

# PRIVAtE: Passive Radar Interpretability using Variational Auto Encoders

Marco Cominelli\*, Paolo Braca<sup>†</sup>, Leonardo M. Millefiori<sup>†</sup>, Lance M. Kaplan<sup>‡</sup>, Mani B. Srivastava<sup>§</sup>

Francesco Gringoli\* and Federico Cerutti\*<sup>¶</sup>

\*University of Brescia, Italy

<sup>†</sup>NATO STO Centre for Maritime Research and Experimentation (CMRE), Italy

<sup>‡</sup>DEVCOM Army Research Laboratory, USA

<sup>§</sup> University of California Los Angeles, USA

<sup>¶</sup>Cardiff University, UK

**Abstract**—This paper aims to present a method to increase interpretability human analysts can have in actionable intelligence from analysing Wi-Fi signals used as passive radar systems for situational understanding. The Passive Radar Interpretability using Variational Auto Encoders (PRIVAtE) method is demonstrated using a recent dataset that estimates the latent distributions of antennas of the same Wi-Fi receiver to perform human activity recognition. The performance theoretical analysis of machine learning binary classification includes error probabilities using a statistical test based on a classification learned in the training phase. Results demonstrate that the compressed data obtained using the Variational Auto-Encoder is statistically very informative for providing situational understanding.

**Index Terms**—passive radar, intelligence analysis, artificial intelligence

## I. INTRODUCTION

Suppose an autonomous agent needs to monitor an environment for security purposes continuously, *e.g.*, a restricted facility that only authorised personnel can access or a dangerous environment where only non-human personnel can enter, such as some areas of a chemical plant. Such a monitoring task could utilise wireless signals from Wi-Fi transmitters already present in the environment as signals of opportunity in a passive radar setup.

Indeed, Wi-Fi devices can effectively be used as passive radar systems that *sense* the surroundings, possibly even discern human activity [1]. In [2], some of the authors of this paper proposed a principled analysis of a recently captured dataset [3] of human activities sensed by commercial Wi-Fi devices. The approach relies on Variational Auto-Encoders (VAEs) [4] for identifying generative relationships with a latent distribution, which can be seen as a compressed view of the original signal (Section II). Indeed, in [3] a method is introduced for carrying out human activity recognition (HAR) by analysing channel state information (CSI)—a measurement

of the properties of the wireless channel. This is achieved using commercial Wi-Fi devices, with the primary objective of detecting and categorising distinct activities executed by an individual within a room.

However, in order for a user to interpret such a system’s classifications results, it is vital to provide an estimate of the classification error probabilities [5]. In this paper, building up on recent work by some of the authors (*i.e.*, [6], [7]), we provide a preliminary analysis of the asymptotic properties of classification performance in HAR applications. The theoretical foundations for the asymptotic analysis of the classification performance are reported in Section III.

In Section IV, we provide a first analysis of the dataset introduced in [3] and analysed in [2] using the statistical analysis technique introduced in [6] and [7]. The results show that the compressed data obtained using VAEs are statistically very informative for providing situational understanding. The presented PRIVAtE method supports remarks suggested in [2] by visual inspection of some graphical representation of latent space distributions, providing a statistically sound methodology for assessing the trustworthiness of the overall HAR system. Section V provides a brief overview of the current trends in CSI-based sensing, highlighting the novelty of the proposed approach. Finally, Section VI concludes the paper.

## II. ANALYSIS OF A SITUATIONAL AWARENESS SCENARIO USING VARIATIONAL AUTO-ENCODERS

### A. Scenario Overview

The work in [3] analyses the CSI collected from a commercial Wi-Fi receiver for HAR applications. Essentially, the CSI encapsulates information about how wireless signals interact with the surrounding environment, manifesting as distinctive interference patterns influenced by factors like room layout, furniture configuration, and even the presence and movement of individuals. Each CSI acts as a unique electromagnetic signature of the surroundings, enabling Wi-Fi receivers to be likened to passive radars under suitable assumptions.

The CSI pertains to the assessment of wireless channel characteristics, calculated by the receiver for each incoming

Thanks to Erik Blasch for participating in the project. The work is partially supported by the European Office of Aerospace Research & Development (EOARD) and the Air Force Office of Scientific Research (AFOSR) under award numbers FA8655-22-1-7017 and FA9550-22-1-0193, and by the US DEVCOM Army Research Laboratory (ARL) under Cooperative Agreement #W911NF2220243. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the United States government.

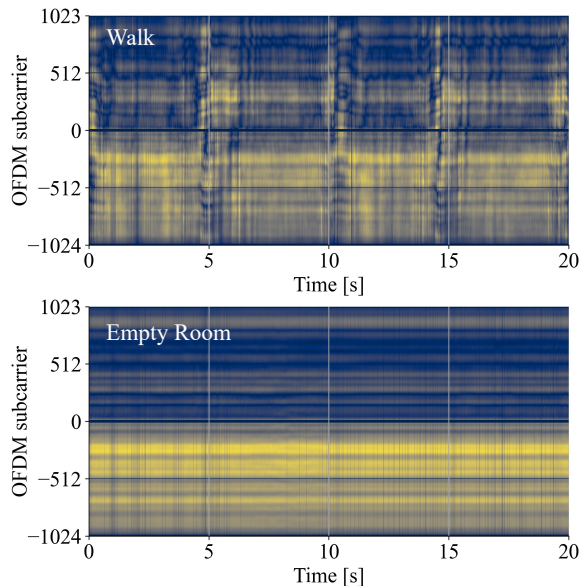


Fig. 1: Visualisation of the CSI magnitude collected by one antenna for two different activities (*Walk* and *Empty Room*) in the same room. Brighter colours represent larger values.

Wi-Fi frame. When considering narrow-band signals, the received signal  $y$  is usually described in the frequency domain as the product of the channel frequency response  $h$  with the transmitted signal  $x$  (we disregard noise in this brief overview). In practice, the channel frequency response  $h$  is a function of both the frequency  $f$  and the time  $t$ ; hence, in wide-band orthogonal frequency-division multiplexing (OFDM) communication systems—like Wi-Fi—the CSI is a crucial element that enables the correction of frequency-selective distortions. Indeed, the CSI  $\mathcal{H}$  is precisely the receiver’s estimate of the channel response  $h(f, t)$  obtained by using a reference signal in the preamble of every Wi-Fi frame [8]. Under the hypothesis that the entire Wi-Fi frame fits within the coherence time of the channel,  $\mathcal{H}$  can be estimated on the preamble and then used to equalise the payload of the frame. However, the CSI can also be used to estimate and track signal variations caused by the multipath induced by the physical surroundings, like a person or an object moving, as depicted in Fig. 1. Therefore, it should be clear that CSI analysis represents an essential element for various Wi-Fi sensing applications [9].

In this study, we rely on a CSI dataset recently made public for HAR [3]. The experimental setup consists of a pair of Asus RT-AX86U routers, each outfitted with four antennas and endowed with multiple-input multiple-output (MIMO) capabilities. Throughout the experimental trials, one of the routers generates dummy Wi-Fi traffic consistently at a fixed rate of 150 frames per second, utilising the injection feature provided by AX-CSI [9]. Meanwhile, the other router, also known as the *monitor*, utilises the same software tool to engage in *sensing*. The sensing process involves the collection of CSI from the received test Wi-Fi frames. Similarly to [2], the focus of this paper is on a singular scenario from the dataset,

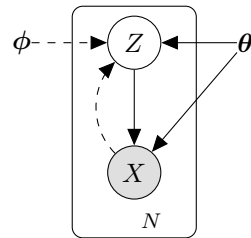


Fig. 2: Plate notation of a VAE [4], with dashed lines denoting the variational approximation.  $\theta$  are the true yet unknown parameters of the  $Z$  distribution which generated the data  $X$ , while  $\phi$  are the learnt parameters.

in which an individual carries out different activities within an indoor space measuring approximately 45 m<sup>2</sup>. For every distinct activity, the monitor accumulates 80 s of CSI data.

Just like in [2], our assumption is that the target individual can engage in only a subset of the potential activities outlined in the original dataset [3], namely: *walk*, *run*, *jump*, *sit*, and *empty room*. We opt for these activities as they constitute the fundamental categories of conceivable actions. These have also been taken into account in related research pertaining to CSI-based HAR (see [3], [10], [11]).

## B. Activity Classification with Variational Auto-Encoders

The CSI, a complex vector estimating the wireless channel’s frequency response, is approached in a simplified manner as suggested in [2]. We discard the phase information and focus solely on the magnitude of the CSI to streamline the process. Using the signal magnitude as a sequence of CSI can be likened to a spectrogram, which we normalise in amplitude relative to the highest value found in the entire dataset. Subsequently, we employ a 3-second sliding window to extract the samples to train four VAEs: VAE-A1, VAE-A2, VAE-A3, and VAE-A4. Each of these models is fed the CSI data from its respective antennas, identified by the model number.

For each of these VAEs — *cf.*, Figure 2 — we make the assumption that the prior distribution over  $Z$  follows a centred isotropic bivariate Gaussian  $p(\mathbf{z}) = \mathcal{N}(\mathbf{z} | \mathbf{0}, \mathbf{I})$ , while  $P(X|Z)$  conforms to a bivariate Gaussian distribution. To approximate the posterior, we consider that the true, yet impractical, posterior can be approximated by a Gaussian form with an approximately diagonal covariance structure. In this context, the variational approximation of the posterior can be defined as a bivariate Gaussian distribution with a diagonal covariance matrix:

$$p(\mathbf{z} | \mathbf{x}, \phi) = \mathcal{N}\left(\mathbf{z} | \boldsymbol{\mu}^{(i)}, \text{diag}\left(\boldsymbol{\sigma}^{2(i)}\right)\right) \quad (1)$$

Here, the mean  $\boldsymbol{\mu}^{(i)}$  and standard deviation  $\boldsymbol{\sigma}^{(i)}$  of the approximate posterior are outputs derived from the encoder.

The results presented in [2] show that a multi-layer perceptron (MLP) processing the latent space information learned from the four VAEs can recognise the human activity with a

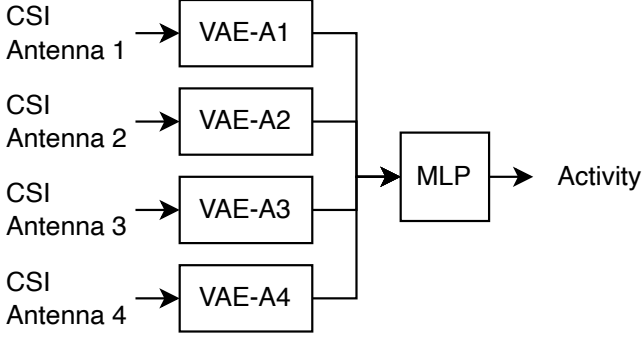


Fig. 3: Architecture proposed in [2] for CSI-based human activity recognition. Each VAE processes the CSI of a separate antenna of the Wi-Fi receiver; then, the latent parameters are aggregated and classified by an MLP.

test accuracy of around 95%. Figure 3 illustrates schematically the proposed architecture. After the VAEs have processed the CSI, their latent space parameters are concatenated into a single vector that becomes the input of the MLP.

### III. PERFORMANCE OF DATA-DRIVEN CLASSIFIERS

In [6], [7], some of the authors of this paper introduce two mathematical frameworks for assessing the efficacy of a general binary machine learning classifier in relation to asymptotic error probabilities. These frameworks draw upon established concepts such as the Central Limit Theorem (CLT) [12] and the Large Deviations Principle (LDP) [13].

Drawing inspiration from [6], [7], let's delve into a scenario where a family of real-valued decision statistics, denoted as  $T^{(n)}$ , processes a sequence of observations  $\mathcal{X}^{(n)}$ , that in our context are provided by the different CSI antennas of the Wi-Fi receiver. With reference to Figure 3, in the following we consider  $n = 4$  as there are four different antennas witnessing the same activity, *cf.*, [2].

The objective revolves around distinguishing between two hypotheses:  $\mathcal{H}_0$  and  $\mathcal{H}_1$ , wherein the observations can be distributed according to  $f_0(\cdot)$  under  $\mathcal{H}_0$ , or  $f_1(\cdot)$  under  $\mathcal{H}_1$ . Considering that the distributions under  $\mathcal{H}_0$  and  $\mathcal{H}_1$  are often either unknown or excessively intricate to derive, we then shift our attention to a scenario where the decision statistic emerges from a learning mechanism operating on a substantial, finite, labelled “training set”  $\mathcal{Y}$ , *i.e.*, a dataset separate from  $\mathcal{X}^{(n)}$  but available for each hypothesis. The decision statistic, denoted as  $T_\omega^{(n)}$ , is called Data Driven Decision Function (D3F) and is characterised by a parameter set  $\omega$  acquired during the training phase.

In our context the D3F is essentially an MLP trained with a binary cross-entropy loss function and a uniform prior (typically in a balanced training set). The D3F is then defined as a difference of logarithmic posterior probabilities as depicted in Equation (2):

$$T_\omega^{(n)} = \frac{1}{n} \left( \log p_\omega^{(\mathcal{H}_1)}(\mathcal{X}^{(n)}) - \log p_\omega^{(\mathcal{H}_0)}(\mathcal{X}^{(n)}) \right), \quad (2)$$

where  $p_\omega^{(\mathcal{H}_k)}$ ,  $k = 0, 1$ , are the outputs of the MLP, interpreted as approximations of the posterior hypothesis probabilities, and  $\mathcal{X}^{(n)}$  is the block of outputs provided by the VAEs.

The statistical test is defined by comparing  $T_\omega^{(n)}$  with a decision threshold  $\gamma_n$ , *i.e.*,

$$\begin{cases} T_\omega^{(n)} \geq \gamma_n : & \text{decide } \mathcal{H}_1, \\ T_\omega^{(n)} < \gamma_n : & \text{decide } \mathcal{H}_0. \end{cases} \quad (3)$$

### IV. EXPERIMENTAL ANALYSIS

In line with what is described in [7], we trained several MLPs, each of which can provide binary classifications considering as input the latent space learnt by the four VAEs, in a fashion very similar to what discussed in [2]. Our MLPs share the same structure. The first layer is a fully-connected layer with ReLU (Rectified Linear Unit) activation function. Subsequently, one more fully-connected layer with ReLU activation functions is added as this intermediate layer enables the model to learn complex patterns in the data and capture higher-level representations. Finally, a fully-connected layer with two units is added, serving as the output layer of the model, with linear activation. This layer's output will be used for binary classification, and the model will make predictions based on the logits produced by this layer. Each MLPs is trained using the Adam optimiser and the *binary cross entropy* loss function operating over logits.<sup>1</sup> The various MLPs operate directly on the latent space computed by the four VAEs — VAE-A1, VAE-A2, VAE-A3, and VAE-A4 — each of which receives the CSI data from its respective antenna: any statistics on the classification error is thus informative of the level of overlaps different classes have.

Figures 4 and 5 illustrate the results of the computation of the D3F statistic  $T_\omega^{(n)}$  when considering two major cases of interests: the case where a *normal* behaviour for the situation to be monitored is that the room should be *Empty*, *cf.*, Figure 4, and one where one should expect that in the room there should be someone *Running*, *cf.*, Figure 5. By inspection of Figure 4, it appears that the latent spaces of the VAEs are statistically very informative for providing classification interpretability. This means that it is easy to determine a decision threshold  $\gamma_n$  to select one of the two target hypotheses. A notable exception is the case of *Sit*, Figure 4a, where it appears that the two classes somehow overlap partially in the latent space, making it difficult for a classifier to distinguish them. This seems to reinforce a remark based on visual inspection of the graphical representation of the latent spaces mentioned in [2].

Another remark — based on a visual inspection of the graphical representation of latent spaces in Figure 5 and discussed in [2] — also receives support, namely that the classes *Run* and *Jump* share some similarities, and that *Run* and *Walk* are the least separable out of all the target classes. Thus, Figures 4 and 5 do not only provide the statistical counterpart to intuitions based upon graphical inspections in [2], but they also provide a first quantitative assessment of

<sup>1</sup>The code used for this analysis is available at <http://tiny.cc/D3FCSI23> (on 25th October 2023).

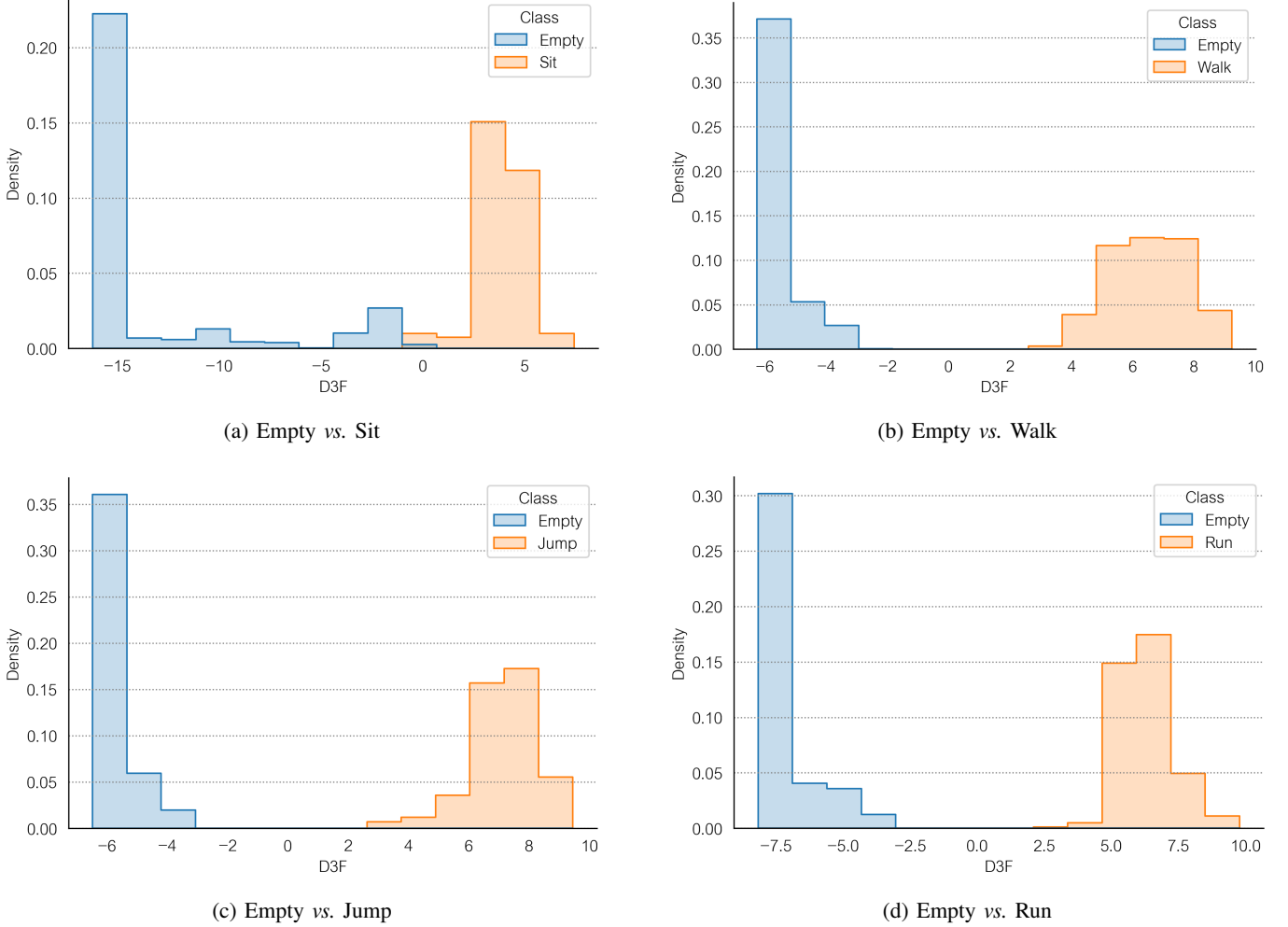


Fig. 4: Empirical density of the D3F's output with  $n = 4$ , computed when considering the four binary classification problems associated with the class *Empty Room*.

the level of confidence linked to the decision threshold  $\gamma_n$ , and the level of type I and II errors deemed acceptable, as they highlight the trade-off between statistical significance and sensitivity in decision-making. The interpretable results would thus enhance trust in decision informativeness facilitating human-AI teaming.

## V. RELATED WORK ON CSI SENSING

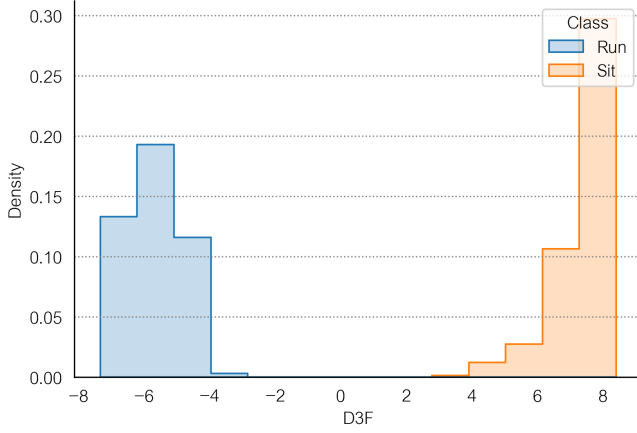
Device-free wireless sensing is a relatively new research field that has experienced significant growth over the last decade [14]. Recently, the research community focused on rendering CSI-based wireless sensing more reliable by devising novel environment-independent sensing techniques. Examples of such techniques include the preprocessing of the CSI data using statistical techniques [15] or deriving physical quantities such as the Micro-Doppler effect [10]. Indeed, one of the most critical limitations of CSI-based sensing is that the model learned in one environment cannot be directly transferred to another because the nodes' location and the furniture layout strongly affect the CSI. On the other hand, this work tries

for the first time to improve Wi-Fi sensing by providing human analysts with interpretable results that are eventually easier to generalise to new scenarios. While some works already started investigating the interpretability of multi-modal sensors for HAR [16], to our knowledge the efforts towards an interpretable framework for Wi-Fi sensing are still thin.

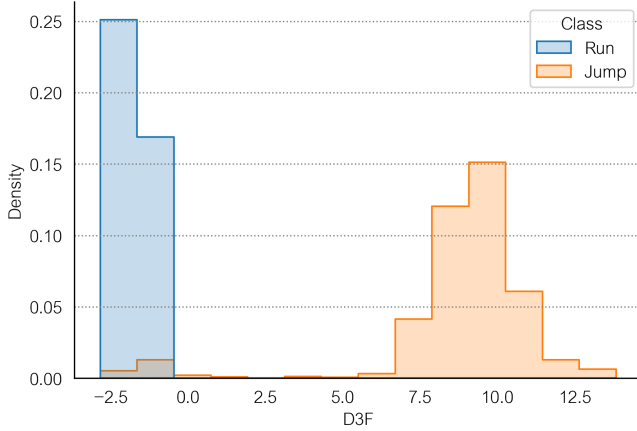
## VI. CONCLUSIONS

This paper provides a first analysis of the dataset introduced in [3] and analysed in [2] using the statistical analysis technique introduced in [6] and [7]. The results show that the compressed data obtained using VAEs are statistically very informative for providing classification interpretability. The paper demonstrates remarks suggested in [2] by visual inspection of some graphical representation of latent space distributions, and provides a statistically sound methodology for assessing the trustworthiness of the overall HAR system.

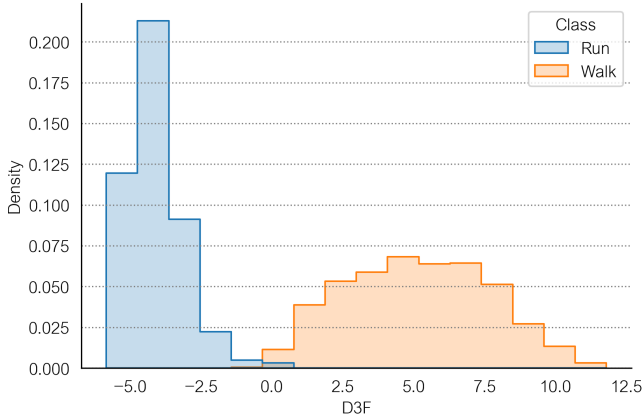
Future work will include an in-depth analysis of the theoretical properties that can be ensured by inspection of the VAEs used. In particular, in [6] and [7] the authors show



(a) Run vs. Sit



(b) Run vs. Jump



(c) Run vs. Walk

Fig. 5: Empirical density of the D3F's output with  $n = 4$ , computed when considering the remaining binary classification problems associated with the class *Run*. The comparison between *Run* and *Empty Room* is already reported in Fig. 4.

the relationship between the parameter  $n$  and the quality of the performance analysis, a connection we could exploit to optimise the process of training the VAEs.

We also aim to expand the training procedure illustrated in [2] in a neuro-symbolic fashion for situation understanding—*i.e.*, merging neural networks with symbolic reasoning to create a synergistic framework for tackling complex cognitive tasks—thus enabling the training of the VAEs to use the class signal to support human analysts. The analysis provided in this paper will be particularly useful for devising a loss function for minimising the error probabilities.

## REFERENCES

- [1] W. Li, R. J. Piechocki, K. Woodbridge, C. Tang, and K. Chetty, "Passive WiFi Radar for Human Sensing Using a Stand-alone Access Point," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 1986–1998, 2020.
- [2] M. Cominelli, F. Gringoli, L. M. Kaplan, M. B. Srivastava, and F. Cerutti, "Accurate Passive Radar via an Uncertainty-Aware Fusion of Wi-Fi Sensing Data," *Proceedings of 26th International Conference on Information Fusion*, June 2023.
- [3] M. Cominelli, F. Gringoli, and F. Restuccia, "Exposing the CSI: A Systematic Investigation of CSI-based Wi-Fi Sensing Capabilities and Limitations," in *2023 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2023, pp. 81–90.
- [4] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," in *ICLR2014*, 2014.
- [5] R. Tomsett, A. Preece, D. Braines, F. Cerutti, S. Chakraborty, M. Srivastava, G. Pearson, and L. Kaplan, "Rapid trust calibration through interpretable and uncertainty-aware AI," *Patterns*, vol. 1, no. 4, 2020.
- [6] P. Braca, L. M. Millefiori, A. Aubry, S. Marano, A. De Maio, and P. Willett, "Statistical hypothesis testing based on machine learning: Large deviations analysis," *IEEE Open Journal of Signal Processing*, vol. 3, pp. 464–495, 2022.
- [7] P. Braca, L. M. Millefiori, A. Aubry, A. De Maio, and P. Willett, "Large deviations for classification performance analysis of machine learning systems," *arXiv preprint arXiv:2301.07104*, 2023.
- [8] E. Khorov, A. Kiryanov, A. Lyakhov, and G. Bianchi, "A Tutorial on IEEE 802.11 ax High Efficiency WLANs," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 197–216, 2018.
- [9] F. Gringoli, M. Cominelli, A. Blanco, and J. Widmer, "AX-CSI: Enabling CSI Extraction on Commercial 802.11ax Wi-Fi Platforms," in *Proceedings of the 15th ACM Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*, 2022, p. 46–53.
- [10] F. Meneghello, D. Garlisi, N. Dal Fabbro, I. Tinnirello, and M. Rossi, "SHARP: Environment and Person Independent Activity Recognition with Commodity IEEE 802.11 Access Points," *IEEE Transactions on Mobile Computing*, pp. 1–16, 2022.
- [11] N. Bahadori, J. Ashdown, and F. Restuccia, "ReWiS: Reliable Wi-Fi Sensing Through Few-Shot Multi-Antenna Multi-Receiver CSI Learning," in *Proceedings of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2022, pp. 50–59.
- [12] E. L. Lehmann, J. P. Romano, and G. Casella, *Testing statistical hypotheses*. Springer, 2005.
- [13] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*. Springer Science & Business Media, 2009, vol. 38.
- [14] Y. Ma, G. Zhou, and S. Wang, "WiFi Sensing with Channel State Information: A Survey," *ACM Computing Surveys*, vol. 52, no. 3, pp. 1–36, 2019.
- [15] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas, W. Xu, and L. Su, "Towards Environment Independent Device Free Human Activity Recognition," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking (Mobicom)*, 2018, p. 289–304.
- [16] L. Yuan, J. Andrews, H. Mu, A. Vakil, R. Ewing, E. Blasch, and J. Li, "Interpretable passive multi-modal sensor fusion for human identification and activity recognition," *Sensors*, vol. 22, no. 15, 2022.