2022 44th Annual International Conference of
the IEEE Engineering in Medicine & Biology Society (EMBC)
Scottish Event Campus, Glasgow, UK, July 11-15, 2022

2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-2782-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/EMBC48229.2022.9870989

Similarity Maps for Ventricular Arrhythmia Classification

Qing Lin[1], Hak-Keung Lam[1], Michael J. Curtis[2] and Zoran Cvetkovic[1]

Abstract— **Ventricular arrhythmias are the primary arrhythmias that cause sudden cardiac death. In current clinical and preclinical research, the discovery of new therapies and their translation is hampered by the lack of consistency in diagnostic criteria for distinguishing between ventricular tachycardia (VT) and ventricular fibrillation (VF). This study develops a new set of features,** *similarity maps*, **for discrimination between VT and VF using deep neural network architectures. The similarity maps are designed to capture the similarity and the regularity within an ECG trace. Our experiments show that the similarity maps lead to a substantial improvement in distinguishing VT and VF.**

I. INTRODUCTION

Ventricular Fibrillation (VF) is one of the cardiovascular diseases with the highest mortality, manifesting as the sudden cardiac death if the patient is not adequately treated within a few minutes [1]. Defibrillation is currently considered the only definitive treatment for VF. However, due to the poor discrimination between VF and ventricular tachycardia (VT), all tachyarrhythmias may be treated with DC shock [2]. Inappropriate shocks can be harmful since the response to interventions varies according to the type of tachyarrhythmias [3]. Therefore, the correct detection and classification of VF and VT are essential for applying adequate intervention, developing new therapies and translational medicine. However, the discovery of new therapies and their translation is hampered by a lack of consistency in diagnostic criteria for distinguishing between VF and VT [4]. Aside from inconsistency over specific definitions, the ECG has intrinsic differences from person to person. Some transient and persistent forms of polymorphic VT may be confounded with VF, as VT is not always monomorphic [5]. In addition, complex arrhythmias are often segueing between different types [4]. This all makes it very challenging to distinguish between VT and VF. Therefore, exploring an automatic method is crucial to objectively distinguish VT and VF from ECG signals.

For detection and classification of ventricular arrhythmias using ECG signals, various classification approaches and feature sets have been developed. Comprehensive reviews of different low-dimensional features proposed for ventricular arrhythmia detection and classification have been presented in [6], [7], along with systematic evaluations of their ability to discriminated between VT and VF. Alwan et al. [5] improved the discrimination between VT and VF by combining most effective features from [6], [7] with high-dimensional spectral features, using temporal ensembles of SVM classifiers. The highest average sensitivity obtained by this study reached 74.7% in the three-class classification between VT, VF, and non-ventricular rhythms (NVR), *i.e.* all other rhythms that apart from VT and VF.

Recently, deep learning has been applied in many different fields, including medicine and bioengineering. This end-to-end deep learning approach has the ability to automatically extract high-level and informative features from the raw input data. Owing to this, many approaches have been elaborated for cardiac arrhythmia classification [8], [9], [10], [11], [12], including a deep convolutional structure with 16 residual blocks that achieved cardiologist-level accuracy in classifying 12 cardiac arrhythmias [9]. Hence, in this work we reassessed the accuracy of classification between VT, VF and NVR using the deep neural architecture proposed in [9], referred to as ResNet34. However, we observed no improvement in comparison with the SVM approach in [5], and similarly to the results in [5] a very high variability in the accuracy depending on the split of the data between train and test records. Inspection of the data revealed inconsistent labelling criteria across and within databases we considered, as well as departures from standard definitions of VT and VF [4]. Towards further assessing and improving the classification accuracy, the data were first relabelled according to Lambeth Conventions [4]. In the spirit of the Lambeth Conventions' definitions of VT and VF, which are based on regularity and similarity between consecutive segments of ECG signals, we then propose *similarity maps* as features for ventricular arrhythmia classification. Experimental results show that as a result of consistent labelling and novel features, the average sensitivity increases to 86.08%.

The remainder of the paper is organised as follows. Section II presents details of used methodologies. Section III discusses the data set setting and experimental evaluation method. Section IV summarises the experimental outcomes. In the end, conclusions were drawn in Section V.

II. METHODS

A. Similarity Maps

In the field of ventricular arrhythmia diagnosis, experts typically consider the regularity of abnormal heartbeats when analyzing patients' ECG records [4]. However, in nearly all CNN-based ventricular arrhythmia classification tasks, the prediction is dependent on the features extracted from an extended sequence of ECG signals in which the rhythm may segue between different type. Therefore, we propose a signal

[1]Qing Lin, Hak-Keung Lam and Zoran Cvetkovic are with Department of Engineering, King's College London, The strand, London, WC2B 4BG, UK. qing.lin@kcl.ac.uk, hak-keung.lam@kcl.ac.uk, zoran.cvetkovic@kcl.ac.uk

[2]Michael Curtis is with School of Cardiovascular Medicine & Sciences, King's College London. michael.curtis@kcl.ac.uk

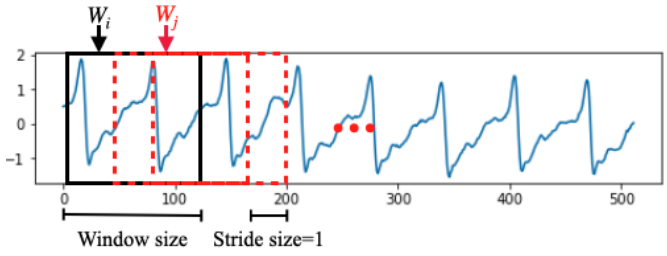

Fig. 1. The generation of similarity maps.

processing technique to capture repetition and similarities within ECG segments.

In order to discover pattern repetition, or the lack thereof, in ECG signals, all possible patterns in the ECG trace need to be collected by means of a sliding window. Given an ECG segment $x = \{x_1, x_2, \ldots, x_N\}$ consisting of N samples, which we will refer to as *observation length*, we first obtain all sub-sequences through the sliding window of length $l+1$ and shift by 1. The sub-sequence W_i obtained in this manner are given by: $W_i = \{x_i, x_{i+1}, \ldots, x_{i+l}\}$. For each pair of sub-sequences, we evaluate their similarity using two measures:

- Euclidean Distance: $I_{i,j} = \sqrt{\sum_{k=1}^{l} |W_{i,k} - W_{j,k}|^2}$
- Normalised Dot Product: $I_{i,j} = \dfrac{< W_i, W_j >}{\|W_i\| \, \|W_j\|}$

For each sub-sequence W_i, we define one channel of features using its similarity values with all other subsequences. The length of each feature channel is equal to L, where $L = N - l$ is the total number of subsequences. The i-th feature channel can be represented as: $I_i = \{I_{i,1}, I_{i,2}, \ldots, I_{i,L}\}$, where $I_{i,j}$ is the similarity value between subsequences W_i and W_j. Figure 1 shows a schematic diagram of this sliding window approach to generate the i-th feature channel.

The similarity map feature $\mathbf{I}$ is then the $L \times L$ matrix of all similarity scores $I_{i,j}$:

$$\mathbf{I} = \begin{bmatrix} I_{1,1} & I_{1,2} & \ldots & I_{1,L-1} & I_{1,L} \\ I_{2,1} & I_{2,2} & \ldots & I_{2,L-1} & I_{2,L} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ I_{L,1} & I_{L,2} & I_{L,3} & I_{L,L-1} & I_{L,L} \end{bmatrix}$$

Similarity feature maps were only used to distinguish VT and VF. This is because the similarity feature maps discard some morphological features of ECG signal, such as RR intervals, QT segments, QRS complexes, which is crucial for the identification of most non-ventricular arrhythmias.

B. Cyclically Shifted Similarity Maps

To obtain the average similarity value of ECG segment, we considered also the circular shift operation for each channel of the similarity feature map, which shifted the first entry to the last position while moving all other entries to the previous position. Given the i-th channel of the similarity feature map $\mathbf{I}$, we repeatedly operate circular shifts shift i times until $I_{i,i}$ reached the first position. We removed the first column of the similarity map after circular shift operation, which gave the following variant of the similarity map:

$$\mathbf{I}' = \begin{bmatrix} I_{1,2} & I_{1,3} & \ldots & I_{1,L-1} & I_{1,L} \\ I_{2,3} & I_{2,4} & \ldots & I_{2,L} & I_{2,1} \\ I_{3,4} & I_{3,5} & \ldots & I_{3,1} & I_{3,2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ I_{L,1} & I_{L,2} & \ldots & I_{L,L-2} & I_{L,L-1} \end{bmatrix}$$

Besides, we also formed a vector $\mathbf{I}'_{avg}$ by calculating the average value of each column of the cyclically shifted similarity map $\mathbf{I}'$.

C. Classification Using Deep Learning

We employed 1D ResNet34 [9] as the backbone classifier for feature extraction and classification. The network architecture consisted of 34 convolutional layers which were grouped into 16 residual blocks. There are 32×2^k filters with 16 filter widths in the convolution layers, where k starts from 0 and is increased by 1 for every 4 residual blocks.

D. Hierarchical Classification

Since the criteria for identification of ventricular tachyarrhythmias and discrimination between VT and VF are different, we introduced a two-stage hierarchical approach that breaks down this 3-class problem into two sub-problems.

For each given ECG segment, decisions are made simultaneously by classifiers C_1 and C_2 with different sets of features. C_1 makes three types of decisions, VF, VT, or NVR, but the decision will only be accepted when the sample is classified as NVR. Otherwise, the decision is taken by the classifier C_2, which is a binary classification model that distinguishes VT and VF. This procedure is described as:

$$C(X_{c1}, X_{c2}) = \begin{cases} C_1(X_{c1}), & if\ C_1(X_{c1}) = \text{NVR} \\ C_2(X_{c2}), & if\ C_1(X_{c1}) \neq \text{NVR} \end{cases} \quad (1)$$

where X_{c1} and X_{c2} are the input features for classifiers C_1 and C_2, respectively.

Classification experiments were performed with the hierarchical structure model, where C_1 and C_2 are both ResNet34. The input features X_{c1} of classifier C_1 were extracted using Parzen band-pass filters [13] with 64 trainable filters. For the classifier C_2, we considered three different feature maps as input features X_{c2}, in particular, similarity map $\mathbf{I}$, cyclically shifted similarity map $\mathbf{I}'$ and cyclically shifted similarity map with averaging $\mathbf{I}'_{avg}$. The feature map X_{c2} was further processed by a 1D convolutional layer with 16 1×4 filters before being passed to the classifier C_2.

III. EXPERIMENTAL SETUP

A. Data and Preprocessing

ECG signals were obtained from five publicly available arrhythmia databases: MIT-BIH Arrhythmia Database (MITDB) [14], the MIT-BIH Malignant Ventricular Arrhythmia Database (VFDB) [15], the European ST-T Database (EDB) [16], the Creighton University Ventricular Tachyarrhythmia Database (CUDB) [17] and the extended American Heart Association Database (AHADB). In this study,

only records containing VT or VF episodes are kept. Since the labeling of ventricular flutter is controversial and labeled in different ways between databases, any records with clearly marked ventricular flutter rhythms are excluded in this work. As a result, 91 records were used in our experiments.

Mislabeling is commonplace in public databases because of a lack of consistency in the use of diagnostic criteria for distinguishing between VF and VT. In our exploration of databases, we have encountered unequivocal examples of the following (using the Lambeth Conventions [4] definitions of arrhythmias as the criteria): VT labeled as VF; VF labeled as VT; electrical noise labeled as VF; asystole labeled as VF. In addition, we have noted that the mislabeling of traces is not systematic or consistent. This is most evident in traces where a rhythm starts out as VT, and then changes to VF, and then changes back to VT or asystole or normal sinus rhythm where the curators have in some instances labeled the different phases as different rhythms, and in other cases labeled the entire bloc of arrhythmia as a single rhythm (i.e., VT or VF). Consequently, we carefully relabeled the collected data according to Lambeth conventions [4]. After relabeling, only 78 of 91 records carried ventricular arrhythmia episodes. Table I shows the total duration of each category present across used records before and after the relabeling. The relabeled data set will be made publicly available.

TABLE I
THE AMOUNT OF DATA BEFORE AND AFTER RELABELING.

Class	Original(s) - 91 records	Relabeled(s) - 78 records
NVR	157366	122405
VT	5348	10021
VF	16229	3175
Total	178943	135601

All collected records were resampled to a 250Hz and processed by a high pass Butterworth filter with a cutoff of 0.5 Hz to remove the ECG baseline wander. Each record was normalized to zero mean and a standard deviation of 1.

ECG records were split into non-overlapping segments for training and testing. The length of ECG segment is referred to as observation length. To deal with the class imbalance problem, we randomly subsampled the largest class (NVR class) to the size of the second-largest class.

B. Training

Classification experiments were performed using 10 bootstraps resamples, with 80% of collected ECG records as training set and the rest held for testing. In order to have a fair comparison between methods, each experiment used the same hyperparameters including learning rate and batch size, etc. The trainable weights in ResNet34 were initialized as described in [18]. Adam optimizer with initialized learning rate $lr = 5 \times 10^{-4}$ was applied to update the weights.

The models were trained with a maximum of 50 epochs and mini-batches of size 32. We aimed to get batches that contained samples from all the classes. For one specific class, we first calculate the percentage of that class in the total number by $P_i = \frac{N_i}{N}$, where $i \in \{VT, VF, NVR\}$. Then, the number of samples of class i in one batch can be calculated by $M_i = P_i \times Batch\_size$. In other words, in each batch of an epoch, there are M_{VT} samples from VT set, M_{VF} samples from VF set and the remaining samples from NVR set. After processing all the samples of the smallest class, we move to the next epoch.

C. Evaluation Method

Due to the class imbalance, we report the sensitivity of each given category, which measures the proportion of correctly classified samples:

$$sn_s = \frac{TP_s}{TP_s + FN_s}, s \in S \tag{2}$$

where the TP_s is the number of examples in class s correctly assigned to class s, and the NF_s is the number of examples in class s incorrectly assigned to other classes.

Another metric to determine the performance of a classifier is the unweighted sensitivity, or average sensitivity, which can be calculated as:

$$SEN_{avg} = \frac{1}{|S|} \sum_{s \in S} sn_s \tag{3}$$

IV. EXPERIMENTAL RESULTS

Table II shows the results obtained using ResNet34 in the three-class case with raw ECG segment as the input features using the original data set and relabeled data set. The observation length is 512 samples (2s).

TABLE II
THE EFFECT OF DATA LABELING ON INDIVIDUAL AND AVERAGE CLASS SENSITIVITIES.

Method	Dataset	VT%	VF%	NVR%	SEN_{avg} %
Study [5]	non-relabeled	50.0	79.8	94.4	74.7
ResNet34	non-relabeled	37.50	81.00	95.33	71.28
ResNet34	relabeled	78.68	55.01	94.17	75.96

The model is biased toward classifying the ventricular arrhythmias as the class with a larger amount. Since relabeling leads to higher average class sensitivity, further experiments are done using the relabeled data. Table III shows classification results of the proposed hierarchical architecture with similarity maps based on the Euclidean distance and the normalised dot product. This classification experiment was performed using $N = 512$-sample (2s) observation length with $l = 128$-sample windows to form the similarity maps. As the input features of classifier C_1 are the same, the sensitivity of NVR is unique when observation length is consistent. It can be observed that classification using similarity maps based on the normalised inner products yields lower average class sensitivity, so similarity maps in further experiments were formed using the Euclidean distance.

Table IV shows the impact of the window length l used to generate the similarity maps. All results were obtained again with the observation length of $N = 512$ samples (2s). We note that the results are not too sensitive to the window length and the particular form of the similarity maps. The

TABLE III

SENSITIVITY VALUES OF THE HIERARCHICAL CLASSIFIER WITH THE SIMILARITY MAPS BASED ON THE EUCLIDEAN DISTANCE AND THE NORMALISED DOT PRODUCT.

Similarity Measure	Input Feature	VT (%)	VF (%)	NVR (%)	SEN_{avg} (%)
Euclidean Distance	Similarity Maps	79.62	72.70	94.84	82.38
	Cyclically Shifted Similarity Maps	84.38	61.72		80.31
	Cyclically Shifted SimilarityMaps with Averaging	76.15	77.65		**82.88**
Normalised Dot Product	Similarity Maps	81.35	65.94		80.71
	Cyclically Shifted Similarity Maps	83.69	61.24		79.93
	Cyclically Shifted Similarity Maps with Averaging	61.33	75.43		77.20

highest average class sensitivity was achieved using $l = 128$-sample windows and the similarity maps with the cyclic shift and averaging. Compared with results shown in Table II, the average sensitivity was increased from 75.96% to 82.88%. The improvement in average class sensitivity mainly comes from the classification between VT and VF.

TABLE IV

SENSITIVITY VALUES OF THE HIERARCHICAL APPROACH WITH THE SIMILARITY MAPS CALCULATED USING WINDOWS OF DIFFERENT LENGTHS.

Input Features	Window Size	VT (%)	VF (%)	NVR (%)	SEN_{avg} (%)
Similarity Maps	64 samples	79.35	70.70	94.84	81.63
	128 samples	79.62	72.70		82.39
	192 samples	80.93	72.30		82.69
	256 samples	78.72	71.97		81.85
Cyclically Shifted Similarity Maps	64 samples	90.18	52.01		79.01
	128 samples	84.39	61.73		80.32
	192 samples	87.15	63.16		81.72
	258 samples	81.31	64.54		80.23
Cyclically Shifted Similarity Maps with averaging	64 samples	66.49	79.65		80.32
	128 samples	76.15	77.65		**82.88**
	192 samples	66.67	80.28		80.60
	256 samples	75.88	73.99		81.57

TABLE V

SENSITIVITIES OF THE HIERARCHICAL APPROACH WITH AVERAGED CYCLICALLY SHIFTED SIMILARITY MAPS, CORRESPONDING TO DIFFERENT OBSERVATION LENGTHS.

Observation length	VT(%)	VF(%)	NVR(%)	SEN_{avg}(%)
512 samples (2s)	76.15	77.65	94.84	82.88
1280 samples (5s)	83.54	80.53	94.18	86.08

Finally we consider the impact of increasing the observation length from $N = 512$ samples (2s) to $N = 1280$ samples (5s). The results shown in Table IV, obtained with similarity maps generated using windows of $l = 128$ samples, with cyclic shift and averaging (the best-performing scenario in the previous experiment), demonstrate further improvement from 82.88% to 86.08% achieved by increasing the observation length.

V. CONCLUSIONS

In present study, a hierarchical deep CNN model was considered to discriminate between VT, VF and NVR using ECG signals. The data used in this study were collected from 5 public data repositories and relabeled according to the Lambeth Conventions (II) [4]. We proposed the similarity maps, novel features for ventricular arrhythmia classification, that aim to capture the existence of repetitive patterns in ECG signals. Our empirical results show that the similarity maps may offer substantially improved accuracy in distinguishing between VT and VF in an ECG trace.

REFERENCES

[1] M.E. Smith, T.W.and Cain. Sudden cardiac death: epidemiologic and financial worldwide perspective. *Journal of Interventional Cardiac Electrophysiology*, 17(3):199–203, 2006.

[2] J.B. van Rees et al. Inappropriate implantable cardioverter-defibrillator shocks: incidence, predictors, and impact on mortality. *Journal of the American College of Cardiology*, 57(5):556–562, 2011.

[3] F. Kette et al. What is ventricular tachycardia for an automated external defibrillator? *J Clin Exp Cardiol*, 5:285, 2014.

[4] M.J. Curtis et al. The lambeth conventions (ii): guidelines for the study of animal and human ventricular and supraventricular arrhythmias. *Pharmacology & therapeutics*, 139(2):213–248, 2013.

[5] Y. Alwan et al. Methods for improved discrimination between ventricular fibrillation and tachycardia. *IEEE Transactions on Biomedical Engineering*, 65(10):2143–2151, 2017.

[6] Q. Li et al. Ventricular fibrillation and tachycardia classification using a machine learning approach. *IEEE Trans. Biomed. Eng.*, 61(6):1607–1613, 2014.

[7] F. Alonso-Atienza et al. Detection of life-threatening arrhythmias using feature selection and support vector machines. *IEEE Trans. Biomed. Eng.*, 61(3):832–840, 2014.

[8] J. Li et al. Deep convolutional neural network based ecg classification system using information fusion and one-hot encoding techniques. *Mathematical Problems in Engineering*, 2018, 2018.

[9] A.Y. Hannun et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine*, 25(1):65, 2019.

[10] G.T. Taye et al. Application of a convolutional neural network for predicting the occurrence of ventricular tachyarrhythmia using heart rate variability features. *Scientific reports*, 10(1):1–7, 2020.

[11] A. Picon et al. Mixed convolutional and long short-term memory network for the detection of lethal ventricular arrhythmia. *PloS one*, 14(5):e0216756, 2019.

[12] R. Panda et al. Detection of shockable ventricular cardiac arrhythmias from ecg signals using ffrewt filter-bank and deep convolutional neural network. *Computers in Biology and Medicine*, 124:103939, 2020.

[13] D. Oglic et al. Learning waveform-based acoustic models using deep variational convolutional neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:2850–2863, 2021.

[14] G.B. Moody and R.G. Mark. The impact of the mit-bih arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3):45–50, 2001.

[15] S.D. Greenwald. *The development and analysis of a ventricular fibrillation detector*. PhD thesis, Massachusetts Institute of Technology, 1986.

[16] A. Taddei et al. The european st-t database: standard for evaluating systems for the analysis of st-t changes in ambulatory electrocardiography. *European heart journal*, 13(9):1164–1172, 1992.

[17] F.M. Nolle et al. Crei-gard, a new concept in computerized arrhythmia monitoring systems. *Computers in Cardiology*, 13:515–518, 1986.

[18] K. He et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1026–1034, 2015.