Improved Deep Learning Based ECG Classification through Automated Feature Selection and Weighted Loss Function

Timo De Waele*, Daniel Peralta*, Eli De Poorter*, Adnan Shahid*
*IDLab, Department of Information Technology, Ghent University - imec
Technologiepark–Zwijnaarde 126, B-9052 Ghent, Belgium
Email: firstname.lastname@ugent.be

***Abstract*—This paper proposes two approaches to significantly improve the detection rate of abnormal heartbeats in an Electrocardiogram (ECG) based deep learning heartbeat classifier. We introduce an automated feature selection procedure using the Kendall rank correlation coefficient to improve the performance of already existing classifier models. Further, we propose a methodology to cope with the class imbalance present in many ECG and other medical datasets by using a weighted loss function. The proposed methods demonstrate a significant improvement in the detection of Supraventricular Ectopic Beat (SVEB) and Ventricular Ectopic Beat (VEB) type heartbeats. Boasting an impressive 20% increase in terms of recall for the SVEB class when compared to state-of-the-art classifiers. This advancement could lead to more reliable and efficient tools for early arrhythmia detection, particularly beneficial in places where professional medical care is not easily accessible.**

***Index Terms*—Deep learning, Electrocardiogram, Arrhythmia detection, Heartbeat classification, Feature selection, Neural networks**

I. Introduction

Cardiovascular diseases (CVD) are the leading cause of death globally. In 2019, an estimated 17.9 million people died from CVD, which represented 32% of all global deaths [1]. A noteworthy fact is that over three-quarters of CVD deaths occur in low- and middle-income countries. This can largely be attributed to the fact that people in these countries have less access to effective healthcare services. As a result, CVD often gets detected late in the course of the disease. In these countries, a scalable, cost-effective way of performing an automated initial screening of heart conditions through ECG could prove a life-saving solution. One method of building these analysis tools could be by combining signal processing techniques and deep learning. Deep learning has already proven useful in early diagnosis of a variety of medical problems. Researchers have successfully applied neural networks to detect and classify breast cancer based on Magnetic Resonance Imagery (MRI) images [2]. Applying deep learning to gene expression data has shown to be useful for the grading and prediction of the survivability of brain tumors [3]. By combining manually extracted medical features such as glucose level, Body Mass Index (BMI), and insulin levels with deep learning models researchers have also managed to automate the detection of diabetes [4]. Using a deep Convolutional Neural Network (CNN) on the data extracted by an Electroencephalogram (EEG) the early diagnosis of Parkinson's disease is now also within the realm of possibilities [5]. Detecting abnormal heartbeat types, which could be the indication of a potential arrhythmia, is also feasible when applying neural networks on the signals coming from an ECG [6].

Due to the significant promise for life-saving diagnostic advancements deep learning has been widely applied for automated ECG classification. Driven in part by the accessibility of numerous high-quality open source datasets like MIT-BIH and PTB-XL [7], [8]. A critical issue in ECG classification research is the use of global accuracy as the sole performance metric. This approach fails to adequately represent the true performance for ECG classifiers, given the inherent class imbalance in datasets, where normal heartbeats significantly outnumber abnormal ones. Such an imbalance can mask the model's performance in detecting rarer, yet clinically crucial, abnormal heartbeats. Another prevalent issue is the common practice in research papers of shuffling and splitting data into training and validation sets without taking into consideration the specific patients. This method often results in the inclusion of ECG records from the same patients in both training and validation datasets. Given the highly individual nature of ECG signal morphology, this approach can lead to misleading results, as it does not accurately assess the model's ability to generalize to new, unseen patient data [9]. Such practices in dataset handling and performance evaluation can significantly limit the clinical applicability and reliability of these deep learning models.

Most successful studies made use of features that were extracted from the ECG signal. By combining coefficients from the Continuous Wavelet Transformation (CWT) with features based on the distance between the subsequent peaks in the signal, called the R-R interval, researchers have managed to obtain a global classification accuracy of 86% [10]. Researchers further improved on this by extracting more frequency domain information and simultaneously transforming the ECG signal from a 1D time-series to a 2D image through the use of the Short-Time Fourier Transform (STFT). This allowed them to apply transfer learning on a ResNet architecture that was pretrained for image classification. Using this approach they managed to achieve a global accuracy of

90% [11]. Other researchers have transformed the ECG signal to a Vectorcardiogram (VCG) and combined this with a feature selection strategy to further increase the global accuracy to 92% [12]. Finally, Wang *et al.* used another type of 1D to 2D and frequency domain transformation based on the CWT, they combined this with scalar features based on R-R intervals and applied a CNN on this data to obtain an overall accuracy of 97.5% [13]. This approach is to the best of our knowledge currently the best-performing algorithm on the MIT-BIH database. Due to the high class imbalance, with normal heartbeats being significantly more prevalent than abnormal ones, one should not focus too much on global accuracy. Since the main purpose of automated ECG analysis is to detect arrhythmias, the recall of the abnormal beats should be the most important metric used to compare results. This number indicates the percentage of abnormal beats correctly identified as being normal. In terms of arrhythmia, 2 types of heartbeats are of particular importance: the VEB and the SVEB. Table I compares the recall performance of the previously mentioned works with the algorithm that we propose in this paper.

TABLE I
RECALL VALUES FOR THE SVEB AND VEB HEARTBEAT TYPES.

	SVEB	**VEB**
Can Ye *et al.* [10]	60.8%	81.5%
Cao *et al.* [11]	9.0%	88.4%
Garcia *et al.* [12]	53.0%	87.3%
Wang *et al.* [13]	75.2%	95.7%
Our work	94.2%	96.8%

This paper builds further on the work done by Wang *et al.*. To enable ease of comparison we used the same training and validation data as well as applied the same preprocessing methods as them. While they already propose a good performing algorithm based on the combination of the CWT and R-R interval features. The performance on beats of the SVEB type still contained room for improvement, with a recall of just 75%. To further enhance the overall performance of the model, and specifically the performance on the SVEB type of beats, we propose two enhancements. Namely, adding more features through an automated feature selection procedure, based on the Kendall rank correlation coefficient, and using a weighted loss function instead of the standard cross entropy loss [14]. This strategy of automated feature calculation and selection has already proven to be successful in other time-series classification problems [15]. This paper clearly shows the benefit of neural networks for the automated analysis of ECG signals. This work could lay the foundation for a wearable tool designed to perform continuous heart monitoring. Giving people who do not have easy access to expert cardiology care the potential to gain direct insights into their health status.

The main contributions of this paper are as follows. First, we propose an automated feature selection procedure for time-series, through the use of the Kendall rank correlation coefficient. Second, we propose a deep learning architecture that uses CWT, RR features, and the new time series features calculated through the Kendall rank correlation coefficient. Third, we further improved the classification performance of the proposed architecture through the use of a weighted loss function, significantly increasing the recall of the abnormal heartbeats. Fourth, we provide an extensive reporting of the results obtained to gain insights into the behavior and trade-offs of the proposed methodology. Finally, we will make the model and the code used for generating the results in this paper publicly accessible (https://github.com/timodw/ecg_classification_ijcnn).

In the following section we provide an overview of the proposed methodology. Section III extensively reports the obtained results. In Section IV we provide further insight into the obtained results. Finally, the paper is concluded in Section V.

II. METHODOLOGY

We first describe a pipeline for calculating time series features from the ECG signal and then present the deep learning architecture used for the ECG classification.

A. Feature calculation and selection

We propose a pipeline for calculating time series features from the ECG signal and automatically selecting the best-suited features based on their Kendall rank correlation coefficient, as shown in Figure 2. Initially, we segmented the 30-minute-long ECG signal for each patient into windows containing the data from a single heartbeat. Subsequently, we utilized the `tsfel` Python library to calculate a wide variety of time-series features on these windows [16]. Table II gives an overview of the 4 categories of features extracted by this library as well as some examples for each category. The full list of features can be found in the documentation of the `tsfel` library (https://tsfel.readthedocs.io/en/latest/descriptions/feature_list.html). These features were then sorted according to their feature significance value obtained by computing the Kendall Rank Correlation Coefficient τ and is calculated as follows:

$$\tau = \frac{n_c - n_d}{\binom{n}{2}}.$$

where n_c denotes the count of concordant pairs, n_d denotes the count of discordant pairs, and n is the total number of pairs. Concordant pairs (α_i, β_i) and (α_j, β_j) satisfy either $\alpha_i > \alpha_j \wedge \beta_i > \beta_j$ or $\alpha_i < \alpha_j \wedge \beta_i < \beta_j$; otherwise they are discordant. This is illustrated in Figure 1. In this paper, class labels are numerically encoded (0 for normal, 1 for SVEB, and 2 for VEB), with Y representing these numerical labels. X_k denotes the feature values for the k-th feature of the i-th

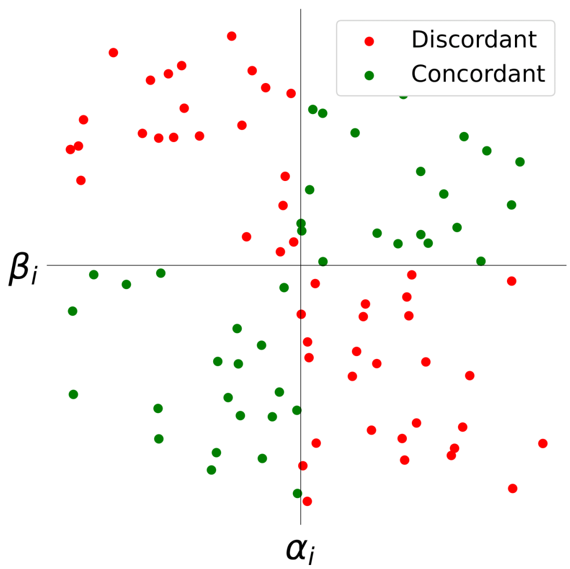

Fig. 1. Concordant and discordant pairs with respect to the pair (α_i, β_i).

window. Concordant and discordant counts for each feature are computed as follows:

$$\begin{aligned} Concordant_{i,k} &= ((X_{i,k} < X_{j,k}) \wedge (Y_i < Y_j)) \\ &\vee ((X_{i,k} > X_{j,k}) \wedge (Y_i > Y_j)) \\ Discordant_{i,k} &= ((X_{i,k} < X_{j,k}) \wedge (Y_i > Y_j)) \\ &\vee ((X_{i,k} > X_{j,k}) \wedge (Y_i < Y_j)) \end{aligned}$$

This calculation ultimately provides us with 2 values. The first is the Kendall τ value, indicating the correlation of the feature to the classification labels. A value of 0 indicates no correlation, 1 indicates full correlation, and -1 indicates full negative correlation. The second is the statistical p-value indicating the statistical significance of this τ value. To select the most relevant features, we initially ranked all features by their p-value and then used the absolute value of τ as a tie-breaker. In the end, we only used the top 5 features as obtained by this procedure in our model, as preliminary results showed no improvement in performance when more features were added. Table III lists these 5 features and their Kendall τ value.

TABLE II
CATEGORIES OF FEATURES EXTRACTED BY THE TSFEL PYTHON LIBRARY.

Statistical	Mean, median, standard deviation, variance,...
Temporal	Autocorrelation, energy, zero-crossing rate,...
Spectral	Spectral entropy, spectral centroid, spectral roll-off,...
Wavelet	Wavelet energy, wavelet entropy,...

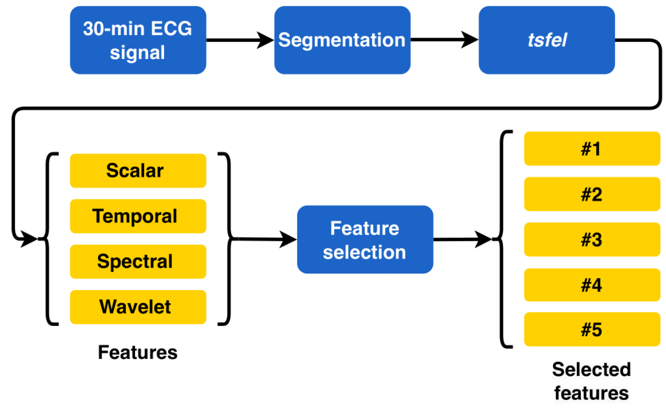

Fig. 2. Graphical overview of the feature selection pipeline. Features are calculated on the ECG windows, the top 5 features are then selected according to their Kendall rank correlation coefficient.

TABLE III
TOP 5 FEATURES AS OBTAINED BY THEIR KENDALL RANK CORRELATION COEFFICIENT

Feature	τ
Kurtosis	-0.2969
Area under the curve	0.2454
Average power	0.2043
Absolute energy	0.2043
Autocorrelation	0.2043

B. Proposed deep learning architecture

Figure 3 shows the proposed deep learning architecture that uses the 5 features calculated in the previous step, along with R-R features and the CWT. We concatenated the 5 features with the 4 R-R interval features: next R-R distance, previous R-R distance, the average of the previous 10 R-R distances, and the ratio between the next and the previous R-R distances. These 9 features were passed through a Deep Neural Network (DNN) with 2 hidden layers of 128 neurons each to obtain 64 output features. These features are then concatenated with the 64 features coming from the CNN head. The CNN architecture is the same as proposed by Wang *et al.*.

This CNN takes the CWT transformed ECG signal as an input and calculates 64 features. The CWT is a type of time-frequency domain transformation. It uses the same base idea as the STFT but makes it more adjustable through scale and translation parameters. Given a signal $x(t)$, the CWT is defined as

$$C_a(b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \cdot \phi\left(\frac{t-b}{a}\right) dt.$$

where a and b represent scale and translation parameters, respectively and $\phi(t)$ is the wavelet function used. The wavelet function $\phi(t)$ used in this paper is represented as follows.

$$\phi(t) = \frac{2}{\sqrt{3\sqrt[4]{\pi}}} \exp\left(-\frac{t^2}{2}\right)\left(1-t^2\right)$$

The scale parameter can also easily be transformed to its corresponding frequency by

$$f = \frac{f_c \cdot f_s}{a}$$

here f_c indicates the center frequency of the used wavelet and f_s is the sampling frequency of the original signal. By iterating over the different time steps while varying the scale parameter, and thus the frequency, one can use this transformation to obtain a 2D scalogram in the time-frequency domain of the original signal. This paper uses the same wavelet as Wang *et al.*, it closely matches the shape of ECG waveform and is already widely used in ECG signal analysis.

The 128 features in total from both heads are then passed through a final DNN containing 3 hidden layers with 256, 64, and 32 neurons respectively. The final output of the classifier model is 3. Table IV lists all layers with their corresponding parameters of the model used. Instead of using the ReLU activation function, we opted to use the Leaky ReLU activation function with a negative slope of 0.2. This was done to address the dying ReLU problem [17] where neurons become inactive and only output zero, thus blocking gradients from flowing back.

Finally, we changed the loss function from a simple Cross Entropy Loss to a Weighted Cross Entropy Loss function to combat the class imbalance that is present in the MIT-BIH database, and most ECG datasets. Over 80% of the heartbeats in the dataset are the normal type, and only 6.6% and 2.5% are of the VEB and SVEB type respectively. So samples that represent a minority class should carry more weight for optimizing the model's weights in comparison to samples from the majority class. In the Weighted Cross Entropy loss function, each term of the regular Cross Entropy loss is weighted according to the occurrence frequency of the true class. This ensures that classes that occur less in the training data carry more weight in the loss function, thus making the model focus more on predicting these samples correctly. The mathematical formulation of this loss function is as follows:

$$loss = -\sum_{c=1}^{C} w_c \cdot y_c \cdot log\frac{exp(x_c)}{\sum_{i=1}^{C} exp(x_i)}.$$

where C is the number of classes in the training dataset, w_c is the weight associated to class c, y_c is either 0 if c is different from the true class of the sample and 1 if c is equal to the true class of the sample, x_i is the i-th value of the output neurons of the classifier. The value of the class weights is calculated as follows:

$$w_c = \frac{N}{\sum_{i=1}^{N} 1 \cdot \{y_i = c\}}.$$

where N is the total number of samples in the dataset. y_i is the class of the i-th sample in the dataset. For this paper, the weights were calculated using the occurrences of each class in the training dataset. This resulted in weights of 0.37, 18.00, 4.44 for the normal type, SVEB, and VEB respectively.

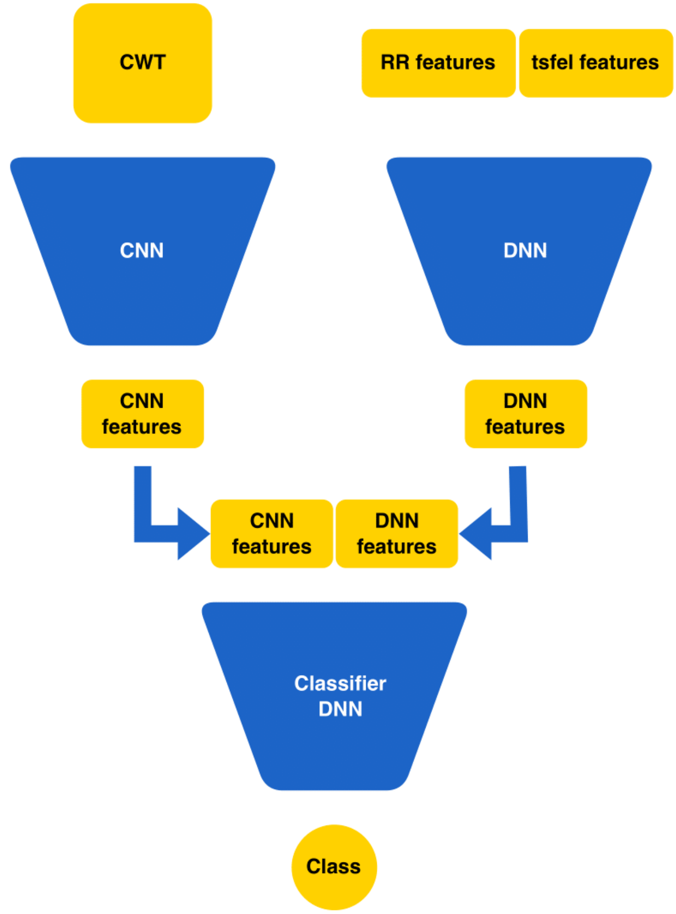

Fig. 3. Overview of the architecture used in this paper. The outputs of both heads are concatenated and fed through a final classifier neural network to obtain the predicted type of heartbeat.

III. Results

We use the results of Wang *et al.* as a baseline for comparing against our results. To make a fair comparison we split the data into the same training and validation dataset, where the split was made on a per-patient basis, so no data from patients that were present in the training dataset was present in the validation data. This was done to ensure that the results we obtained were reflective of the generalization capabilities of the proposed approach toward data from new patients. Figure 4 plots the results from the reference paper as a confusion matrix, we compiled this confusion matrix based on the confusion matrix given in the original paper. The normal class is indicated by the letter N, SVEB is represented by S and VEB is shown as a V. We see that both the N and V classes show good classification accuracy, the S class however shows a different result. Only 75% of the S class instances are correctly classified, 18% is misinterpreted as being a normal heartbeat, and just under 7% is interpreted as being a V-type heartbeat. The main metric of importance was the recall value of the abnormal beats, as the correct detection of abnormal heartbeats is significantly more important than the correct detection of

TABLE IV
MODEL ARCHITECTURE

	Layer Type	**Parameters**
Beat Head	2D Convolutional	Input channels: 1, output channels: 16, kernel size: 7
	Batch normalization	–
	LeakyReLU	Slope: 0.2
	Max pooling	Size: 5
	2D Convolutional	Input channels: 16, output channels: 32, kernel size: 3
	Batch normalization	–
	LeakyReLU	Slope: 0.2
	Max pooling	Size: 3
	2D Convolutional	Input channels: 32, out channels: 64, kernel size: 3
	Batch normalization	–
	LeakyReLU	Slope: 0.2
	Adaptive max pooling	Output size: 1x1
	Flatten	–
Feature Head	Dense	Input size: 9, output size: 128
	LeakyReLU	Slope: 0.2
	Dense	Input size: 128, output size: 128
	LeakyReLU	Slope: 0.2
	Linear	Input size: 128, output size: 64
	LeakyReLU	Slope: 0.2
Classifier	Dense	Input size: 128, output size: 256
	LeakyReLU	Slope: 0.2
	Dense	Input size: 256, output size: 64
	LeakyReLU	Slope: 0.2
	Dense	Input size: 64, output size: 32
	LeakyReLU	Slope: 0.2
	Dense	Input size: 32, output size: 3

TABLE V
OVERVIEW OF THE HYPERPARAMETERS USED DURING TRAINING.

Learning rate	10^{-4}
Epochs	200
Batch Size	4096
L2 regularization	None
LR scheduling	None
Checkpoint	Best validation loss

normal heartbeats. The recall is defined as follows:

$$Recall = \frac{TP}{TP + FN}$$

With TP being the instances that got classified correctly and FN being the incorrectly classified instances.

Figure 5 shows the confusion matrix obtained by the first configuration proposed in this paper. Adding the top 5 `tsfel` features as ranked by their Kendall rank correlation coefficient. This shows a 10% increase in terms of recall for the S-type heartbeats, with now around 12% of the instances being

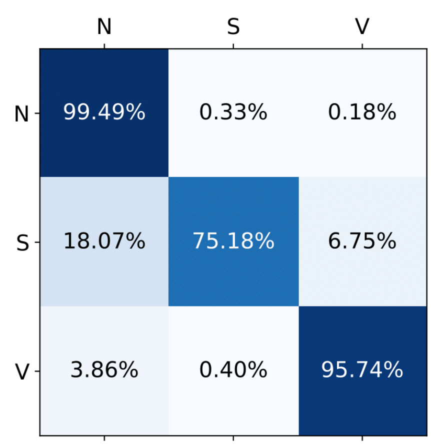

Fig. 4. Confusion matrix visualizing the classification results of Wang *et al.* [13].

misinterpreted for a normal heartbeat and just under 3% that is classified as a V-type heartbeat. One thing of notice for this configuration is the lower accuracy for the normal heartbeats, and thus more normal beats are being interpreted as abnormal beats.

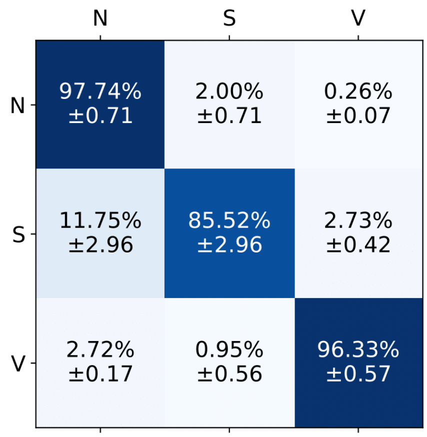

Fig. 5. Confusion matrix visualizing the classification results of our approach combined with 5 time-series features.

Finally, we trained the previous model with a weighted cross entropy loss function. The results for this configuration are given in Figure 6. This approach further enhances the classification accuracy of the S type by another 10%. However, almost 7% of the normal heartbeats now get misinterpreted as an S-type heartbeat. Leading to a lower precision and f1-score.

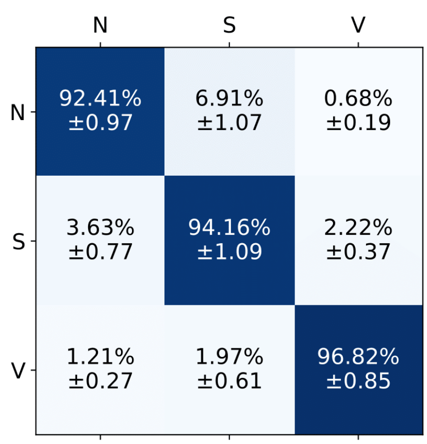

Fig. 6. Confusion matrix visualizing the classification results of our approach combined with 5 time-series features and trained using weighted cross entropy.

Table VI shows the extensive numerical results for our work as well as other published research. Here we observe that both of our proposed improvements show a significant increase over the baseline in terms of the balanced accuracy, which is the mean of the recall over all classes. If we look purely at the recall of the abnormal type of heartbeats, S and V, we see that adding extra features, combined with a weighted loss function, shows the best results. This configuration also has the highest precision on the normal type of heartbeats.

IV. Discussion

By just adding 5 extra features in addition to the 4 R-R interval features we managed to increase the recall on the SVEB type heartbeats by 10%. This is in contrast to many other publications, which struggle to get the recall on these types of beats above 70%. The VEB recall also increased slightly. However, our recall on the normal heartbeats is reduced by around 2%, with now 2% of normal heartbeats being classified as being of the SVEB type. Indicating that the model has become more sensitive to the SVEB class, preferring this class over the normal beats to obtain a higher recall on the abnormal beats, at the cost of a lower recall on the normal beats. This is preferable in the case of arrhythmia detection as it is preferable to have a higher recall on potentially life-threatening arrhythmias [18].

By adding class weights to the Cross Entropy loss function we managed to make the model focus more on the harder samples when optimizing the weights. Due to the low occurrence rate of the SVEB type beat in the training data, these types of beats carried the most weight when the loss values were computed. This shows in the results from this configuration, with now just over 94% of the SVEB beats being predicted correctly, almost another 10% increase over the first configuration and a 20% increase when comparing it against the baseline model. However, this performance increase on the SVEB type of beats came at a cost. Now just under 7% of the normal heartbeats get classified as being SVEB. When applying this in a clinical setting, this would mean 7% of all normal beats would incorrectly be attributed to be an indication of cardiac arrhythmia. This number of false positives is not ideal, this indicates that around 1 in 14 cases where the heartbeat is normal it would get classified as being abnormal and thus triggering a ”false alarm”. A high number of false positives or so-called ”false alarms” could potentially result in alarm fatigue when applied in a clinical setting [19]. In the case of arrhythmia detection, a potentially missed arrhythmia event is significantly worse than triggering a false alarm. To this end, obtaining a high recall of abnormal heartbeats should be the highest priority when designing beat classification algorithms.

V. Conclusion and Future Work

The goal of this study was to improve the performance of automated ECG analysis by improving the accuracy of an ECG based heartbeat classifier. We started by getting the baseline results from a model developed by Wang *et al.* on the popular MIT-BIH heartbeat database. This was, to the best of our knowledge, the best-performing classifier on the MIT-BIH database. However, this approach, as many other published research, still struggles with the SVEB type of heartbeats,

TABLE VI
DETAILED COMPARISON OF THE CLASSIFICATION RESULTS OF RELATED PAPERS AND OUR 2 PROPOSED IMPROVEMENTS.

		N			**S**			**V**		
	Balanced accuracy	**Precision**	**Recall**	**f1-score**	**Precision**	**Recall**	**f1-score**	**Precision**	**Recall**	**f1-score**
Can Ye *et al.* [10]	72.8%	98.0%	94.6%	96.3%	52.3%	60.8%	56.2%	81.5%	63.1%	71.1%
Cao *et al.* [11]	64.2%	95.3%	95.1%	95.2%	13.0%	9.0%	10.6%	68.2%	88.4%	77.0%
Garcia *et al.* [12]	81.1%	98.0%	94.0%	96.0%	53.0%	62.0%	57.1%	59.4%	87.3%	70.7%
Wang *et al.* [13]	90.1%	81.9%	**99.5%**	89.9%	**99.0%**	75.2%	**85.5%**	93.3%	95.7%	94.5%
5 features	93.2% ± 1.0	99.3% ± 0.1	97.7% ± 0.7	**98.5% ± 0.4**	64.2% ± 7.3	85.5% ± 3.0	73.0% ± 4.8	**94.9% ± 0.8**	96.3% ± 0.6	**95.6% ± 0.7**
5 features and weighted loss	**94.5% ± 0.4**	**99.7% ± 0.1**	92.4% ± 1.0	95.9% ± 0.5	36.0% ± 3.2	**94.2% ± 1.1**	52.0% ± 3.2	90.2% ± 2.4	**96.8% ± 0.9**	93.4% ± 1.6

only achieving a 75% recall on this type of beat. By adding automatically extracted features and an extra feature head to the deep learning we managed to obtain an SVEB recall of 85%. By changing the loss function used during training to a weighted Cross Entropy loss function we further managed to increase the recall performance on the SVEB beats to 94%. This increase in recall on the SVEB class came at the cost of the recall on normal beats, going down from 99% to 92%. In a clinical setting, when this algorithm would be used for detecting arrhythmias, the number of missed arrhythmias is of more importance than the number of people without arrhythmia who get falsely flagged as having an arrhythmia. As this paper achieved an almost 20% increase in recall on the SVEB type and a 1% increase in the recall of the VEB type of heartbeat, we significantly outperform other published research that was evaluated on the MIT-BIH dataset. This deep learning classifier enables real-world applications by providing remote arrhythmia detection capabilities to patients who lack direct access to cardiologists, thereby democratizing access to essential diagnostic insights into cardiac health.

Future work should focus on further increasing the recall of the abnormal type of heartbeats while keeping the number of normal beats that get misinterpreted as being abnormal within an acceptable range. This can be accomplished in a variety of manners. The architecture of the model can be extended or changed from a CNN based model to a Recurrent Neural Network (RNN) based one, or even a Transformer based model. More advanced features could also be added, by not only using the CWT as an input but other different transformations as well, such as the STFT or the Gramian Angular Field (GAF). As the R-R interval features already incorporate some more long-term information next to just the local heartbeat, extra features extracted from the surrounding beats could also be integrated into the model. Next to architectural improvements efforts should also be made to combine the MIT-BIH dataset with other publicly available datasets such as INCART and PTB-XL. These datasets could extend the training data available to the model but could also be used to evaluate the generalizability of the developed models.

ACKNOWLEDGEMENTS

This work was supported by the Fund for Scientific Research Flanders, Belgium, FWO-Vlaanderen, FWO-SB, under Grant 1S33924N.

This project was supported by the Erasmus+ program of the European Commission through the ‘Capacity Building for Digital Health Monitoring and Care Systems in Asia (DigiHealth-Asia)’ under grant agreement number 619193-EPP-1-2020-BE-EPPKA2-CBHE-JP.

REFERENCES

[1] World Health Organization, “Cardiovascular diseases (cvds).” World Health Organization, 2021. Accessed: 2024-01-15.

[2] B. S. Abunasser, M. R. J. Al-Hiealy, I. S. Zaqout, and S. S. Abu-Naser, “Convolution neural network for breast cancer detection and classification using deep learning,” *Asian Pacific journal of cancer prevention: APJCP*, vol. 24, no. 2, p. 531, 2023.

[3] L. P. Petalidis, A. Oulas, M. Backlund, M. T. Wayland, L. Liu, K. Plant, L. Happerfield, T. C. Freeman, P. Poirazi, and V. P. Collins, “Improved grading and survival prediction of human astrocytic brain tumors by artificial neural network analysis of gene expression microarray data,” *Molecular cancer therapeutics*, vol. 7, no. 5, pp. 1013–1024, 2008.

[4] M. Rahman, D. Islam, R. J. Mukti, and I. Saha, “A deep learning approach based on convolutional lstm for detecting diabetes,” *Computational biology and chemistry*, vol. 88, p. 107329, 2020.

[5] S. L. Oh, Y. Hagiwara, U. Raghavendra, R. Yuvaraj, N. Arunkumar, M. Murugappan, and U. R. Acharya, “A deep learning approach for parkinson’s disease diagnosis from eeg signals,” *Neural Computing and Applications*, vol. 32, pp. 10927–10933, 2020.

[6] M. D. Mazhar Qureshi, D. Peralta, E. De Poorter, R. Mumtaz, A. Shahid, I. Moerman, and T. De Waele, “Multiclass heartbeat classification using ecg signals and convolutional neural networks,” in *ICoDT2, the 2nd International Conference on Digital Futures and Transformative Technologies*, IEEE, 2022.

[7] G. B. Moody and R. G. Mark, “The impact of the mit-bih arrhythmia database,” *IEEE engineering in medicine and biology magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[8] P. Wagner, N. Strodthoff, R.-D. Bousseljot, D. Kreiseler, F. I. Lunze, W. Samek, and T. Schaeffter, "Ptb-xl, a large publicly available electrocardiography dataset," *Scientific data*, vol. 7, no. 1, p. 154, 2020.

[9] E. J. d. S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "Ecg-based heartbeat classification for arrhythmia detection: A survey," *Computer methods and programs in biomedicine*, vol. 127, pp. 144–164, 2016.

[10] C. Ye, B. V. Kumar, and M. T. Coimbra, "Heartbeat classification using morphological and dynamic features of ecg signals," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 10, pp. 2930–2941, 2012.

[11] M. Cao, T. Zhao, Y. Li, W. Zhang, P. Benharash, and R. Ramezani, "Ecg heartbeat classification using deep transfer learning with convolutional neural network and stft technique," in *Journal of Physics: Conference Series*, vol. 2547, p. 012031, IOP Publishing, 2023.

[12] G. Garcia, G. Moreira, D. Menotti, and E. Luz, "Inter-patient ecg heartbeat classification with temporal vcg optimized by pso," *Scientific reports*, vol. 7, no. 1, p. 10543, 2017.

[13] T. Wang, C. Lu, Y. Sun, M. Yang, C. Liu, and C. Ou, "Automatic ecg classification using continuous wavelet transform and convolutional neural network," *Entropy*, vol. 23, no. 1, p. 119, 2021.

[14] H. Abdi, "The kendall rank correlation coefficient," *Encyclopedia of Measurement and Statistics. Sage, Thousand Oaks, CA*, pp. 508–510, 2007.

[15] T. De Waele, A. Shahid, D. Peralta, A. Eerdekens, M. Deruyck, F. A. Tuyttens, and E. De Poorter, "Time-series based feature selection and clustering for equine activity recognition using accelerometers," *IEEE Sensors Journal*, 2023.

[16] M. Barandas, D. Folgado, L. Fernandes, S. Santos, M. Abreu, P. Bota, H. Liu, T. Schultz, and H. Gamboa, "Tsfel: Time series feature extraction library," *SoftwareX*, vol. 11, p. 100456, 2020.

[17] L. Lu, Y. Shin, Y. Su, and G. E. Karniadakis, "Dying relu and initialization: Theory and numerical examples," *arXiv preprint arXiv:1903.06733*, 2019.

[18] T. Le Nguyen and L. Thomas, "Supraventricular ectopic activity: When excessive it is not all benign!," *Journal of Atrial Fibrillation*, vol. 3, no. 2, 2010.

[19] M. Cvach, "Monitor alarm fatigue: an integrative review," *Biomedical instrumentation & technology*, vol. 46, no. 4, pp. 268–277, 2012.