilated-gated CNN (DG-CNN) [23] and multitask learning in [58,90].

2.4. Unsupervised learning-based models

Because the rarity of anomalous data is one of the main challenges in anomalous SED, unsupervised learning has been chosen over supervised approaches to deal with the absence of labeled and evenly balanced amounts of data for each class. Autoregressive neural networks, such as autoencoders, and metric learning approaches, particularly density estimation, are potentially well suited to dealing with such an issue.

Recently, Wei et al. [89] suggested a reconstruction autoencoder to compute the anomaly score using metric learning for both types of autoencoders evaluated, namely feedforward and variational autoencoders. More recently, Purohit et al. [61] presented a deep autoencoder trained to learn GMM distributions to detect anomalies in audio signals. Besides, unsupervised density estimation has also been used to describe the underlying probability density of independent and identically distributed data sets in [36,53]. In [6], a GMM model was trained on purely normal data and the Kullback-Leibler (KL) divergence between the input and the output was estimated. This strategy was refined in [7] by removing a fourth of the mixture model's most divergent Gaussian distributions to increase KL divergence performance.

At last, it is worth noting that the authors have recently published a comprehensive review about anomalous SED, including the state-of-the-art methods, datasets and applications [50]. It is also interesting to notice that fuzzy clustering is rarely used is sound event detection, and particularly in anomalous SED.

3. Methods

Since one should have no *a priori* knowledge about anomalous sounds, it was decided in this work to deal with the addressed topic of anomalous sound event detection as a one-class classification problem. Such an approach necessitates developing models on normal data only. Then, a set of criteria should be set to distinguish anomalous samples from normal ones, during the test phase.[1]

3.1. State-of-the-art method: one-class SVM

The choice was made to start with a brief description of the state-of-the-art method, *i.e.*, One-Class Support Vector Machines (OC-SVM), because its is the one that inspired the idea of treating anomalous and normal data samples separately [72]. OC-SVM is a variant of the SVM algorithm that aims at estimating a function with positive values on a half-space and negative values on its complement. OC-SVM, in general, divides the input space into normal data and outliers. However, training is carried out on normal data only. The final decision is made using the sign function $g(x)$, which is calculated as follows:

$$g(x) = \text{sgn}(w^T \phi(x) - \rho), \tag{1}$$

where ϕ denotes the Gaussian kernel, w the orthogonal vector to the separating hyperplane, and ρ a bias term. If this function is positive for each sample, the sample is referred to as normal; otherwise, the sample is considered as an outlier. In this paper, OC-SVM is primarily used to compare performance with the proposed methods for one-class classification. A detailed description of the

[1] The code of the proposed methods is available at https://github.com/zied-mnasri/Uncertainty-modeling-anomalous-SED.

one-class classification problem formulation with SVM can be found in [72]. However, it should be noted that the aforementioned formulation of OC-SVM ignores data uncertainty. Some novel OC-SVM formulations that account for uncertainty have recently been proposed in [80,91].

Regarding the parametrization of OC-SVM, the main parameters to set are: a) the Kernel function, *e.g. Guassian*, *RBF*, *Polynomial*, … *etc.*; b) The *kernel coefficient* (γ), which is usually taken as the inverse of the number of features; and c) ν which is an upper bound on the fraction of training errors and a lower bound of the fraction of support vectors, that has a value $0 < \nu \leq 1$.

However, it should be noted that the *a priori* division of the training set into normal data and outliers, and using the normal data ratio as a parameter in the training process may have a deep effect on anomaly detection. Also, [14] showed that high dimensionality kernel methods may have a negative impact on the learning process, due to the high variance of meaningless, yet numerous, noisy samples.

3.2. Baseline method: normal event-based variational autoencoders

The autoencoder is a neural network developed with the goal of estimating the identity function. It is a popular unsupervised learning technique, used either for extracting features from unlabeled data or for reconstructing input data. To accomplish this, the autoencoder optimizes the weights to minimize the mean square error between the given input and the obtained output. The value of a hidden layer is then used as an encoded representation of the input. This latter layer, known also as the code layer, provides an encoded representation of the input. Such an encoded representation can be used either as latent features of the input, *i.e.*, feature extraction, or as an input to the second half of the autoencoder, *i.e.*, the decoder, to yield a reconstructed image of the input [54].

The autoencoder has been proposed as a semi/weakly-supervised anomaly detection method in several works, especially dealing with image processing. Thus, training is performed using *normal* data only, then the reconstruction error is compared to a predefined threshold to decide about outlierness (cf (4). The rationale behind such a reasoning lies in the assumption that only *normal* samples should apply to the trained model, hence any anomalous/outlier pattern is more likely to present a reconstruction error higher than that of *normal* ones, thus than the threshold. This method has been used for anomaly detection in several works, *e.g.* for cyber-security [11], space telemetry [70], medical imaging [75] and video processing [93]. However, it should be noted that the threshold used to asses anomaly in (4) is mostly chosen manually using (6).

3.2.1. Variational autoencoder

The variational autoencoder (VAE) is a reconstruction neural network as well, because it learns a compressed representation of the input to reconstruct the output. However, the code layer of VAE stores the parameters of a probability distribution, such as mean and variance, in a latent space, reflecting the input. The decoder then employs the probability distribution to build an approximation of the input data. As a result, the encoder approximates the identity function's probability distribution. A detailed description of the VAE learning theory can be found in [41].

In this work, the VAE is chosen as the baseline method instead of simple or deep autoencoders for two main reasons. First, the proposed method aims to improve its performance for anomaly detection by using fuzzy clustering instead of simple thresholding, as will be details in the next sections; secondly, the VAE approximates a probability distribution in its code layer, which makes the output reconstruction taking into account the uncertainty of the input.

This form of uncertainty is expressed by the VAE's loss function, which differs from that of the simple and deep autoencoders by the addition of a Kullback-Leibler (KL) distance, which measures the difference between the input probability density and the approximated one. In fact, The VAE attempts to find $P(Z)$ using the *a priori* distribution $P(Z|X)$, which is obtained using variational inference by minimizing the loss given by:

$$\log P(X) = -\{||X - \hat{X}||_2 + \mathrm{KL}(Q(Z|X)||P(Z))\}, \tag{2}$$

where $||.||_2$ denotes the L^2 norm and KL the Kullback-Leibler divergence, given by:

$$\mathrm{KL}(A||B) = \int p_A(x) \log \frac{p_B(x)}{p_A(x)} dx. \tag{3}$$

As a result, the purpose of VAE is to train the encoder output $Q(Z|X)$ so that the divergence between $Q(Z|X)$ and $P(Z|X)$ is as small as possible. For example, if $P(Z)$ is a Gaussian distribution, the encoder generates the mean and variance, which are then used to calculate $P(Z|X)$.

3.2.2. Anomaly detection using VAE

As mentioned, the baseline VAE is trained on normal data, *i.e.*, audio clips that contain only non-hazardous events, in a manner similar to OC-SVM. The output RMSE error is then calculated for the aforementioned terms, as follows:

$$\epsilon = \sqrt{\frac{\sum_{k=1}^{m} (x_k - \tilde{x_k})^2}{m}}, \tag{4}$$

where $X = \{x_k\}_{k=1,\ldots,m}$ and $\hat{X} = \{\hat{x}_k\}_{k=1,\ldots,m}$ are the input and the output feature vectors to and from the VAE, respectively. Finally, the comparison of the VAE output error (ϵ) to a predetermined threshold (τ_0) reveals whether the input sample (i) is normal or anomalous, as follows:

$$Event(i) = \begin{cases} normal & \text{if } 0 \le \epsilon \le \tau_0, \\ anomalous & \text{if } \epsilon > \tau_0, \end{cases} \tag{5}$$

where the threshold (τ_0) is calculated as

$$\tau_0 = \overline{\epsilon_{tr}} = \frac{1}{M} \times \sum_{m=1}^{M} \epsilon_{st_m}, \tag{6}$$

where $\epsilon_{st_m} = (\epsilon - \epsilon_{min})/(\epsilon_{max} - \epsilon_{min})$ is the standardized reconstruction error of the autoencoder on the training set $\{x_{tr_m}\}_{(m=1,\dots,M)}$.

Finally, it should be noted that a similar approach based on spectrogram reconstruction using the VAE has recently been used to assess anomaly in urban sounds [45]. In that work, the threshold used to assess anomaly is computed as the root mean square error (RMSE) of the average reconstruction error on the training set, whereas in our approach, τ_0 is determined by (6).

3.3. Proposed method: interval-valued fuzzy sets based on VAE's reconstruction error

The baseline VAE anomaly detection model provides a primary/naive method to predict whether the input sample (i) is normal or anomalous. In fact, such a baseline approach does not take into consideration the issue of uncertainty in audio data, that has already been explained in the introduction (cf. Section 1, §Motivation: The uncertainty problem). As a result, we proceed to a more elaborated method, where an interval-valued fuzzy membership function is obtained from the VAE's reconstruction error (ϵ). Therefore, fuzzy sets, including Type-1 FS, Type-2 FS, Interval-Valued FS and Interval Type-2 FS are defined hereafter.

3.3.1. Fuzzy sets

The soft/fuzzy clustering approach consists in using a membership function instead of a hard/crisp membership decision. Then an object belongs to all clusters, but with different membership degrees, having values between 0 and 1 [5].

Type-1 fuzzy sets. Type-1 fuzzy clustering problem can be stated as follows: Given a set $X = \{x_1, ..., x_n\}$ of data objects, a set $\Omega = \{\omega_1, ..., \omega_c\}$ and a membership function $u(x, \omega)$, find $\omega \in \Omega$ such that $\forall x \in X$, $\forall \omega \in \Omega$, $0 \le u(x, \omega) \le 1$. In the following, $u(x, \omega)$ will denote the primary membership.

Type-2 fuzzy sets. Type-2 fuzzy sets [47], also known as General Type-2 fuzzy sets [21,46], are defined as a 2-variable membership function $\mu_A(x, u)$ where $\forall\, x \in X$, $\forall\, u \in J_x \subseteq [0, 1]$, $\tilde{A} = \{(x, u), \mu_A(x, u) | \forall x \in X, \forall u \in J_x \subseteq [0, 1]\}$ and $0 \le \mu_{\tilde{A}}(x, u) \le 1$. $\mu_{\tilde{A}}(x, u)$ is named the second grade. Thus, the General Type-2 Fuzzy Sets are defined as a generalization of Type-1 fuzzy sets, as in [84,85].

Interval-valued fuzzy sets (IVFS). An interval-valued fuzzy set (IVFS) is a particular variant of type-2 fuzzy sets, and therefore called also interval type-2 fuzzy sets. An IVFS A on the universe $U \neq \emptyset$ is a mapping $\underline{A} : U \to L([0, 1])$, such that the membership degree of $u \in U$ is given by $A(u) = [\underline{A}(u), \overline{A}(u)] \in L([0, 1])$, where $\underline{A} : U \to [0, 1]$ and $\overline{A} : U \to [0, 1]$ are mappings defining the lower and the upper bounds of the membership interval $A(u)$, respectively [8].

Interval Type-2 fuzzy sets. In the above definition of Type-2 fuzzy sets, if all second grades $\mu_{\tilde{A}}(x, u)$ are equal to 1, then $\tilde{A}$ is named an interval type-2 fuzzy set [47]. Therefore, in [9,48], interval type-2 fuzzy sets are defined as a generalization of interval-valued fuzzy sets (IVFS).

Footprint of uncertainty (FOU). Uncertainty in the primary membership of a type-2 fuzzy set, $\tilde{A}$, consists of a bounded region known as the footprint of uncertainty (FOU). It is expressed as the union of all primary memberships, *i.e.*,

$$FOU(\tilde{A}) = \bigcup_{x \in J} J_x. \tag{7}$$

As argued by [47], the term footprint of uncertainty is extremely useful because it not only focuses the attention on the uncertainties inherent in a specific type-2 membership function, the shape of which is a direct result of the nature of these uncertainties, but it also provides a very convenient verbal description of the entire domain of support for all secondary grades of a type-2 membership function. Another advantage also advanced by [47] consists in the ability of the FOU region to depict a type-2 fuzzy set graphically in two-dimensions instead of three dimensions, and in so doing helps to overcome the three-dimensional nature of type-2 fuzzy sets which makes them very difficult to draw.

The FOU region, e.g. the region delimited by μ_L and μ_U in Fig. 3, indicate the presence of a distribution on top of it, *i.e.*, the new third dimension of type-2 fuzzy sets. The shape of that distribution is determined by the secondary grade level chosen. To date, these are the most often used type-2 fuzzy sets, as reported by [47].

Interpretation of interval-valued fuzzy sets. The fuzzy formalism adopts continuous truth values, or set memberships, instead of a binary membership. It can be used to model partially true concepts, avoiding decisions that are weakly supported. However, according to [47,8], when it comes to model data or membership uncertainty, the basic type-1 fuzzy sets seem unfit. Therefore, type-2 fuzzy sets have been designed, and in particular interval-valued type-2 fuzzy sets have been proposed to deal with such issues. In [8], two different interpretations of IVFS are proposed:

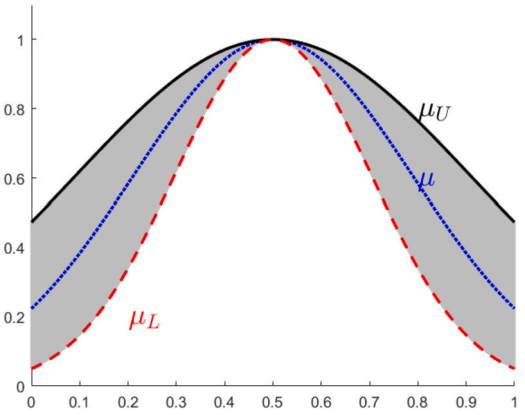
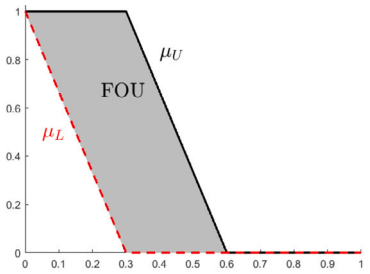
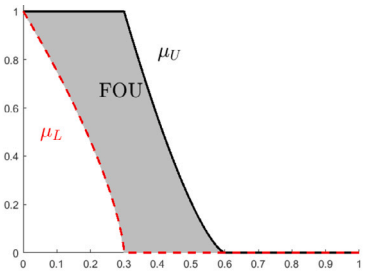


Fig. 3. Illustration of the interval-valued type-2 fuzzy sets using a Gaussian membership function: μ represents the membership function without uncertainty; μ_L and μ_U represent the lowest and the highest bounds, respectively, of the membership function when uncertainty is taken into account; the area comprised between μ_L and μ_U (in grey) represents the footprint of uncertainty (FOU), cf. (7).



(a) Piecewise linear membership function ($\alpha_{j,L} = \alpha_{j,U} = 1$ in (8,9))



(b) Piecewise nonlinear membership function ($0 < \alpha_{j,L}, \alpha_{j,U} < 1$ in (8,9))

Fig. 4. Proposed 2-component piecewise linear and nonlinear membership functions; the dashed and the continuous lines indicate the lower membership $\mu_L(\epsilon)$ and the upper membership $\mu_U(\epsilon)$ components, respectively, cf. (8), (9), (ϵ is the input VAE error given by (4)); the area comprised between μ_L and μ_U (in grey) is the footprint of uncertainty (FOU), cf. (7).

a) As mathematical entities in which each element's membership in the set is determined by a closed interval. Obviously, because two functions $\underline{A}$ and $\overline{A}$ are considered in this interpretation, the difficulty of working with fuzzy sets is duplicated. This interpretation is relevant only from a theoretical standpoint.
b) Each interval denotes that the expert has a good idea of a lower and an upper bound for an element's membership degree in the fuzzy set, but does not know the exact value of that membership. This interpretation is more relevant when working with imprecise information about the degree of membership. In comparison to standard fuzzy sets, when IVFS are processed, this uncertainty is present, making the results less specific but more credible.

In general, membership functions of IVFS are less specific than those of fuzzy sets, although this lack of specificity makes them more realistic in some applications. Their benefit is that they allow us to be skeptical about identifying a precise membership function, as argued by [8].

3.3.2. *Proposed fuzzy membership functions*

In this work, we used some particular forms of fuzzy membership function based on the output error of the baseline VAE, trained on normal data only, to model the membership to each class $j \in \{0,1\}$ where 0 and 1 stand for the classes *normal* and *anomalous*, respectively. Thus, for each class, the corresponding membership function is composed of: i) a Pessimistic/Lower component $\mu_{j,L}$, and ii) an Optimistic/Upper membership $\mu_{j,U}$ ($j \in \{0,1\}$) cf. (8), (9).

Piecewise linear/nonlinear fuzzy membership function. First, we set a piecewise linear fuzzy membership function made of a lower/pessimistic and an upper/optimistic components given by (8) and (9), respectively, where $\alpha_{j,L} = \alpha_{j,U} = 1$ yields the piecewise linear membership, and $0 < \alpha_{j,L}, \alpha_{j,U} < 1$ ($j \in \{0,1\}$) gives the nonlinear shape of the membership functions as depicted in Fig. 4a and Fig. 4b, respectively. In both figures, the footprint of uncertainty (FOU) for each pair of membership functions $\{\mu_{j,L}, \mu_{j,U}\}$ is represented by the area comprised between the curves of the lower and the upper components. It should be noted that the parameters $\{a_j, b_j\}$ ($j \in \{0,1\}$) in (8) and (9) are set empirically, following two main criteria: i) The performance of anomaly detection on

Table 2
Membership functions used for evaluation of IVFS with respect to VAE-error's uncertainty (ϵ-based uncertainty): $0 < \omega_0, \omega_1 < 1$ are the weights of the classes *normal* and *anomalous*, respectively, such that $\omega_0 + \omega_1 = 1$; the value of τ is empirically set in $]0, 1[$ as $\tau = (1 - \omega_0)\tau_0$ where τ_0 is the threshold used in (5).

Membership function type	Generic membership function	Lower membership ($\mu_{j,L}$) parameters	Upper membership ($\mu_{j,U}$) parameters
Triangular	$\mu(\epsilon) = \begin{cases} 0 & \text{if} \quad \epsilon \le 0, \\ \frac{\epsilon - a}{m - a} & \text{if} \quad a < \epsilon < m, \\ \frac{b - \epsilon}{b - m} & \text{if} \quad m \le \epsilon < b, \\ 0 & \text{if} \quad \epsilon \ge b. \end{cases}$	$\begin{cases} a = \omega_j \tau, \\ b = 1 - \omega_j \tau, \\ m = \frac{b+a}{2}. \end{cases}$	$\begin{cases} a = 0, \\ b = 1, \\ m = \frac{b+a}{2}. \end{cases}$
Trapezoidal	$\mu(\epsilon) = \begin{cases} 0 & \text{if} \quad (\epsilon \le a) \text{ or } (\epsilon \ge d), \\ \frac{\epsilon - a}{b - a} & \text{if} \quad a < \epsilon < b, \\ 1 & \text{if} \quad b \le \epsilon < c, \\ \frac{d - \epsilon}{d - c} & \text{if} \quad c \le \epsilon < d. \end{cases}$	$\begin{cases} a = 0, \\ b = \tau, \\ c = 1 - \tau, \\ d = 1. \end{cases}$	$\begin{cases} a = 0, \\ b = \omega_j \tau, \\ c = 1 - \omega_j \tau, \\ d = 1. \end{cases}$
Linear Γ-function	$\mu(\epsilon) = \begin{cases} 0 & \text{if} \quad (\epsilon \le a), \\ \frac{\epsilon - a}{b - a} & \text{if} \quad a < \epsilon < b, \\ 1 & \text{if} \quad \epsilon \ge b. \end{cases}$	$\begin{cases} a = 0, \\ b = \tau. \end{cases}$	$\begin{cases} a = 0 \\ b = \omega_j \tau. \end{cases}$
Piecewise linear function	$\mu_L(\epsilon) = \begin{cases} 1 - \frac{\epsilon}{a} & \text{if} \quad 0 \le \epsilon < a, \\ 0 & \text{if} \quad \epsilon \ge a. \end{cases}$ $\mu_U(\epsilon) = \begin{cases} 1 & \text{if} \quad 0 \le \epsilon < a, \\ \frac{b - \epsilon}{a} & \text{if} \quad a \le \epsilon < b, \\ 0 & \text{if} \quad \epsilon \ge b. \end{cases}$	$a = \omega_j \tau$	$\begin{cases} a = \omega_j \tau, \\ b = 2\omega_j \tau. \end{cases}$

the training set, ii) The type of uncertainty modeling, either with respect to the input, *i.e.*, the VAE's reconstruction error (ϵ-based uncertainty) (cf. Table 2), or with respect to the primary membership (μ-based uncertainty) (cf. Table 3):

$$\mu_{j,L}(\epsilon) = \begin{cases} (1 - \frac{\epsilon}{a_j})^{\alpha_{j,L}} & \text{if} \quad 0 \le \epsilon < a_j, \\ 0 & \text{if} \quad \epsilon \ge a_j. \end{cases} \tag{8}$$

and

$$\mu_{j,U}(\epsilon) = \begin{cases} 1 & \text{if} \quad 0 \le \epsilon < a_j, \\ (\frac{b_j - \epsilon}{a_j})^{\alpha_{j,U}} & \text{if} \quad a_j \le \epsilon < b_j, \\ 0 & \text{if} \quad \epsilon \ge b_j, \end{cases} \tag{9}$$

where $j \in \{0, 1\}$ denotes the classes *normal* and *anomalous*, respectively, and $\{\alpha_{j,L}, \alpha_{j,U}\} \subset [0, 1]$ defines the linearity/nonlinearity of the membership function.

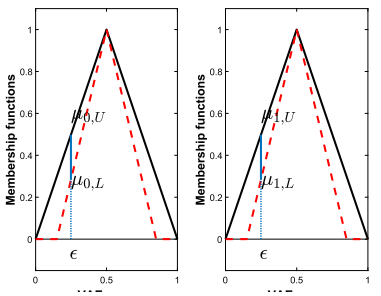
The parameters $\{a_j, b_j\}$ are used to set the upper boundary and the slope of the linear piecewise membership function, as a function of the hypothetic weight of the normal and the anomalous class ω_0 and ω_1, respectively, whereas the parameters $\alpha_{j,L}, \alpha_{j,U}$ are intended to control the tightness of the FOU area in the nonlinear piecewise membership function. It should also be noted that these parameters are specific to the proposed method, and were only introduced in [52] that describes the preliminary results of the proposed method.

Other basic membership functions. In addition to the proposed piecewise linear/nonlinear membership functions, other conventional membership functions are also utilized to evaluate uncertainty, including piecewise linear functions such as triangular, trapezoidal and linear Γ-function, as shown in Fig. 5, as well as nonlinear ones, such as Gaussian, Laplacian and tangent-hyperbolic functions, as depicted in Fig. 6. Similarly to the proposed piecewise linear/nonlinear membership function, each of the aforementioned functions will undergo some modification to extract a lower component and an upper one in order to evaluate uncertainty, either with respect to the input, *i.e.*, the VAE reconstruction error (ϵ) (cf. Table 2) or to the value of the primary membership (cf. Table 3).

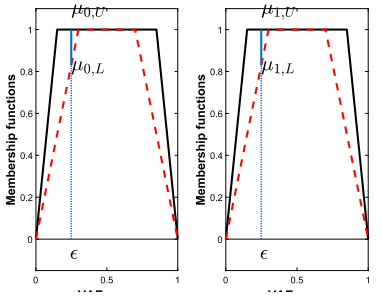
3.3.3. Classwise membership functions

To create a FOU region, each membership function is decomposed into a lower/pessimistic component and an upper/optimistic one. To model the membership to each class, *i.e.*, *normal* or *anomalous*, the parameters of each utilized membership function are modified in order to be adapted to each class. This can be done using an arbitrary weight $\omega_j \in [0, 1]$ for each class $j \in \{0, 1\}$. Hence, for each membership function component, *i.e.*, μ_L and μ_U, a classwise term is created, *i.e.*, $\mu_{j,L}$ and $\mu_{j,U}$ ($j \in \{0, 1\}$). Then, for each class, FOU is the region delimited by the curves of $\mu_{j,L}$ and $\mu_{j,U}$ (cf. Fig. 5 and Fig. 6).

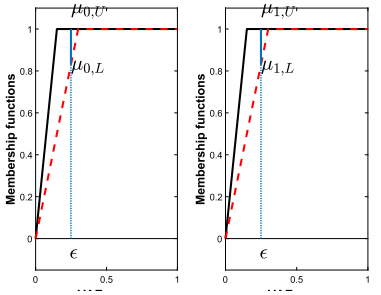
Thus, the so created FOU region allows evaluating the uncertainty. At this level, two options are available: i) Modeling the uncertainty with respect to the input, *i.e.*, the VAE reconstruction error (ϵ), so that the 2-D membership function is evaluated on the



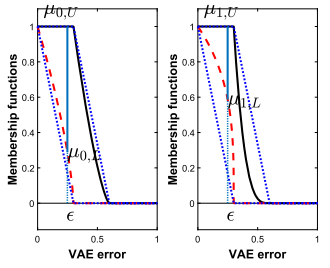
(a) Triangular membership function



(b) Trapezoidal membership function



(c) Γ-Linear membership function



(d) Piecewise linear membership function

Fig. 5. Evaluation of IVFS with respect to the input VAE error uncertainty (ϵ-based uncertainty): The continuous and the dashed lines indicate the upper and the lower membership components, *i.e.*, $\mu_{j,L}$ and $\mu_{j,U}$ ($j \in \{0,1\}$), respectively, as detailed in Table 2); the vertical straight line between the lower and the upper membership components indicates the interval $[\mu_{j,L}(\epsilon), \mu_{j,U}(\epsilon)]$: interval comparison is performed between $[\mu_{0,L}(\epsilon), \mu_{0,U}(\epsilon)]$ and $[\mu_{1,L}(\epsilon), \mu_{1,U}(\epsilon)]$ to determine the corresponding class using (10)-(14).

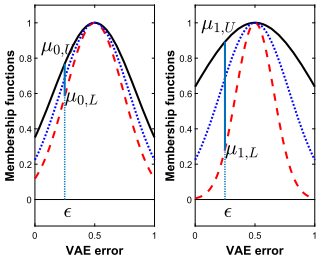
Table 3
Membership functions used for evaluation of IVFS with respect to the primary membership's uncertainty (μ-based uncertainty): $0 < \omega_0, \omega_1 < 1$ are the weights of the classes *normal* and *anomalous*, respectively, such that $\omega_0 + \omega_1 = 1$; the value of τ is empirically set in $]0,1[$ as $\tau = (1-\omega_0)\tau_0$ where τ_0 is the threshold used in (5).

Membership function type	Generic membership function	Lower membership ($\mu_{j,L}$) parameters	Upper membership ($\mu_{j,U}$) parameters
Gaussian	$\mu(\epsilon) = (\exp\{-0.5(\frac{(\epsilon-m)}{\sigma})^2\})^\alpha$	$\alpha = \frac{1}{\omega_j}$	$\alpha = \omega_j$
Laplacian	$\mu(\epsilon) = (\exp\{-0.5\frac{\lvert\epsilon-m\rvert}{\sigma}\})^\alpha$	$\alpha = \frac{1}{\omega_j}$	$\alpha = \omega_j$
Tanh	$\mu(\epsilon) = (\frac{e^{\frac{\epsilon}{\tau}} - e^{-\frac{\epsilon}{\tau}}}{e^{\frac{\epsilon}{\tau}} + e^{-\frac{\epsilon}{\tau}}})^\alpha$	$\alpha = \frac{1}{\omega_j}$	$\alpha = \omega_j$
Piecewise nonlinear function	$\begin{cases} \mu_L(\epsilon) = \begin{cases} (1-\frac{\epsilon}{a})^\alpha_{j,L} & \text{if } 0 \le \epsilon < a, \\ 0 & \text{if } \epsilon \ge a. \end{cases} \\ \mu_U(\epsilon) = \begin{cases} 1 & \text{if } 0 \le \epsilon < a, \\ (\frac{b-\epsilon}{a})^\alpha_{j,U} & \text{if } a \le \epsilon < b, \\ 0 & \text{if } \epsilon \ge b. \end{cases} \end{cases}$	$\begin{cases} a = \omega_j\tau, \\ \alpha_{j,L} = \frac{1}{\omega_j}. \end{cases}$	$\begin{cases} a = \omega_j\tau, \\ b = 2\omega_j\tau. \\ \alpha_{j,U} = \omega_j. \end{cases}$

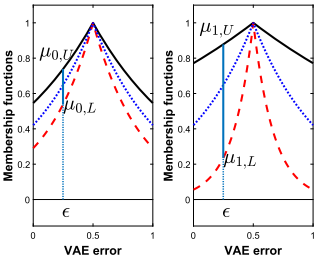
first dimension; ii) Creating a secondary membership function that depends on the variation on the primary membership, *i.e.*, with respect to the second dimension, *i.e.*, $\mu(\epsilon)$, as follows:

ϵ-based uncertainty. The first option consists in modeling the uncertainty with respect to the VAE reconstruction error, *i.e.*, ϵ-axis. This can be obtained through the variation of the parameters of the membership functions on the ϵ-axis. Therefore, for each membership function in Table 2, the parameters are set using the weights $\{\omega_0, \omega_1\} \subset]0,1]$ so that the obtained lower and upper membership components describe the uncertainty of the VAE reconstruction error (ϵ). Table 2 mentions the required parameterization to obtain such a configuration of the membership functions.

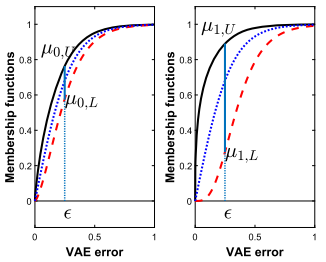
μ-based uncertainty. The second option is to model uncertainty with respect to the primary membership values, *i.e.*, the μ-axis. Thus, for each class, the corresponding membership function is composed of: i) a Pessimistic/Lower component $\mu_{j,L}$, and ii) an



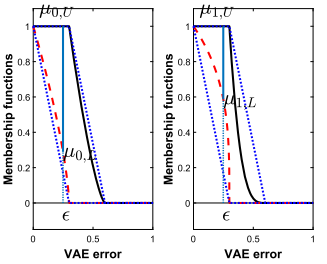
(a) Gaussian membership function



(b) Laplacian membership function



(c) Tanh membership function



(d) Piecewise nonlinear membership function

Fig. 6. Evaluation of IVFS with respect to the primary membership's uncertainty (μ-based uncertainty): The continuous and the dashed lines indicate the upper and the lower membership components, *i.e.*, $\mu_{j,L}$ and $\mu_{j,U}$ ($j \in \{0,1\}$), respectively, as detailed in Table 3); the vertical straight line between the lower and the upper membership components indicates the interval $[\mu_{j,L}(\epsilon), \mu_{j,U}(\epsilon)]$: interval comparison is performed between $[\mu_{0,L}(\epsilon), \mu_{0,U}(\epsilon)]$ and $\left[\mu_{1,L}(\epsilon), \mu_{1,U}(\epsilon)\right]$ to determine the corresponding class using (10)-(14).

Optimistic/Upper membership $\mu_{j,U}$ ($j \in \{0,1\}$). The parameterization of each membership component corresponding to each class $j \in \{0,1\}$ is presented in Table 3.

Yielded footprint of uncertainty. Using such a parametrization, with respect either to the input, *i.e.*, the VAE reconstruction error (ϵ), or to the membership value ($\mu(\epsilon)$), as given in Table 2) and Table 3, respectively, both membership functions components define an area between the lower and the upper membership function curves, *i.e.*, $\mu_{j,L}$ and $\mu_{j,U}$, that represents the FOU region for ϵ-based and μ-based uncertainty, as shown in Fig. 5 and Fig. 6, respectively. In type-2 fuzzy sets, the FOU region indicates where the secondary membership, *i.e.*, the third dimension of the type-2 fuzzy set, has non-zero value for the 2-D input formed by $(\epsilon, \mu(\epsilon))$. In particular, for interval-valued fuzzy sets, the FOU region corresponds to the region where the secondary membership equals 1 for all 2-D inputs $(\epsilon, \mu(\epsilon))$.

3.3.4. *Interval comparison method*

The membership functions calculated using both the ϵ-based and the μ-based uncertainty yield a couple of intervals $\left[\mu_{j,L}, \mu_{j,U}\right]$ for each class index $j \in \{0,1\}$. Therefore, it is necessary to perform interval comparison to evaluate the interval-valued fuzzy membership. Interval comparison, known also as degree of preference, can be understood like a measure of confidence. In fact, the smaller the interval, the greater the confidence, and hence the tighter the membership function. Interval comparison is a particular case of fuzzy number comparison that has been studied extensively for several years [28,37], utilizing either probabilistic [26,73] or possibilistic [32] approaches, as well as fuzzy set theory [92]. For further information on interval and fuzzy number comparison, see [86,87].

The goal of interval comparison is to rank real-number intervals or fuzzy numbers based on their boundary values. According to [88], a developed heuristic strategy provides the distinct advantage of not relying on midpoints for interval comparison. This makes sense, especially when dealing with fuzzy values or confidence intervals. To compare two intervals $A = \left[a_1, a_2\right]$ and $B = \left[b_1, b_2\right]$, [88], as reported in [73], defines the degree of preference of A over B, denoted $P(A > B)$, using:

$$P(A > B) = \frac{\max(0, a_2 - b_1) - \max(0, a_1 - b_2)}{(a_2 - a_1) + (b_2 - b_1)}. \tag{10}$$

Reciprocally, the degree of preference of B over A is defined as:

$$P(B > A) = \frac{\max(0, b_2 - a_1) - \max(0, b_1 - a_2)}{(a_2 - a_1) + (b_2 - b_1)}, \tag{11}$$

so that

$$P(A > B) + P(B > A) = 1, \tag{12}$$

and

$$\begin{cases} \text{if} \quad A \equiv B \quad \text{then} \quad P(A > B) = P(B > A) = 0.5, \\ \text{if} \quad a_2 < b_1 \quad \text{then} \quad P(B > A) = 1. \end{cases} \tag{13}$$

The secondary membership function as a degree of preference of intervals can be computed using (10), (13). To do so, the pessimistic/lower and optimistic/upper membership functions for the input VAE reconstruction error of a given sample (i), *i.e.*, (ϵ), are calculated for each class $j \in \{0, 1\}$, and then used to build an interval $\left[\mu_{j,L}(\epsilon), \mu_{j,U}(\epsilon)\right]$. Then, for each sample, the so-formed intervals, *i.e.*, $\left[\mu_{0,L}(\epsilon), \mu_{0,U}(\epsilon)\right]$ and $\left[\mu_{1,L}(\epsilon), \mu_{1,U}(\epsilon)\right]$, are compared. Finally, defuzzification entails matching the event class to the interval chosen as the least favored, as specified by (14):

$$Event(i) = \arg\min_{j=0,1} \{P(A_j > A_{k \neq j})\}, \tag{14}$$

where

$$\begin{cases} A_j & = & [\mu_{L,j}(\epsilon), \mu_{U,j}(\epsilon)], \\ A_{k \neq j} & = & [\mu_{k,L}(\epsilon), \mu_{k,U}(\epsilon)] \; \forall \; k \neq j. \end{cases}$$

The proposed method's principle is depicted in Fig. 5 and Fig. 6 for different shapes of membership functions, tailored for ϵ-based and μ-based uncertainty, respectively.

3.3.5. Methodology

The proposed approach is applied to the addressed problem, i.e. anomaly detection in audio signals, following these steps (also cf. Algorithm 1):

- Each audio chunk, of approximately 1 sec, is segmented into overlapping frames
- For each frame, the spectrogram is computed, then the {MFCC, Δ-MFCC, Δ-Δ-MFCC} vector is extracted.
- The {MFCC, Δ-MFCC, Δ-Δ-MFCC} vectors extracted for all the frames are used to build the input feature matrix of the audio chunk.
- A variational autoencoder model is trained on the subset containing only normal samples, *i.e.*, ordinary street noise without any anomalous event.
- For each input audio clip in the test phase, the VAE reconstruction error is calculated between the input MFCC matrix, and the reconstructed one obtained at the output of the VAE network, using (4), (5).
- The obtained VAE reconstruction error is used to calculate a pair of lower and upper membership functions, corresponding to each class, *i.e.*, *normal* and *anomalous*, respectively, using (8), (9) for the proposed piecewise linear/nonlinear membership functions, or the equations cited in Table 2 and Table 3 for the other basic membership functions.
- The intervals yielded for the classes *normal* and *anomalous*, *i.e.*, $[\mu_{0,L}(\epsilon), \mu_{0,U}(\epsilon)]$ and $[\mu_{1,L}(\epsilon), \mu_{1,U}(\epsilon)]$, respectively, are compared using the method described in (10)-(13).
- The chunk-level event class is returned by (14).

4. Experiments and results

The experimental protocol is presented below. It includes the data processing stages, the experimental parameters of the different models and the evaluation measures selected. The results are then presented, analyzed and discussed.

4.1. Audio database

Recently, a complete inventory of datasets for anomalous sound event identification has been presented in [50], including audio traffic datasets such as the AXA database [71], WASN [3], and MIVIA dataset [15]. The latter has the advantage of being the only open-access audio traffic monitoring database —at the time of elaboration of this work—. It includes nearly one hour of traffic sound signals captured in a real road environment at 23 different sites in the region of Salerno, Italy, including city centers, highways, and country roads.

The audio signals were recorded using an Axis P8221Audio Module and an Axis T83 omnidirectional microphone for audio surveillance applications. Signals were sampled at 32 KHz and encoded at 16 bits per PCM sample. The audio clips are available online in WAV file format [1]. Table 4 shows the composition of the dataset. The original dataset is divided into 57 one-minute audio clips that were manually annotated. The annotation file contains the event names, such as car accident and tire skidding, whereas background noise, comprises all other uninteresting events, such as pedestrian voices, horn blowing, or simply street noise. Also, the onset and offset time of each event are available for timely audio tagging. It should be noted that this database has been used in several related works, such as in [16,19,20,49,51,66,69,79], during the last few years.

Algorithm 1: Anomaly detection for each audio chunk.

Result: Event of each audio chunk (*normal / anomalous*)
for *Each audio chunk (k)* **do**
  for $n \leftarrow 1 : N$ *(i.e Number of overlapping frames)* **do**
    Sp(n) $\leftarrow$ Mel-Spectrogram(*Frame*) (i.e extract the spectrogram in the Mel-frequency domain);
    MFCC(n) $\leftarrow$ DCT(log(|**Sp(n)**|);
    Mel-Features(n) $\leftarrow$ [MFCC$_{n,0} \dots$ MFCC$_{n,12}$, Δ-MFCC$_{n,0} \dots \Delta$-MFCC$_{n,12}$, Δ-Δ--MFCC$_{n,0} \dots \Delta$-Δ-MFCC$_{n,12}$];
  end
  S(k) $\leftarrow$ [**Mel-Features**$(1) \dots$ **Mel-Features**(N)];
  $\hat{\mathbf{S}}(\mathbf{k}) \leftarrow$ VAE(**S(k)**);
  $\epsilon \leftarrow ||\mathbf{S(k)} - \hat{\mathbf{S}}(\mathbf{k})||_2$;
  if *Method* $==$ *Baseline* **then**
    if $\epsilon \leq \tau$ **then**
      Class$(k) \leftarrow 0$ (*normal*);
    else
      Class$(k) \leftarrow 1$ (*anomalous*);
    end
  else
    // Method $==$ IVFS ;
    if *Uncertainty* $==$ ϵ*-based (i.e., on the VAE's error)* **then**
      Select a membership function ($\mu(\epsilon)$) from Table 2
    else
      if *Uncertainty* $==$ μ*-based (i.e., on the membership function)* **then**
        Select a membership function ($\mu(\epsilon)$) from Table 3
      end
    end
    for $j \leftarrow 0 : 1$ **do**
      $A_j \leftarrow [\mu_{j,L}(\epsilon), \mu_{j,U}(\epsilon)]$;
      (i.e the interval-valued membership function to each class)
    end
    Class$(k) \leftarrow \arg \min_{j \in \{0,1\}} P(A_j) > P(A_{j \neq k})$ (cf. (10)-(14));
  end
end
Event $\leftarrow$ Class(k);

Table 4
Composition of MIVIA road traffic audio dataset [1] before and after data augmentation.

Before data augmentation			After data augmentation		
Label	Count	Duration(s)	Label	Count	Duration(s)
Background noise		2732	Normal	6911	6911
Car crashes	200	326.28	Anomalous	1012	1012
Tire skidding	200	522.50			

4.2. Data augmentation

Data augmentation was performed by segmenting the audio clips into short overlapping chunks. Each chunk has a duration of 1 second with a 50% overlap on the previous one. In this way, the amount of short chunks was increased by 100%. Furthermore, by using this technique, we ensure that each chunk contains at most one anomalous event, so that the chunk can only be labeled as *normal* or *anomalous*. Anomalous events are only car accidents, while all other events, such as pedestrian voices, traffic sound and tire skidding, or simply street noise, are considered normal. Therefore, the binary labeling into *normal* and *anomalous* events and the segmentation of the original audio clips into overlapping chunks increased the number of audio data from 57 multi-labeled 1-minute clips to 7923 single labeled 1-second chunks.

Table 4 shows the increase in the number of audio clips and associated events due to the data augmentation.

4.3. Feature set

The authors established in a previous work [49] that the discriminatory strength of some of the common features used for sound event detection, in particular for DCASE challenges [78], is unsuitable for the problem of road audio surveillance. Actually, uncertainty is one of the most significant issues in audio modeling. This issue is exacerbated when the sound of some target events is inherently mixed with background noise, as in the case of audio traffic.

In [66], following a thorough examination of the discriminatory potential of the standard feature set proposed at the DCASE'2013 challenge for audio event detection [78], the energy and cepstral coefficients were determined to be the most discriminatory [66]. Indeed, the efficacy of MFCC coefficients for speech and speaker recognition has long been demonstrated due to their high ability to capture the broad spectral properties of an audio event [62]. Following this feature selection process, the feature set was limited to

Table 5
Convolutional variational autoencoder (CVAE) architecture; (*) in the validated architecture, Bottleneck is set to 40.

Part of the autoencoder	Hidden layers	Size×number of conv. filters	Transfer function	Size of stride
Encoder network	Conv2D	(3×3)×32	Relu	(2×2)
	Conv2D	(3×3)×64	Relu	(2×2)
	Conv2D	(3×3)×128	Relu	(2×2)
Code layer	Fully connected	Bottleneck*		
Decoder network	Transposed Conv2D	(3×3)×128	Relu	(2×2)
	Transposed Conv2D	(3×3)×64	Relu	(2×2)
	Transposed Conv2D	(3×3)×32	Relu	(2×2)
Output			Sigmoid	

Table 6
CVAE training parameters.

Training parameters	Value
Execution Environment	"Auto" ("GPU" or "CPU")
Learning algorithm	ADAM
Number of Epochs	100
Mini-batch size (M)	32
Learning rate	1e-3
Gradient decay	0.9
Number of training images (N)	6911
Number of Iterations	$\lfloor \frac{N}{M} \rfloor$

a set of 13 MFCC coefficients, as well as log-energy, and their first and second derivatives (Δ and Δ-Δ), that are extracted from the Mel-log spectrum, using the method detailed in [2].

To extract MFCC features and their first and second derivatives, each audio clip in the database is segmented into overlapping chunks. Then, for each chunk, the mel-spectrogram is computed using 64 mel-bands and a Hann window of length 1024 points with 50% hop rate. For each feature, *i.e.*, MFCC, Δ-MFCC and Δ-Δ-MFCC, a matrix containing all vectors at the chunk level is built. Finally a 3-D matrix composed of the so-obtained matrices is constructed.

4.4. VAE network's architecture

Convolutional layers were used to build the VAE network. This architecture, also known as convolutional VAE (CVAE), has been used in a number of anomaly detection algorithms, including those for image, video, and, more recently, audio processing [50]. The fundamental idea is to offer the input as a 2-D or 3-D matrix that convolutional layers can process. Thus, in the case of audio streams, the input features are extracted at chunk level, to form a 2-D or a 3-D matrix. Then, either at the encoder or decoder side, two or more convolutional layers are stacked. The code layer is generally taken as a bottleneck fully-connected layer. The final VAE's architecture and the set of training hyperparameters are detailed in Tables 5, 6, respectively. Both were set up empirically as those giving the best evaluation results. It is also worth mentioning that this VAE's architecture has been tailored for this work, hence not relying on pretrained models.

4.5. Experimental protocol

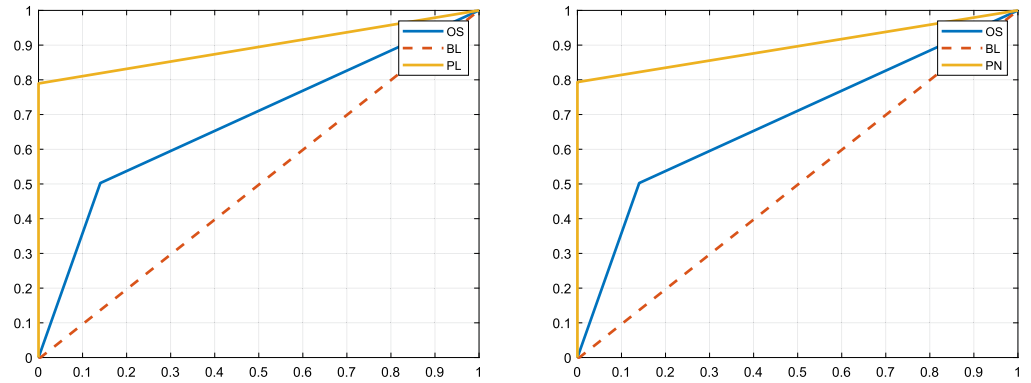
The goal of the experimental work is to detect anomalous audio events on roads, in particular car accidents, in the following steps: i) Extracting features from the chosen audio traffic database, *i.e.*, MIVIA DB [15]; ii) training the VAE model on normal events' data only; iii) calculating the VAE reconstruction error for each sample in the test set using (4); iv) Evaluating the lower and the upper membership function components to form the intervals bounded by the lower and the upper membership values, using any of the linear or nonlinear membership functions mentioned in Table 2 or Table 3, respectively; v) implementing and evaluating the interval-valued fuzzy membership functions using (10)-(13); and finally, vi) executing defuzzification to detect the predicted class using (14).

As far as feature extraction is concerned, it was discovered that the proportion of non-hazardous (normal) event samples is substantially higher than that of hazardous (anomalous) occurrences, implying that the model is heavily biased towards the class *normal*. To address this issue, data augmentation was used, which allows for more data to be acquired by segmenting audio signals into short chunks with a duration of 1sec and an overlap rate of 50%, so that each chunk contains at most one anomalous event.

As a result, about 7400 chunks were obtained from the original 57 × 1min audio clips. Approximately 80% of the segmented chunks belong to the normal event class, and 20% to the anomalous event class, from which 60% belong to car accident category. Nonetheless, it is important to note that all training chunks, whether normal or anomalous, contain background street noise.

Table 7
Results of anomalous sound event detection metrics (cf. (15), (16)) using the benchmarking OC-SVM method cf. (1), the baseline VAE error thresholding cf. (4), (5) and the proposed linear/nonlinear membership function cf. (8), (9) using the following settings: for OC-SVM, a Gaussian kernel with γ-parameter set to 0.04, was used (for its high performance on the evaluation set); for the baseline VAE and the piecewise linear/nonlinear membership functions, $\tau_0 = 0.5$ is the threshold in (5), and $\{\omega_0, \omega_1\}$ are the weights of the classes *normal* and *anomalous*, respectively, such that $\omega_1 = 1 - \omega_0$; $p = 0.2$ for *p-AUC*; (*) NaN values are due to zero correctly estimated samples.

Method	ω_0	Acc	P_0	P_1	R_0	R_1	$F1_0$	$F1_1$	AUC	$p\text{-}AUC$
OC-SVM		**0.81**	**0.92**	**0.36**	**0.86**	**0.50**	**0.89**	**0.42**	**0.68**	**0.06**
VAE only (Baseline) cf. (4), (5) with $\tau_0 = 0.5$		**0.88**	**0.88**	**0.00**	**1.00**	**0.00**	**0.94**	**NaN***	**0.50**	**0.02**
VAE-IVFS with piecewise linear	0.6	0.59	1.00	0.24	0.53	0.99	0.69	0.39	0.76	0.04
membership function cf. (8), (9)	0.7	0.79	0.99	0.39	0.76	0.94	0.86	0.55	0.85	0.08
with $a_j = (1-\omega_j)(1-\omega_0)\tau$ and	0.8	0.90	0.98	0.57	0.90	0.90	0.94	0.70	0.90	0.14
$b_j = 2(1-\omega_j)(1-\omega_0)\tau$;	0.9	**0.97**	**0.97**	**1.00**	**1.00**	**0.77**	**0.98**	**0.87**	**0.89**	**0.12**
$\alpha_{j,L} = \alpha_{j,U} = 1$ $(j \in \{0,1\})$										
VAE-IVFS with piecewise nonlinear	0.6	0.59	0.99	0.24	0.54	0.95	0.69	0.39	0.74	0.04
membership function cf. (8), (9) and	0.7	0.64	0.99	0.25	0.59	0.97	0.74	0.40	0.78	0.05
with $a_j = (1-\omega_j)(1-\omega_0)\tau$,	0.8	0.91	0.98	0.62	0.91	0.89	0.94	0.73	0.90	0.13
$b_j = 2(1-\omega_j)(1-\omega_0)\tau$;	0.9	**0.97**	**0.97**	**1.00**	**1.00**	**0.80**	**0.98**	**0.89**	**0.90**	**0.12**
$\alpha_{j,L} = \omega_j$ and $\alpha_{j,U} = \frac{1}{\omega_j}$ $(j \in \{0,1\})$										



(a) ROC curves for ϵ-based uncertainty modeling

(b) ROC curves for μ-based uncertainty modeling

Fig. 7. Comparison of ROC curves for the benchmarking methods, *i.e.*, (OS): OC-SVM cf. (1), (BL): Baseline VAE error thresholding cf. (4), (5), (PL): Piecewise linear membership and (PN): Piecewise nonlinear membership (using $\omega_0 = 0.9$), cf. (8), (9).

Finally, convolutional variational autoencoder networks were built using the architecture detailed in Table 5, and utilizing an input feature vector composed of log-energy features MFCC(0), spectral features MFCC(1-13), and their first and second derivatives (Δ and Δ-Δ). 80% percent of the extracted data was used for training and validation, while the remaining 20% was used for test.

4.6. Analysis of results

The test results are presented in three levels: i) benchmarking of the proposed piecewise linear membership function with the state of the art (cf. Table 7), ii) modeling uncertainty with respect to the input, *i.e.*, the VAE reconstruction error, *i.e.*, ϵ-based uncertainty (cf. Table 8), and iii) modeling uncertainty with respect to the primary membership values, *i.e.*, μ-based uncertainty (cf. Table 9). The values of the event weights $\{\omega_0, \omega_1\}$ were fine-tuned to find out the tradeoff between data distribution and overall performance. It should also be emphasized that to provide a significant meaning of using weights, the condition $\omega_0 + \omega_1 = 1$ was set, so that every weight reflects the *hypothetic* proportion of each class.

4.6.1. Evaluation metrics

Evaluation has been performed using two types of metrics: First, standard measures for classification problems, such as overall accuracy (Acc), and classwise *Precision* (P_j), *Recall* (R_j) and *F1-score* ($F1_j$), were computed as follows:

$$Acc = \frac{\sum_{j=0}^{1} c_j}{\sum_{j=0}^{1} r_j}, P_j = \frac{c_j}{e_j}, \quad R_j = \frac{c_j}{r_j}, \quad F1_j = \frac{2P_jR_j}{P_j + R_j}, \tag{15}$$

Table 8
Results of anomalous event detection metrics (cf. (15), (16)) using linear membership functions (cf. Table 2) to evaluate ϵ-based uncertainty; For the proposed piecewise linear membership function, cf. (8), (9): $\tau_0 = 0.5$ is the threshold in (5), and $\{\omega_0, \omega_1\}$ are the weights of the classes *normal* and *anomalous*, respectively, such that $\omega_1 = 1 - \omega_0$; (*) NaN values are due to zero correctly estimated samples.

Method	w_0	*Accuracy*	P_0	P_1	R_0	R_1	$F1_0$	$F1_1$	AUC	p-AUC
Triangular	0.6	0.87	0.87	NaN*	1.00	0.00	0.93	NaN*	0.50	0.02
	0.7	0.88	0.88	NaN*	1.00	0.00	0.93	NaN*	0.50	0.02
	0.8	0.85	0.85	NaN*	1.00	0.00	0.92	NaN*	0.50	0.02
	0.9	**0.87**	**0.87**	**NaN***	**1.00**	**0.00**	**0.93**	**NaN***	**0.50**	**0.02**
Trapezoidal	0.6	0.98	0.98	1.00	1.00	0.84	0.99	0.91	0.92	0.13
	0.7	0.97	0.97	1.00	1.00	0.79	0.98	0.88	0.89	0.12
	0.8	**0.96**	**0.96**	**1.00**	**1.00**	**0.69**	**0.98**	**0.82**	**0.84**	**0.11**
	0.9	0.96	0.95	1.00	1.00	0.67	0.98	0.80	0.83	0.11
Γ-linear	0.6	0.97	0.98	0.92	0.99	0.89	0.98	0.90	0.94	0.14
	0.7	**0.97**	**0.97**	**1.00**	**1.00**	**0.78**	**0.98**	**0.87**	**0.89**	**0.12**
	0.8	0.96	0.95	1.00	1.00	0.68	0.98	0.81	0.84	0.11
	0.9	0.96	0.95	1.00	1.00	0.66	0.98	0.80	0.83	0.11
VAE-IVFS with piecewise linear membership function cf. (8), (9) with $a_j = (1-\omega_j)(1-\omega_0)\tau$ and $b_j = 2(1-\omega_j)(1-\omega_0)\tau$; $\alpha_{j,L} = \alpha_{j,U} = 1$ $(j \in \{0,1\})$	0.6	0.59	1.00	0.24	0.53	0.99	0.69	0.39	0.76	0.04
	0.7	0.79	0.99	0.39	0.76	0.94	0.86	0.55	0.85	0.08
	0.8	0.90	0.98	0.57	0.90	0.90	0.94	0.70	0.90	0.14
	0.9	**0.97**	**0.97**	**1.00**	**1.00**	**0.77**	**0.98**	**0.87**	**0.89**	**0.12**

Table 9
Results of anomalous event detection metrics (cf. (15), (16)) using nonlinear membership functions (cf. Table 3) to evaluate μ-based uncertainty; For the proposed piecewise nonlinear membership function, cf. (8), (9): $\tau_0 = 0.5$ is the threshold in (5), and $\{\omega_0, \omega_1\}$ are the weights of the classes *normal* and *anomalous*, respectively, such that $\omega_1 = 1 - \omega_0$.

Method	w_0	*Accuracy*	P_0	P_1	R_0	R_1	$F1_0$	$F1_1$	AUC	p-AUC
Gaussian	0.6	0.92	0.95	0.75	0.96	0.70	0.95	0.73	0.83	0.11
	0.7	0.94	0.97	0.77	0.97	0.77	0.97	0.77	0.87	0.12
	0.8	**0.95**	**0.98**	**0.77**	**0.96**	**0.86**	**0.97**	**0.82**	**0.91**	**0.13**
	0.9	0.94	0.97	0.74	0.95	0.83	0.96	0.79	0.89	0.13
Laplacian	**0.6**	**0.95**	**0.96**	**0.87**	**0.98**	**0.75**	**0.97**	**0.80**	**0.86**	**0.12**
	0.7	0.94	0.94	0.86	0.98	0.63	0.96	0.73	0.81	0.10
	0.8	0.95	0.96	0.83	0.98	0.74	0.97	0.78	0.86	0.12
	0.9	0.94	0.96	0.80	0.98	0.70	0.97	0.75	0.84	0.11
Tanh	0.6	**0.97**	**0.96**	**1.00**	**1.00**	**0.74**	**0.98**	**0.85**	**0.87**	**0.12**
	0.7	0.96	0.96	1.00	1.00	0.73	0.98	0.85	0.87	0.12
	0.8	0.96	0.96	1.00	1.00	0.64	0.98	0.78	0.82	0.10
	0.9	0.94	0.94	1.00	1.00	0.57	0.97	0.73	0.79	0.09
VAE-IVFS with piecewise nonlinear membership function cf. (8), (9) with $a = (1-\omega_j)(1-\omega_0)\tau$ and $b_j = 2(1-\omega_j)(1-\omega_0)\tau$; $\alpha_{j,L} = \omega_j$ and $\alpha_{j,U} = \frac{1}{\omega_j}$; $j \in \{0,1\}$	0.6	0.59	0.99	0.24	0.54	0.95	0.69	0.39	0.74	0.04
	0.7	0.64	0.99	0.25	0.59	0.97	0.74	0.40	0.78	0.05
	0.8	0.91	0.98	0.62	0.91	0.89	0.94	0.73	0.90	0.13
	0.9	**0.97**	**0.97**	**1.00**	**1.00**	**0.80**	**0.98**	**0.89**	**0.90**	**0.12**

where r_j, e_j and c_j $(j \in \{0,1\})$ denote the number of ground-truth, estimated and correctly detected events for each of the classes *normal* and *anomalous*, respectively.

Whereas *Precision* identifies the ability of the model to find only the relevant samples for the considered class, *Recall* finds all the relevant samples corresponding to the class for which it is computed. Besides, both measures are interested only in measuring the *True Positive Rate* (TPR), and neglect the *False Positive Rate* (FPR). Therefore, even high *Precision* and *Recall* may not be enough to provide a complete evaluation of the performance of the model.

Therefore, the *Receiver-Operator Curve* (ROC), and its *Area-Under-the-Curve* (AUC) may be useful to provide an answer about how the TPR varies in function of FPR. Also, ideally, we would like to see a high TPR for a low FPR. Therefore we compute the *partial Area-Under-the-Curve* (p-AUC) for a pre-determined low FPR rate, to check that TPR is high not only for a high FPR value, but also for a low value. Both AUC and p-AUC are given by (16), where $p = 1$ stands for full AUC, *i.e.*, for FPR = 1.

$$p\text{-AUC} = \frac{1}{\lfloor pN_- \rfloor N_+} \sum_{i=1}^{\lfloor pN_- \rfloor} \sum_{i=j}^{N_+} \mathcal{H}(\mathcal{A}(x_j^+) - \mathcal{A}(x_i^-)), \tag{16}$$

where: $\mathcal{A}(x)$ is the anomaly score such that $\mathcal{H}(x) = 1$ if $x > 0$ and 0 otherwise; $\{x_i^-\}_{i=1}^{N_-}$ and $\{x_j^+\}_{j=1}^{N_+}$ are the normal and anomalous test samples, respectively, sorted in descending order of anomaly scores; N_- and N_+ are the number of total normal and anomalous samples respectively; and $0 < p \leq 1$ is a predefined false positive rate ($p = 1$ for the total AUC). For more details about the evaluation metrics for anomalous sound event detection, cf. [50].

Finally, it is worth noting that standard classification metrics, such as global accuracy, precision, recall and F1 score, generally indicate the global tendency of the classifier, whether being biased or balanced. However, in case of highly imbalanced datasets — as in our case — the AUC score provides an outlook of the true positive rate (sensitivity) regarding specificity. Furthermore, the AUC is sometimes misleading since a higher absolute AUC does not always guarantee higher sensitivity for higher specificity (lower false positive rate). Therefore, calculating the partial AUC (p-AUC) for a low – and realistic – false positive rate allows checking the presence of such a behavior.

4.6.2. Results of benchmarking with the state of the art

Table 7 shows the effectiveness of utilizing an interval-valued fuzzy membership function to improve anomaly detection, in comparison to other state-of-the-art methods of anomaly detection, *i.e.*, OC-SVM and baseline VAE error thresholding. The main advantages of employing the proposed approach are as follows:

- Both suggested methods, *i.e.*, piecewise linear and nonlinear membership functions, surpass the state-of-the-art OC-SVM, in terms of overall accuracy, reaching 97% for the suggested technique vs. 81% for OC-SVM.
- For the suggested membership functions, precision, recall, and F1 score are more evenly distributed between both classes, *i.e.*, *normal* and *anomalous*, and clearly improve the results of the baseline VAE error thresholding method. Actually, despite the good overall accuracy of the baseline method, reaching 88%, simple thresholding is unable to detect the class *anomalous*.
- Similarly, both piecewise linear and nonlinear membership functions outperforms the benchmarking OC-SVM and VAE thresholding methods in terms of AUC, as depicted in Figs. 7a and 7b. This proves that, unlike the benchmarking methods, the high accuracy of both proposed methods does not hide a high false positive rate.
- Finally, the effect of utilizing imbalanced weights is more evidenced, with higher accuracy when w_j is much greater for the class *normal*, e.g. for $\omega_0 \geq 0.8$, hence $\omega_1 \leq 0.2$. In fact, even though these class weights are set empirically, they seem to boost precision, recall, and F1 scores when they are carefully adjusted. This justifies their use in setting the values of the proposed method's parameters as they may reflect the real normal and anomalous data distribution.

4.6.3. Analysis of modeling uncertainty with respect to the VAE reconstruction error (ϵ-based uncertainty)

The results obtained using the ϵ-based uncertainty methods allow making the following observations:

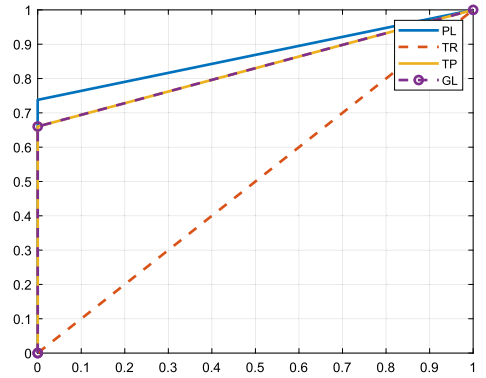
- First, the comparison of the results mentioned at Table 8 to those at Table 7, shows the improvement brought by modeling uncertainty with respect to the input, i.e. the VAE reconstruction error (ϵ) for anomaly detection, for any type of linear membership function, except for the triangular one, over the benchmarking methods, *i.e.*, OC-SVM and baseline VAE error thresholding.
- Secondly, the comparison of the individual performance of each type of membership function (cf. Table 8), shows a slight improvement when using the proposed piecewise linear membership function over the trapezoidal and Γ-linear membership function. However, the overall performance of this set of methods confirms the improvement brought by modeling uncertainty based on the input, *i.e.*, the VAE reconstruction error.
- Thirdly, Table 8 and Fig. 8a shows that the proposed piecewise linear membership function outperforms the similar membership functions, either triangular, trapezoidal or Γ-linear, in terms of AUC. This proves that the suggested piecewise linear function is more effective in modeling the FOU region with respect to the input (ϵ).

4.6.4. Analysis of modeling uncertainty with respect to the primary membership values (μ-based uncertainty)

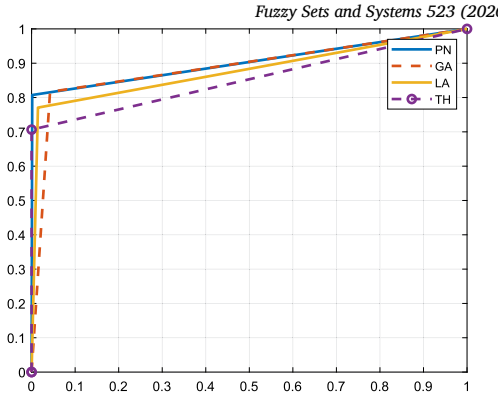
Table 9 illustrates the results of the proposed interval-valued fuzzy sets approach, using nonlinear membership functions, such as the proposed piecewise nonlinear function and other nonlinear membership forms, e.g. Gaussian, Laplacian and tangent-hyperbolic. The results show the contribution of the aforementioned membership functions to improving anomaly detection results, especially using the adequate weights, *i.e.*, those reflecting the test data composition. In particular, the proposed piecewise nonlinear membership function yields the best results, when the normal data weight is very high, *i.e.*, $\omega_0 = 0.9$. In addition, Table 9 and also Fig. 8b show that the proposed piecewise nonlinear membership function gives the best AUC outcome. This proves its effectiveness to model the FOU region with respect to the primary membership, in comparison to other nonlinear functions used to model this type of uncertainty, such as Gaussian, Laplacian and tanh functions.

4.6.5. Analysis of modeling uncertainty with respect to the class weights

The class weights $\{\omega_0, \omega_1\}$ are introduced to reflect the *hypothetic* normal and anomalous data distribution. Their main usefulness consists in weighting the values of the different parameters of the proposed piecewise linear/nonlinear functions, as detailed in Tables 2 and 3. However, since the problem is meant to be unsupervised, the real data distribution is unknown, and so are the values of $\{\omega_0, \omega_1\}$. To this end, Tables 8, 9, and Figs. 9a, 9b show that for both piecewise linear and nonlinear membership functions, it is enough to set $\omega_0 > \omega_1$ to obtain a good AUC result. This may simplify the task, since the class *anomalous* has just to be treated as a minority, without a prior knowledge of its exact proportion.

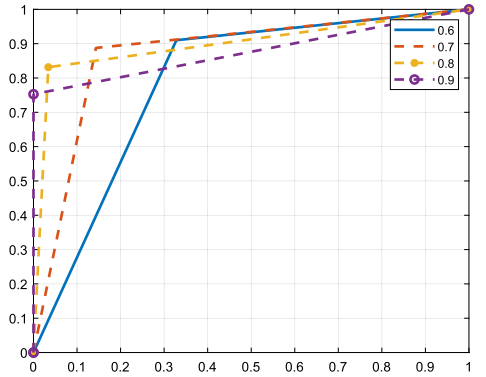


(a) ROC curves for ϵ-based uncertainty using: piecewise linear (PL), triangular (TR), trapezoidal (TP) and Γ-linear (GL) membership functions (cf. Table 2)

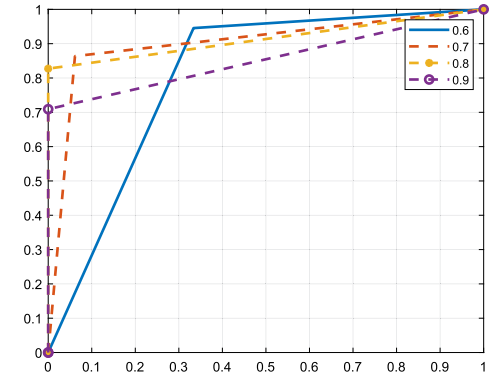


(b) ROC curves for μ-based uncertainty using: piecewise nonlinear (PN), Gaussian (GA), Laplacian (LA) and Tanh (TH) membership functions (cf. Table 3)

Fig. 8. Comparison of ROC curves for the membership function used to model ϵ-based uncertainty in (8a), and μ-based uncertainty in (8b), using different linear and nonlinear membership functions, respectively, with class weights $\omega_0 = 0.9$ and $\omega_1 = 0.1$.



(a) ROC curves for ϵ-based uncertainty using the piecewise linear membership for $\omega_0 \in \{0.6, 0.7, 0.8, 0.9\}$



(b) ROC curves for μ-based uncertainty using the piecewise nonlinear membership for $\omega_0 \in \{0.6, 0.7, 0.8, 0.9\}$

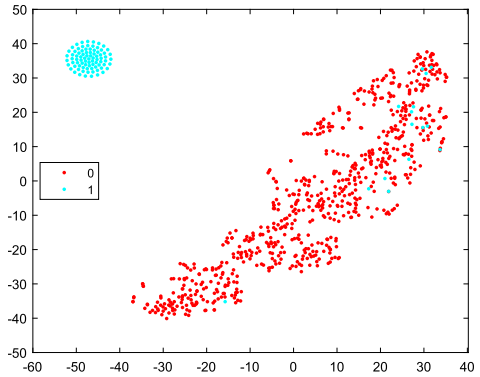
Fig. 9. Comparison of ROC curves for the piecewise linear membership function in (9a) and nonlinear one in (9b) to model ϵ-based and μ-based uncertainty, respectively, using different values of class weights $\{\omega_0, \omega_1\}$ (such that $\omega_1 = 1 - \omega_0$).

4.6.6. *Analysis of the choice and calibration of the membership functions*

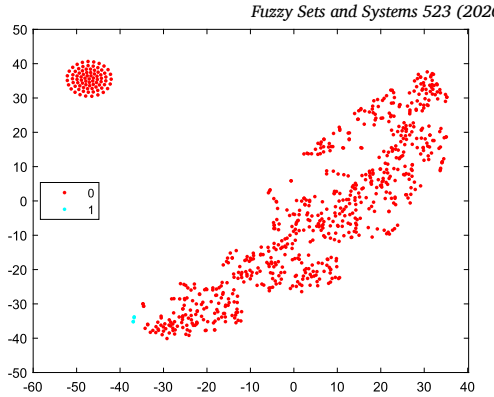
The results in Tables 8 and 9 show that the values of the evaluation metrics depend more on the choice of the membership function used to model uncertainty than on the calibration of the internal parameters. This can be explained as follows:

- In Table 7, it is evident that the use of a membership function, whether stepwise-linear or stepwise-nonlinear, significantly improves the results compared to OC-SVM or the VAE-baseline.
- Within the choice of using a membership to model uncertainty, it is difficult to say how calibration can improve results. In fact, if a membership function returns good metric results, e.g. *Trapezoidal*, Γ-*Linear* and *Piecewise-Linear* in Table 8, the difference between the calibration of different parameters, which depends on the hypothetical normal and anomalous proportions, ω_0 and ω_1, respectively, is not very distinguishable. However, when a membership function does not perform well, e.g. Triangular in Table 8, this behavior is not improved by the use of a different calibration, by varying ω_0 and ω_1.
- The same can be seen in Table 9, where all the membership functions used, *i.e.*, *Gaussian*, *Laplacian*, *Hyperbolic-Tangent* and *Piecewise-Nonlinear* give good results for almost all values of ω_0 and ω_1, thus for all the related parameters $\{a_j, b_j, \alpha_{j,L}, \alpha_{j,U}\}$ $(j \in \{0, 1\})$.

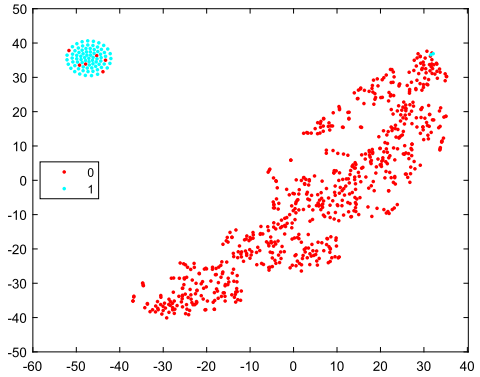
Nevertheless, it should be noted that other membership functions may not give such good results. This shows that the choice of membership function is more important than the calibration of the parameters. This can also be seen as a positive aspect of the



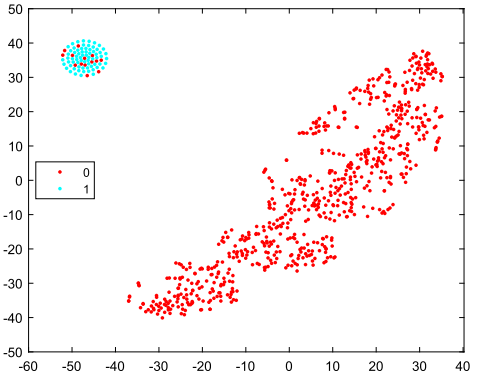
(a) t-SNE distribution of input features with target labels



(b) t-SNE distribution of input features with the labels predicted using the baseline VAE error thresholding method cf. (4,5)



(c) t-SNE distribution of input features with the labels predicted using the piecewise linear membership for ϵ-based uncertainty modeling cf. (8,9) using $\omega_0 = 0.9$



(d) t-SNE distribution of input features with the labels predicted using the piecewise nonlinear membership for μ-based uncertainty modeling cf. (8,9) using $\omega_0 = 0.9$

Fig. 10. t-SNE distribution of input features of the evaluation set using the target and the predicted labels for each of the proposed methods.

proposed method, as the calibration is based on the hypothetical estimation of normal/anomalous proportion, and thus should not affect the quality of the classification results too much.

4.6.7. Analysis of the distribution of input features with respect to the target and predicted labels

Fig. 10 depicts the t-SNE distribution of the input features, *i.e.*, MFCC, Δ-MFCC and Δ-Δ-MFCC on the test set, in correspondence to the target and the predicted labels by each of the proposed methods. First Fig. 10a shows that the target *normal* and *anomalous* clusters are not totally disjoint, which has already been noticed in Fig. 1, and has motivated modeling uncertainty in audio data in this study.

The results show that two separate clusters are obtained. However, only the proposed methods, i.e. using the linear (see Fig. 10c) and non-linear membership functions (see Fig. 10d), succeed in reducing the number of outliers in each cluster. In contrast, the baseline method based on VAE reconstruction, without IVFS, yields two clusters but without any discrimination between the classes (see Fig. 10b).

Therefore, we believe that the proposed method succeeds in two main tasks: Firstly, separating the data into two distinct clusters, where the baseline method fails; secondly, minimizing the number of outliers in each cluster, thus improving the discriminatory power by coping with the uncertainty caused by the presence of background noise.

4.6.8. Concluding remark

In summary, the obtained results attest the relevance of taking into account the fluctuation of the membership function while modeling uncertainty. Actually, the FOU region created by separating the membership function into an upper and lower components proves that interval-valued type-2 fuzzy sets offer a reliable alternative to model uncertainty, whether based on the input, *i.e.*, ϵ-based uncertainty, or on the primary membership, *i.e.*, μ-based uncertainty, in such one-class classification problems.

5. Discussion

Through the analysis of the proposed method and its outcomes, the following comments can be drawn:

5.1. Improvement over the baseline VAE model

The use of the VAE reconstruction error threshold to decide the anomaly in the baseline model is motivated by previous work in the state of the art, particularly for unsupervised/semi-supervised/weakly supervised methods, in which generative networks or a deep/variable autoregressive autoencoder are used for anomaly detection (see Table 1). The main idea of the baseline model is to state that if a sample is anomalous, its reconstructed image will not fit the model, learned only on normal data, and will therefore return a higher reconstruction error. However, while the normal data confirm this hypothesis, as 88% of the normal data confirms this rule for the basic VAE model (see Table 7), the outlier data refute it, with a null precision for the outlier class (see Table 7).

This means that the simple threshold cannot be used to isolate outliers. Therefore, we use the proposed IVFS method based on type-2 fuzzy sets. In this way, the VAE reconstruction error is evaluated not simply using a static threshold, but as an interval of membership values relative to each class, as given by (8)-(9). Then, both intervals are compared to find the most probable class, using (10)-(14).

Thus, the main contribution of combining VAE with IVFS can be summarized as:

- The use of variational autoencoders instead of simple/deep autoencoders provides a further modeling of uncertainty of the input (to the membership function), as the VAE model learns the probability distribution of the signal's features, i.e. the mean and variance, thus quantifying the output reconstruction error as an average/fluctuating value instead of a crisp value.
- Modeling uncertainty by interval-valued type-2 FS requires that both the input and the output of the membership functions be fluctuating. In our case, the input is represented by the VAE reconstruction error. In this case, using VAE – based on modeling the probability distribution of the input features – allows delivering an average / fluctuating reconstruction error instead of a crisp one, as it would have been obtained if a simple AE or OC-SVM were used.

5.2. Effect of parameter calibration

It is obvious that the proposed piecewise linear/nonlinear membership functions help separating between the majority of normal data and the minority of anomalous ones. However, such a result cannot be obtained by simply setting a threshold τ_0 in the baseline VAE error thresholding method, or by just fine-tuning the class weights $\{\omega_0, \omega_1\}$ in the proposed membership functions, but also by taking into account the fluctuation of either the input (ϵ) or the primary membership (μ), using the interval-valued fuzzy sets, as both may be sources of uncertainty.

5.3. Choice of the membership functions

The proposed linear/nonlinear membership functions succeed to boost the anomaly detection scores. Nevertheless, the analysis of the distribution of the input features with respect to the predicted labels reveals that both methods are still unable to detect those *anomalous* samples which share many of their characteristics with normal ones (and that is why they are located in the *normal* cluster). Such shared characteristics must be due to the background noise that predominates a large part of the audio recordings in the dataset (and that should exist in a real-world scenario).

5.4. Effect of modeling uncertainty

It should be noted that modeling uncertainty with respect to the input (ϵ) or to the primary membership (μ) can theoretically be done using linear or nonlinear membership functions, respectively. However, the experiments showed lower results (and therefore were not mentioned) when ϵ-uncertainty was modeled by the nonlinear memberships mentioned in Table 3, and vice-versa, *i.e.*, when linear memberships (cf. Table 2) were used for μ-based uncertainty. This calls for analyzing further the effect of the different functions parameters on such a behavior.

5.5. Effect of imbalanced data

The predominance of the normal class is the typical problem in anomaly detection problems, which explains why most classification methods do not work, mainly because the model is biased towards the dominant class. This is what makes anomaly detection a more complex task than standard classification and requires customized methods. Furthermore, the presence of intrinsic noise in audio signals makes ASD even more complex. This is one of the reasons why we proposed using interval-valued type-2 FS to provide a solution to this problem, as Type-2 FS is designed to model uncertainty, partly due to the presence of noise in all segments, both normal and anomalous.

5.6. Comparison to type-1 fuzzy sets

Type 1 fuzzy sets (T1FS) can be considered as a special case of type 2 fuzzy sets, where the membership function for each class $j \in \{0, 1\}$, is the average of the lower and upper membership functions, as illustrated in Fig. 6. Therefore, the boundaries of the interval are merged, and the event is simply assigned to the class with the highest membership value $Event(i) = \arg\max_{j=0,1}(\mu_j)$. Thus, T1FS do not consider the fluctuation of the input, *i.e.*, the reconstruction error provided by the VAE) or the output *i.e.*, the value of the membership function itself. Consequently, the degree of membership returned by such a T1FS membership function is not a fuzzy number and therefore cannot be used to assess uncertainty.

5.7. General remarks

Finally, it should be emphasized that the primary aim of this work is not showing that the proposed piecewise linear/nonlinear membership functions are performing better than the other basic ones, e.g. Gaussian, Laplacian, etc., but rather to demonstrate the effectiveness of using interval-valued fuzzy sets for modeling uncertainty in improving anomaly detection in such noisy data. Therefore we believe that the added value of the proposed method consists in making basic membership functions, such as those mentioned on Table 2 and Table 3, able to model uncertainty with respect to both dimensions, *i.e.*, the input (ϵ) and the primary membership (μ), through the use of interval-valued fuzzy sets.

6. Conclusion

Based on interval-valued fuzzy sets (IVFS), this paper suggested a novel approach of anomaly detection, that takes care of modeling uncertainty when input data are highly noisy. An immediate application to audio road traffic surveillance allows detecting hazardous events such as car accidents. The proposed approach is based on integrating two anomaly detection methods, namely variational autoencoders (VAE) and interval-valued fuzzy sets (IVFS). First, a VAE model is trained on data belonging to the class *normal* only, and then the VAE reconstruction error is used to compute a lower/pessimistic component and an upper/optimistic component, which serve to generate the interval-valued fuzzy membership function for each class, namely *normal* and *anomalous*. Finally, for defuzzification, *i.e.*, detecting the corresponding class, a probabilistic interval comparison method, known as degree of preference, is used. As a result, the class *anomalous* corresponds to the least preferred/smallest interval. To model uncertainty, the lower and upper membership function components are adjusted to evaluate uncertainty with respect to either the first dimension, *i.e.*, the input VAE reconstruction error (ϵ) or the second dimension, *i.e.*, the membership value ($\mu(\epsilon)$). Standard classification criteria such as overall accuracy, precision, recall, F1-score were used for evaluation, in addition to unsupervised learning-specific metrics such as AUC and $p\text{-}AUC$.

The results yielding from the various experiments allow stating the following comments:

- From an audio signal processing standpoint, standard audio spectrogram-extracted features, *i.e.*, MFCC, Δ-MFCC and Δ-Δ-MFCC, are the best fitted to approach such a problem.
- IVFS appear to be more efficient than crisp one-class SVM at detecting anomaly.
- The double evaluation of uncertainty, *i.e.*, based on the VAE reconstruction error (ϵ) or on the primary membership value (μ), allows modeling uncertainty in audio signals at different levels, due to either the variability of the input features, in case of ϵ-based uncertainty, or the ambiguity of modeling classes for audio signals in a noisy environment, in case of μ-based uncertainty.

As for the comparison with crisp classification in general, and one-class SVM in particular, we notice that modeling uncertainty with interval-valued type-2 FS allows modeling the input's uncertainty, *i.e.*, the VAE's reconstruction error, which is primarily due to the input data, including the inherent noise, and by the output's uncertainty, *i.e.*, the fluctuation of the membership function, represented as an interval. Such a modeling choice in totally absent in crisp/hard classification methods such as OC-SVM, and partially ignored in type-1 FS, where only the uncertainty of the input is considered, whereas the output's uncertainty, *i.e.*, the membership function, is not taken in account.

Finally, we believe that in addition to the specific task of anomalous SED, the obtained performance can be used to address the uncertainty problem that characterizes audio data distribution. In fact, most spontaneous sounds, whether normal or anomalous, are naturally contaminated by background noise, making crisp approaches less efficient in creating a discriminative model. As a result, in such a real-world scenario, IVFS represent an alternative for dealing with uncertainty in audio signals. In the future, the proposed method can be improved to be completely unsupervised, by removing any *a priori* knowledge or hypothesis about the use of labels and class weights.

CRediT authorship contribution statement

Zied Mnasri: Writing – original draft, Investigation, Formal analysis, Data curation, Conceptualization. **Stefano Rovetta:** Writing – review & editing, Project administration, Methodology. **Francesco Masulli:** Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work has been supported by the University of Genova, Italy, and the University of Tunis El Manar, Tunisia.

Data availability

Data has been made available as mentioned in the manuscript

References

[1] MIVIA dataset, https://mivia.unisa.it/datasets/audio-analysis/mivia-road-audio-events-data-set/. (Accessed 8 July 2024).
[2] S. Adavanne, T. Virtanen, A report on sound event detection with different binaural features, in: Proceedings of the Detection and Classification of Acoustic Scenes and Events 2017 Workshop (DCASE2017), Munich, Germany, November 2020.
[3] R.M. Alsina-Pagès, F. Orga, F. Alías, J.C. Socoró, A wasn-based suburban dataset for anomalous noise event detection on dynamic road-traffic noise mapping, Sensors 19 (11) (2019) 2480.
[4] F. Aurino, M. Folla, F. Gargiulo, V. Moscato, A. Picariello, C. Sansone, One-class svm based approach for detecting anomalous audio events, in: 2014 International Conference on Intelligent Networking and Collaborative Systems, IEEE, 2014, pp. 145–151.
[5] R. Babuška, H.B. Verbruggen, An overview of fuzzy modeling for control, Control Eng. Pract. 4 (11) (1996) 1593–1606.
[6] N. Borges, G.G. Meyer, Unsupervised distributional anomaly detection for a self-diagnostic speech activity detector, in: 2008 42nd Annual Conference on Information Sciences and Systems, IEEE, 2008, pp. 950–955.
[7] N. Borges, G.G. Meyer, Trimmed KL divergence between Gaussian mixtures for robust unsupervised acoustic anomaly detection, in: Interspeech 2009, IEEE, 2009, pp. 2555–2558.
[8] H. Bustince, Interval-valued fuzzy sets in soft computing, Int. J. Comput. Intell. Syst. 3 (2) (2010) 215–222.
[9] H. Bustince, J. Fernandez, H. Hagras, F. Herrera, M. Pagola, E. Barrenechea, Interval type-2 fuzzy sets are generalization of interval-valued fuzzy sets: toward a wider view on their relationship, IEEE Trans. Fuzzy Syst. 23 (5) (2014) 1876–1882.
[10] C. Chen, P. Chen, L. Yang, J. Mo, H. Song, Y. Xie, L. Ma, Acoustic anomaly detection via latent regularized Gaussian mixture generative adversarial networks, in: Proceedings of the Detection and Classification of Acoustic Scenes and Events 2020 Workshop (DCASE2020), Tokyo, Japan, November 2020, http://dcase.community/challenge2020/index, Preprint: https://arxiv.org/pdf/2002.01107.pdf.
[11] Z. Chen, C.K. Yeo, B.S. Lee, C.T. Lau, Autoencoder-based network anomaly detection, in: 2018 Wireless Telecommunications Symposium (WTS), IEEE, 2018, pp. 1–5.
[12] A. Dang, T.H. Vu, J.-C. Wang, Deep learning for dcase2017 challenge, Workshop on DCASE2017 Challenge, Tech. Rep. 2017.
[13] S.M. Erfani, S. Rajasegarar, S. Karunasekera, C. Leckie, High-dimensional and large-scale anomaly detection using a linear one-class svm with deep learning, Pattern Recognit. 58 (2016) 121–134.
[14] P.F. Evangelista, M.J. Embrechts, B.K. Szymanski, Taming the curse of dimensionality in kernels and novelty detection, in: Applied Soft Computing Technologies: The Challenge of Complexity, Springer, 2006, pp. 425–438.
[15] P. Foggia, N. Petkov, A. Saggese, N. Strisciuglio, M. Vento, Audio surveillance of roads: a system for detecting anomalous sounds, IEEE Trans. Intell. Transp. Syst. 17 (1) (2015) 279–288.
[16] P. Foggia, A. Saggese, N. Strisciuglio, M. Vento, N. Petkov, Car crashes detection by audio analysis in crowded roads, in: 2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), IEEE, 2015, pp. 1–6.
[17] A.B. Gardner, A.M. Krieger, G. Vachtsevanos, B. Litt, One-class novelty detection for seizure analysis from intracranial eeg, J. Mach. Learn. Res. 7 (Jun 2006) 1025–1044.
[18] R. Giri, F. Cheng, K. Helwani, S.V. Tenneti, U. Isik, A. Krishnaswamy, Group masked autoencoder based density estimator for audio anomaly detection, in: Proceedings of the Detection and Classification of Acoustic Scenes and Events 2020 Workshop (DCASE2020), Tokyo, Japan, November 2020, pp. 51–55, http://dcase.community/documents/workshop2020/proceedings/DCASE2020Workshop_Giri_66.pdf.
[19] A. Greco, N. Petkov, A. Saggese, M. Vento, Aren: a deep learning approach for sound event recognition using a brain inspired representation, IEEE Trans. Inf. Forensics Secur. 15 (2020) 3610–3624.
[20] A. Greco, A. Roberto, A. Saggese, M. Vento, Denet: a deep architecture for audio surveillance applications, Neural Comput. Appl. (2021) 1–12.
[21] M. Hao, J.M. Mendel, Similarity measures for general type-2 fuzzy sets based on the α-plane representation, Inf. Sci. 277 (2014) 197–215.
[22] P. Hayton, S. Utete, D. King, S. King, P. Anuzis, L. Tarassenko, Static and dynamic novelty detection methods for jet engine health monitoring, Philos. Trans. R. Soc. A, Math. Phys. Eng. Sci. 365 (1851) (2007) 493–514.
[23] K.-X. He, W.-Q. Zhang, J. Liu, Y. Liu, Dilated-gated convolutional neural network with a new loss function on sound event detection, in: 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), IEEE, 2019, pp. 1491–1495.
[24] T. Heittola, A. Mesaros, A. Eronen, T. Virtanen, Context-dependent sound event detection, EURASIP J. Audio Speech Music Process. 2013 (1) (2013) 1.
[25] K. Heller, K. Svore, A.D. Keromytis, S. Stolfo, One Class Support Vector Machines for Detecting Anomalous Windows Registry Accesses, 2003.
[26] V.-N. Huynh, Y. Nakamori, J. Lawry, A probability-based approach to comparison of fuzzy numbers and applications to target-oriented decision making, IEEE Trans. Fuzzy Syst. 16 (2) (2008) 371–387.
[27] K. Imoto, N. Tonami, Y. Koizumi, M. Yasuda, R. Yamanishi, Y. Yamashita, Sound event detection by multitask learning of sound events and scenes with soft scene labels, in: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2020, pp. 621–625.
[28] M. Jiménez, Ranking fuzzy numbers through the comparison of its expected intervals, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 4 (04) (1996) 379–388.
[29] C.-C. Kao, W. Wang, M. Sun, C. Wang, R-crnn: region-based convolutional recurrent neural network for audio event detection, arXiv preprint arXiv:1808.06627, 2018.
[30] C.-C. Kao, M. Sun, W. Wang, C. Wang, A comparison of pooling methods on lstm models for rare acoustic event classification, in: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2020, pp. 316–320.
[31] P. Karczmarek, A. Kiersztyn, W. Pedrycz, E. Al, K-means-based isolation forest, Knowl.-Based Syst. 195 (2020) 105659.
[32] A. Kasperski, A possibilistic approach to sequencing problems with fuzzy parameters, Fuzzy Sets Syst. 150 (1) (2005) 77–86.
[33] Y. Kawachi, Y. Koizumi, N. Harada, Complementary set variational autoencoder for supervised anomaly detection, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2018, pp. 2366–2370.

[34] Y. Koizumi, S. Murata, N. Harada, S. Saito, H. Uematsu, Sniper: few-shot learning for anomaly detection to minimize false-negative rate with ensured true-positive rate, in: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019, pp. 915–919.
[35] Y. Koizumi, M. Yasuda, S. Murata, S. Saito, H. Uematsu, N. Harada, Spidernet: attention network for one-shot anomaly detection in sounds, in: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2020, pp. 281–285.
[36] H.-P. Kriegel, P. Kröger, J. Sander, A. Zimek, Density-based clustering, Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 1 (3) (2011) 231–240.
[37] E. Lee, R.-J. Li, Comparison of fuzzy numbers based on the probability measure of fuzzy events, Comput. Math. Appl. 15 (10) (1988) 887–896.
[38] Y. Li, X. Li, The seie-scut systems for ieee aasp challenge on dcase 2017: deep learning techniques for audio representation and classification, in: Proc. Detection Classification Acoustic Scenes Events 2018 Workshop, 2017.
[39] H. Lim, J. Park, Y. Han, Rare sound event detection using 1d convolutional recurrent neural networks, in: Proceedings of the Detection and Classification of Acoustic Scenes and Events 2017 Workshop, 2017, pp. 80–84.
[40] F.T. Liu, K.M. Ting, Z.-H. Zhou, Isolation-based anomaly detection, ACM Trans. Knowl. Discov. Data 6 (1) (2012) 1–39.
[41] W. Liu, R. Li, M. Zheng, S. Karanam, Z. Wu, B. Bhanu, R.J. Radke, O. Camps, Towards visually explaining variational autoencoders, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 8642–8651.
[42] J. Ma, S. Perkins, Time-series novelty detection using one-class support vector machines, in: Proceedings of the International Joint Conference on Neural Networks, 2003, vol. 3, IEEE, 2003, pp. 1741–1745.
[43] M. Mandel, J. Salamon, D.P.W. Ellis, Proceedings of the Detection and Classification of Acoustic Scenes and Events 2019 Workshop (DCASE2019), New York University, NY, USA, October 2019.
[44] E. Marchi, F. Vesperini, S. Squartini, B. Schuller, Deep recurrent neural network-based autoencoders for acoustic novelty detection, Comput. Intell. Neurosci. 2017 (2017).
[45] L. Melgar-García, M. Hosseini, A. Troncoso, Identification of anomalies in urban sound data with autoencoders, in: International Conference on Hybrid Artificial Intelligence Systems, Springer, 2023, pp. 27–38.
[46] J.M. Mendel, General type-2 fuzzy logic systems made simple: a tutorial, IEEE Trans. Fuzzy Syst. 22 (5) (2013) 1162–1182.
[47] J.M. Mendel, R.I.B. John, Type-2 fuzzy sets made simple, IEEE Trans. Fuzzy Syst. 10 (2) (2002) 117–127, https://doi.org/10.1109/91.995115.
[48] J.M. Mendel, H. Hagras, H. Bustince, F. Herrera, Comments on "interval type-2 fuzzy sets are generalization of interval-valued fuzzy sets: towards a wide view on their relationship", IEEE Trans. Fuzzy Syst. 24 (1) (2015) 249–250.
[49] Z. Mnasri, S. Rovetta, F. Masulli, Audio surveillance of roads using deep learning and autoencoder-based sample weight initialization, in: 2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON), IEEE, 2020, pp. 99–103.
[50] Z. Mnasri, S. Rovetta, F. Masulli, Anomalous sound event detection: a survey of machine learning based methods and applications, Multimed. Tools Appl. (2021) 1–50.
[51] Z. Mnasri, S. Rovetta, F. Masulli, A. Cabri, Dealing with uncertainty in anomalous audio event detection using fuzzy modeling, in: Advances in Computational Intelligence Systems: Contributions Presented at the 20th UK Workshop on Computational Intelligence, September 8–10, 2021, Aberystwyth, Wales, UK 20, Springer, 2022, pp. 496–507.
[52] Z. Mnasri, S. Rovetta, F. Masulli, Anomalous sound event detection based on one-class classification using variational autoencoders and interval type-2 fuzzy sets, in: 2023 31st European Signal Processing Conference (EUSIPCO), IEEE, 2023, pp. 171–175.
[53] B. Nachman, D. Shih, Anomaly detection with density estimation, Phys. Rev. D 101 (7) (2020) 075042.
[54] Andrew Ng, Sparse autoencoder, https://web.stanford.edu/class/cs294a/sparseAutoencoder_2011new.pdf, 2011. (Accessed 29 March 2020), Online.
[55] H.-L. Nguyen, Y.-K. Woon, W.-K. Ng, A survey on data stream clustering and classification, Knowl. Inf. Syst. 45 (3) (2015) 535–569.
[56] S. Ntalampiras, I. Potamitis, N. Fakotakis, Probabilistic novelty detection for acoustic surveillance under real-world conditions, IEEE Trans. Multimed. 13 (4) (2011) 713–719.
[57] H. Phan, M. Krawczyk-Becker, T. Gerkmann, A. Mertins, Dnn and cnn with weighted and multi-task loss functions for audio event detection, in: Proc. DCASE 2017-Workshop Detect. Classification Acoust. Scenes Events, 2017.
[58] H. Phan, M. Krawczyk-Becker, T. Gerkmann, A. Mertins, Weighted and multi-task loss for rare audio event detection, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2018, pp. 336–340.
[59] H. Phan, O.Y. Chén, P. Koch, L. Pham, I. McLoughlin, A. Mertins, M. De Vos, Unifying isolated and overlapping audio event detection with multi-label multi-task convolutional recurrent neural networks, in: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019, pp. 51–55.
[60] M.A. Pimentel, D.A. Clifton, L. Clifton, L. Tarassenko, A review of novelty detection, Signal Process. 99 (2014) 215–249.
[61] H. Purohit, R. Tanabe, T. Endo, K. Suefusa, Y. Nikaido, Y. Kawaguchi, Deep autoencoding gmm-based unsupervised anomaly detection in acoustic signals and its hyper-parameter optimization, in: Proceedings of the Detection and Classification of Acoustic Scenes and Events 2020 Workshop (DCASE2020), Tokyo, Japan, November 2020, http://dcase.community/challenge2020/index, Preprint: https://arxiv.org/pdf/2009.12042.pdf.
[62] L.R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proc. IEEE 77 (2) (1989) 257–286.
[63] L.R. Rabiner, B.H. Juang, Fundamentals of Speech Recognition, Prentice-Hall, Inc., ISBN 0-13-015157-2, 1993.
[64] P. Rai, H. Daumé III, S. Venkatasubramanian, Streamed learning: one-pass svms, arXiv preprint arXiv:0908.0572, 2009.
[65] D.A. Reynolds, T.F. Quatieri, R.B. Dunn, Speaker verification using adapted Gaussian mixture models, Digit. Signal Process. 10 (1–3) (2000) 19–41.
[66] S. Rovetta, Z. Mnasri, F. Masulli, Detection of hazardous road events from audio streams: an ensemble outlier detection approach, in: 2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS), IEEE, 2020, pp. 1–6.
[67] S. Rovetta, Z. Mnasri, F. Masulli, A. Cabri, Anomaly detection based on interval-valued fuzzy sets: application to rare sound event detection, in: The 13th International Workshop on Fuzzy Logic and Applications (WILF 2021), CEUR-WS.org, 2021, pp. 1–8.
[68] E. Rushe, B. Mac Namee, Anomaly detection in raw audio using deep autoregressive networks, in: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019, pp. 3597–3601.
[69] A. Saggese, N. Strisciuglio, M. Vento, N. Petkov, Time-frequency analysis for audio event detection in real scenarios, in: 2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), IEEE, 2016, pp. 438–443.
[70] M. Sakurada, T. Yairi, Anomaly detection using autoencoders with nonlinear dimensionality reduction, in: Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis, 2014, pp. 4–11.
[71] M. Sammarco, M. Detyniecki, Crashzam: sound-based car crash detection, in: Proceedings of Vehicle Technology and Intelligent Transport Systems (VEHITS), 2018, pp. 27–35.
[72] B. Schölkopf, R.C. Williamson, A. Smola, J. Shawe-Taylor, J. Platt, Support vector method for novelty detection, Adv. Neural Inf. Process. Syst. 12 (1999) 582–588.
[73] P. Sevastianov, Numerical methods for interval and fuzzy number comparison based on the probabilistic approach and Dempster–Shafer theory, Inf. Sci. 177 (21) (2007) 4645–4661.
[74] K. Shimada, Y. Koyama, A. Inoue, Metric learning with background noise class for few-shot detection of rare sound events, in: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2020, pp. 616–620.
[75] N. Shvetsova, B. Bakker, I. Fedulova, H. Schulz, D.V. Dylov, Anomaly detection in medical imaging with deep perceptual autoencoders, IEEE Access 9 (2021) 118571–118583.

[76] A.A. Sodemann, M.P. Ross, B.J. Borghetti, A review of anomaly detection in automated surveillance, IEEE Trans. Syst. Man Cybern., Part C, Appl. Rev. 42 (6) (2012) 1257–1272.
[77] Y. Song, Z. Wen, C.-Y. Lin, R. Davis, One-Class Conditional Random Fields for Sequential Anomaly Detection, 2013.
[78] D. Stowell, D. Giannoulis, E. Benetos, M. Lagrange, M.D. Plumbley, Detection and classification of acoustic scenes and events, IEEE Trans. Multimed. 17 (10) (2015) 1733–1746.
[79] N. Strisciuglio, M. Vento, N. Petkov, Learning representations of sound using trainable cope feature extractors, Pattern Recognit. 92 (2019) 25–36.
[80] L.V. Utkin, Y.A. Zhuk, An one-class classification support vector machine model by interval-valued training data, Knowl.-Based Syst. 120 (2017) 43–56.
[81] F. Vesperini, D. Droghini, D. Ferretti, E. Principi, L. Gabrielli, S. Squartini, F. Piazza, A hierarchic multi-scaled approach for rare sound event detection, in: Proc. DCASE 2017-Workshop Detect. Classification Acoust. Scenes Events, 2017.
[82] T. Virtanen, A. Mesaros, T. Heittola, M. Plumbley, P. Foster, E. Benetos, M. Lagrange, Proceedings of the Detection and Classification of Acoustic Scenes and Events 2016 Workshop (DCASE2016), Tampere University of Technology, Department of Signal Processing, 2016.
[83] T. Virtanen, A. Mesaros, T. Heittola, A. Diment, E. Vincent, E. Benetos, B.M. Elizalde, Proceedings of the Detection and Classification of Acoustic Scenes and Events 2017 Workshop (DCASE2017), Tampere University of Technology, Laboratory of Signal Processing, 2017.
[84] C. Wagner, H. Hagras, Toward general type-2 fuzzy logic systems based on zslices, IEEE Trans. Fuzzy Syst. 18 (4) (2010) 637–660.
[85] C. Wagner, S. Miller, J.M. Garibaldi, D.T. Anderson, T.C. Havens, From interval-valued data to general type-2 fuzzy sets, IEEE Trans. Fuzzy Syst. 23 (2) (2014) 248–269.
[86] X. Wang, E.E. Kerre, Reasonable properties for the ordering of fuzzy quantities (i), Fuzzy Sets Syst. 118 (3) (2001) 375–385.
[87] X. Wang, E.E. Kerre, Reasonable properties for the ordering of fuzzy quantities (ii), Fuzzy Sets Syst. 118 (3) (2001) 387–405.
[88] Y.-M. Wang, J.-B. Yang, D.-L. Xu, A preference aggregation method through the estimation of utility intervals, Comput. Oper. Res. 32 (8) (2005) 2027–2049.
[89] Q. Wei, Y. Liu, Auto-Encoder and Metric-Learning for Anomalous Sound Detection Task, November 2020, http://dcase.community/challenge2020/index, Preprint: http://dcase.community/documents/challenge2020/technical_reports/DCASE2020_Wei_49_t2.pdf.
[90] X. Xia, R. Togneri, F. Sohel, Y. Zhao, D. Huang, Multi-task learning for acoustic event detection using event and frame position information, IEEE Trans. Multimed. 22 (3) (2019) 569–578.
[91] Y. Xiao, B. Liu, S.Y. Philip, Z. Hao, A robust one-class transfer learning method with uncertain data, Knowl. Inf. Syst. 44 (2) (2015) 407–438.
[92] C.-H. Yeh, H. Deng, A practical approach to fuzzy utilities comparison in fuzzy multicriteria analysis, Int. J. Approx. Reason. 35 (2) (2004) 179–194.
[93] Y. Zhao, B. Deng, C. Shen, Y. Liu, H. Lu, X.-S. Hua, Spatio-temporal autoencoder for video anomaly detection, in: Proceedings of the 25th ACM International Conference on Multimedia, 2017, pp. 1933–1941.