1 Introduction

1.1 Outline

In this report, the deep convolutional neural network model for electrocardiogram (ECG) heartbeat classification proposed in [18] will be presented. The models results on the MIT-BIH arrhythmia [15] and St. Petersburg 12-lead arrhythmia (INCART) [8] data-sets will then be discussed and reproduced. Further, the validity of the model will be extended by evaluating its performance on a third data-set, the PTB Diagnostic database [3]. Finally, issues surrounding data-set pre-processing will be discussed, and improvements will be suggested.

1.2 Background

A cardiac arrhythmia constitutes any heartbeat or heart-rate that is abnormal (fast, slow, early, irregular). While many are harmless - sinus bradycardia (low resting heart rate) is even common in endurance athletes [7] - many others pose a serious health risk.

Heartbeats can be non-invasively recorded using an electrocardiogram (ECG) which records the hearts electrical activity using 12 electrodes placed at different superficial positions on the body. The standard ECG heartbeat morphology can be described by multiple intervals (PR, QT, RR), segments (PR, ST), waves (P,T,U) and one complex (QRS), as shown in Figure 1. Each of these measures encodes different information about the heart and abnormalities can be reflective of arrhythmia's. For example, the PR interval reflects the time required for the heart's electrical impulse to pass from where it originates to the ventricles, and a prolonged PR interval is indicative of an arrhythmia called first-degree atrioventricular block [6]. Of particular importance to this work, these measures can also used to segment ECG recordings into individual heartbeats (see below).

Given the danger associated with certain arrhythmia's and the ease of obtaining ECG recordings, developing accurate classification algorithms for ECG heartbeats is an important task and an active area of research at the intersection of machine learning and medicine.

1.3 Related Work and Motivation

Any ECG classification algorithm involves four main steps; (i) de-noising, (ii) segmentation of beats, (iii) feature selection, and (iv) classification. Owing to their ability to perform feature selection automatically in synergy with classification, most recently proposed ECG heartbeat classification algorithms, including the proposed method, are based on convolutional neural networks (CNNs) [1, 2, 9, 16, 5]. CNNs are composed of an input, output and various hidden layers, making them a class of deep neural networks. Each hidden layer of a CNN performs a specific task on the output of the previous layer, depending on its type. Convolutional layers learn filters that are convolved with their input to produce a new feature map of the data. Pooling layers perform an operation such as max, average or L2 norm pooling, where the maximum, average or L2 norm of a filter region is computed, respectively. Finally, fully connected layers flatten the learned feature map into a vector before the output layer is applied for prediction.

Having said this, before ECG heartbeats can be classified by a CNN or any other model, two obstacles must first be addressed; (i) segmentation of ECG recordings into individual heartbeats (ii) significant class imbalance. ECG recordings must first be segmented since recordings can last several minutes or even hours, and hence may contain hundreds or even thousands of heartbeats. The most commonly used segmentation technique in the literature is the Pan-Tompkins algorithm, where the heartbeats are segmented such that the QRS complex is at the center of the signal. However, not only does this approach make assumptions about the signal morphology and require re-sampling to ensure all heartbeats have the same length, but it also removes a key feature from the signal, the RR interval or heart rate.

Further, ECG classification data-sets are highly imbalanced since diseased samples are much more limited than normal ones. This is a problem because it leads to ineffective training and a classifier that favours the majority class. Since it is the minority (diseased) classes that are of interest in this case, class imbalance must be addressed to develop a meaningful classifier. Traditional approaches involve augmenting the smaller classes using synthetic data before training. However, this increases computational cost and introduces new training data that may not match the true data.

1.4 Goal

In this work, the authors aimed to classify ECG recordings from the MIT-BIH arrhythmia and the St. Petersburg 12-lead arrhythmia (INCART) data-sets into five classes (N, S, V, F, Q) as defined by the Association for Advancement of Medical Instrumentation (AAMI). Their goal was to improve upon existing results by; (i) developing a deep CNN model for ECG heartbeat classification (ii) implementing a novel segmentation algorithm, first proposed by [9], that avoids re-sampling and assumptions about the signal morphology and maintains all important features, and (iii) exploiting a novel loss function, focal loss, first proposed by [13], to handle the class imbalance problem without the need for synthetic data.

2 Methods

2.1 Segmentation

As aforementioned, the traditional method for ECG segmentation, the Pan-Tompkins algorithm, extracts individual heartbeats by identifying the R-peak and centering the signal around it, as shown in Figure 2a. However, in order to include the RR interval as a classification feature, as well as avoid the need to re-sample

the data, the authors chose to use the segmentation algorithm proposed in [9], where a heart beat is defined as starting at an R-peak and ending after 1.2 times the median R-R interval in a 10-s window, as shown in Figure 2b. To extract heartbeats of this form, the following algorithm is suggested. First, the ECG signal is split into 10s windows and the amplitude of the signal in each window is normalized between 0 and 1. This allows R-peaks to be identified by finding the set of all local maximums and selecting those above a threshold of 0.9 (the R-peak has the highest amplitude in any ECG signal). The median RR interval (T) of the current 10s window is then identified, and a signal lasting 1.2T is extracted after each R-peak. After padding with zeros to ensure the signal length is equal to a predefined fixed value, the signal now constitutes a single heartbeat and can be used as input into the model. Having said this, a version of both raw data-sets used in this paper, where all R-peaks have been previously annotated, is freely available on the Physionet website [8]. This improves the above algorithm, and hence was used in our implementation of it, which was inspired by [9, 14]. It is unclear whether this annotated version was also used by the authors (see below).

2.2 Focal Loss

To avoid the need for synthetic data, the authors implement a novel loss function, focal loss, first proposed by [13] to handle class imbalance. In principle, focal loss is an extension of balanced cross-entropy loss that adjusts the contribution of each observation to the loss based on its classification error.

Cross entropy loss can be defined as

$$CE(y, p_c) = -log(P_c) \text{ where } P_c = \begin{cases} p_c \text{ if } y = c \\ 1 - p_c \text{ otherwise} \end{cases} \tag{1}$$

where p_c is the predicted probability of the observation of being in class c (i.e. of $y = c$).

Balanced cross-entropy loss is an extension of cross-entropy loss that addresses class imbalance explicitly by adding a factor α_c that weighs the loss associated with each observation relative to its class. In particular,

$$CE(y, P_c) = -\alpha_c \, log(P_c) \tag{2}$$

α_c can be chosen as the inverse of the class frequency or as a hyper-parameter. In this case,

$$\alpha_c = \begin{cases} \alpha \text{ if } y = c \\ 1 - \alpha \text{ otherwise} \end{cases}$$

where $\alpha \in [0, 1]$ is fixed. Indeed, α was chosen to be 0.25 by cross-validation (see Table 1).

Focal loss extends this even further by including a second weighting factor $(1 - P_c)^\gamma$ to handle class imbalance. In particular, miss-classified points are set to contribute more to the loss, while correctly classified points are made to contribute less. In this way, the function will implicitly on the challenging classes. Indeed,

$$FL(y, P_c) = -\alpha_c (1 - P_c)^\gamma log(P_c) \tag{3}$$

where γ is a hyper-parameter that determines the rate at which samples are down-weighted.

2.3 CNN Architecture

As aforementioned, the proposed ECG heartbeat classification model, described in Figure 3, is based on a CNN. In particular, the model contains two convolutional blocks, each composed of three 1-dimensional convolutional layers with 256 filters activation by ReLU activation functions. Each convolutional block is then followed by batch normalization and dropout operations.

Batch normalization normalizes the output of each convolutional block by subtracting the batch mean and dividing by the standard deviation. It then shifts and scales this normalized input using two learned parameters. By doing this each subsequent layer learns on a more stable distribution of inputs, which is believed to reduce co-variate shift, improve the stability and performance of the classifier and accelerate learning.

Dropout prevents over-fitting by randomly ignoring a certain proportion of nodes/neurons throughout training. A dropout value of 0.5 was shown to be optimal in this case (see Table 2).

Following the second round of batch normalization and dropout, the feature maps global average is computed by a global average pooling layer and a 128-node fully dense layer converts the map back into a 1D tensor. Finally, the $softmax$ activation function is used to compute the probability that the input belongs to each class.

2.4 Performance Metrics

While accuracy is the most common metric for evaluating ECG classifier performance, it is likely significantly distorted by the majority (N) class given the significant class imbalance. Hence, the use of F1-score, precision and recall, which can be defined for each individual class, are recommended by AAMI when evaluating ECG heartbeat classifiers. Hence, five performance metrics; accuracy, area under the curve (AUC), F1-score, precision and recall, were used to evaluate the performance of the model.

All metrics are based on a models confusion matrix, which summarizes the true positives (diagonal elements), false positives (column-sums, excluding the diagonal), false negatives (row-sums, excluding the diagonal) and true negatives (diagonal-sum, excluding the class of interest). In summary, high precision reflects a low false positive rate, high recall reflects a high true positive rate and F1-score reflects the balance between the two. AUC is the area under the receiver operator characteristic (ROC) curve, which plots the

true positive rate (recall) of a binary classifier as a function of the false positive rate. The per-class AUC in a multi-label classification problem can be obtained by taking a one-vs-rest approach, where all classes except the one of interest are grouped together.

3 Results

Both the original and reproduced ECG heartbeat classifier were developed using the Python programming language and the Keras deep learning library. The model was trained using the Adam optimizer, and a complete list of hyper-parameters used can be found in Table 1.

3.1 Data-set Descriptions

This paper focused on the MIT-BIH arrhythmia database [15] and the St. Petersburg 12-lead arrhythmia database (INCART) [8]. The MIT-BIH arrhythmia database is composed of 44 two-channel 360Hz ECG recordings, pre-divided into a training (DS1) and test set (DS2), each containing 22 records. In contrast, the INCART data-set contains 75 12-lead 257Hz ECG recordings that have not been pre-divided. In both data-sets, each heartbeat has been labeled as corresponding to 1 of 19 classes. These can be grouped into the 5 broader AAMI classes N, S, V, F, and Q using Table 1 from [9]. These are the classes of interest in this work, and are highly imbalanced, with an approximate inter-class ratio of $113 : 3.46 : 9.01 : 1 : 10$ (N : S : V : F : Q). As shown in Figure 4, which plots the first 50 heartbeats from the MIT-BIH arrhythmia database in each of the AAMI classes, there is also a large amount of intra-class variability.

In addition, a third data-set was used in this report, the PTB Diagnostic database [3], to extent the results of the paper. It is also freely available on the Physionet website [8], and contains 549 12-lead ECG records. However, unlike the aforementioned data-sets, each heartbeat has been labeled as corresponding to only two classes - normal (N) vs abnormal (A) with an inter-class ratio of $3003 : 7911$ (A : N). The first 50 heartbeats in each class are visualized in Figure 5, along with the average beat.

It is important to note that pre-processed and segmented versions of both the MIT-BIH arrhythmia database and the PTB Diagnostic database are available at [10] and were used to generate our results on these data-sets. In contrast, segmentation of the INCART data-set was done using our own implementation of the algorithm described in section 2.1.

3.2 Results for MIT-BIH Data-set

The classification results of the complete model (with focal loss) on the MIT-BIH test data-set [15] are visualized in Figure 6. However the main results presented in this paper focused on deducing the contribution of focal loss to ECG heartbeat classification. This was done by evaluating the proposed models performance without (model 1) and with (model 2) focal loss, where model 1 is trained using standard cross-entropy loss.

The authors showed that the model with focal loss (model 2) had a lower initial loss and converged must faster than the model without focal loss (model 1). This was consistent in our reproduction of the model, as shown in Figure 7. It should be noted that the slight sudden decrease in loss observed at the 80 epoch occurs due to fine tuning of the Adam optimizer, where the learning rate is reduced from 0.001 to 0.0001 (see Table 1).

As summarized by the confusion matrices shown in Table 2, focal loss, in general, improved the classification of minority classes. In particular, the number of correctly classified observations (blue diagonals) for the minority classes increased from Model 1 to Model 2.

However, the number of observations from the majority class N(0) that were miss-classified as belonging to one of the minority classes (false negatives) increased with focal loss in both the original paper and our reproduction. Despite this, the precision, recall and F1-score associated with the majority class N(0) all increased with focal loss (see Table 3). Hence, although the number of correctly classified normal beats (true positives) decreased, the number of abnormal beats that were incorrectly classified as normal (false positives) decreased even more. This was also consistent with our reproduction of the model.

The reproduced model gave similar results to the original paper on the MIT-BIH data-set with the exception of class S(1) (see Table 3). In particular for Model 1, the precision was much higher in the reproduction (blue), while the recall was much lower (red). This was also reflected by the confusion matrices presented in Table 2, where the number of false positives and true positives (for class S(1)) are both significantly lower in the reproduction of Model 1. However, the recorded variance associated with these metrics in the original paper are quite large (4.88 for precision, 1.73 for recall), and only a single example trial was trained in the reproduction. Hence, it may not be reflective of the average metric provided.

Moreover, once focal loss was added, the reproduced precision and recall of class S(1) significantly decreased and increased, respectively, becoming more comparable to the values listed in the original paper. Importantly, the F1-score increased with focal loss (from 78.25 to 79.18). Hence, congruent with the findings of the paper, the reproduced model with focal loss still outperforms the model without focal loss on the MIT-BIH database.

Finally, the authors estimated the ROC-curves for each of the 5 classes by computing the per-class true positive and true negative rate from the trial confusion matrix for Model 2 (Table 2). As shown in Figure 8 however, this approach significantly under-estimated the ROC-curves (and hence, AUC) of the minority classes when a binary classifier is trained using a one-vs-rest approach (solid curves).

3.3 Results for INCART Data-set

As aforementioned, unlike the MIT-BIH data-set, a pre-processed and segmented version of the INCART data-set was not freely available. Hence, to reproduce the papers results, the segmentation algorithm sug-

gested by the authors and described in Section 2.1 was implemented ourselves. However, the reproduced model performed poorly on this data-set, returning an accuracy of 79.08. This is very low in comparison to the mean accuracy value listed in the paper, namely 97.07. Given that the reproduced models results agreed with those of the paper on the pre-processed MIT-BIH data-set from [10], disparities between the pre-processing described in [9] and the pre-processing executed by the same authors on the data-sets from [10] were then explored.

Firstly, despite the claim that the proposed segmentation algorithm does not require re-sampling, the pre-processed data-sets from [10] were both down-sampled to 125Hz. To see if this led to improved results, down-sampling to 125 Hz was added to our implementation of the segmentation algorithm. The down-sampled and segmented heartbeats obtained using our implementation of the proposed algorithm on the INCART data-set are plotted in Figure 9. Despite the heartbeats appearing qualitatively similar to those of the MIT-BIH pre-processed data-set (Figure 4), the model's performance remained very poor, with an accuracy of 77.87 and an F1-score for all minority classes near 0 (S: 0.0065, V: 0.1181, F:0, Q:0). Moreover, the weighted AUC calculated using a one-vs-rest approach was approximately 0.5, meaning that the model performed no better than a random guess.

We then hypothesized that the model may be learning noise instead of key features, given that de-noising is often the first step of any ECG heartbeat classification algorithm and was not performed here (as it was not mentioned in the paper). Figure 10 plots a segmented heartbeat from the pre-processed data-set from [10] alongside a down-sampled heartbeat from the original MIT-BIH data-set segmented using the suggested technique. It is clear that noise has in fact been removed from the data that gave congruent results to those in the paper. This suggests that perhaps the paper used a de-noised data-set, however, given that no noise removal technique is mentioned, it was not possible to reproduce their results. De-noising was attempted using the Savitsky-Golay filter [20], however this did not improve the results of the model.

3.4 Results for PTB Diagnostic Data-set

Given that the PTB diagnostic data-set is available in the desired pre-processed form, we extended the results of this paper by validating the model on this third data-set. As shown in the confusion matrices given in Tables 4 and 5, the model performed very well with only 22 out of 3638 test data points miss-classified and returning an accuracy and weighted average F1-score of 99.40.

4 Discussion

In this paper, the authors presented a CNN-based architecture for ECG heartbeat classification that exploits a novel loss function, focal loss, to handle class imbalance and uses the segmentation algorithm proposed by [9] to avoid assumptions about the signal morphology and include the R-R interval as a feature. The model was shown to out-perform almost all benchmarks in the field on the MIT-BIH data-set in terms of accuracy, with the exception of models proposed by [19, 21, 12, 17, 5]. However, the model outperformed [12, 17, 5] in terms of F1-score, average precision and/or recall, which are better metrics for heavily imbalanced data-sets. In addition, the model proposed by [19] was a binary classifier and hence had an easier classification task. Moreover, as shown in this report, the proposed model performs very well on a binary classification task, returning a higher accuracy on the PTB Diagnostic data-set (99.41) than [19]. Finally, the model proposed in [21] had a higher computational cost and was trained on a larger training set.

Moreover, the proposed models results were shown to be reproducible on pre-processed and segmented data-sets available from [10], however non-reproducible on data-sets produced using the pre-processing techniques that were outlined in the paper. This was shown using the INCART data-set, where such pre-processed and segmented data-sets are unavailable and must be generated. This report highlighted to obtain the results shown, noise was likely removed from the ECG recordings before training in some way. It is also likely that the recordings were down-sampled from the their original frequencies to 125Hz. If the pre-processed and segmented data available from [10] were used to generate the results shown in the paper as our report highly suggests, then this should have been highlighted by the authors. However, even if this was not the case, more detail concerning the pre-processing methods used is required.

5 Conclusion

In conclusion, the authors presented a CNN-based approach for ECG heartbeat classification that exploits a novel loss function (focal loss) and segmentation algorithm [9] that reproducibly out-performs many state-of-the-art models in the literature when trained on freely accessible pre-processed (de-noised, segmented and down-sampled) data from the MIT-BIH database [10, 15]. However, given that details concerning these pre-processing techniques are not included in the report, reproducing their results verbatim was not feasible. This was especially clear when the model was trained using the INCART data-set. Future work should be done to see if implementing a noise removal technique [11] would improve the results on the INCART data-set. Moreover, model results on pre-processed data with focal loss could be compared to those obtained using more standard techniques to handle class imbalance, such as synthetic data generation. Finally, as each ECG recording contains up to 12 signals, it would be interesting to train the proposed model on each lead (instead of only one) and classify each beat according to the class received the most votes across all leads. For example, if 7/12 leads were classified as S, and 5/12 were classified as Q, then the beat would ultimately be given an S label.

6 Appendix

6.1 Figures

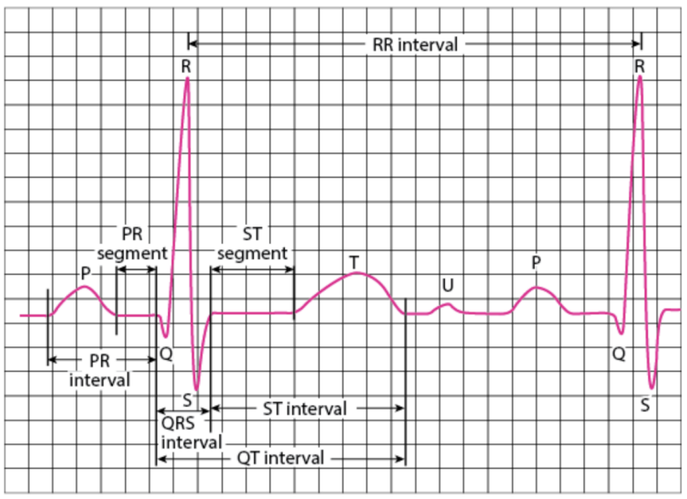

Figure 1: Schematic of an ECG Heartbeat. Figure obtained from [4]

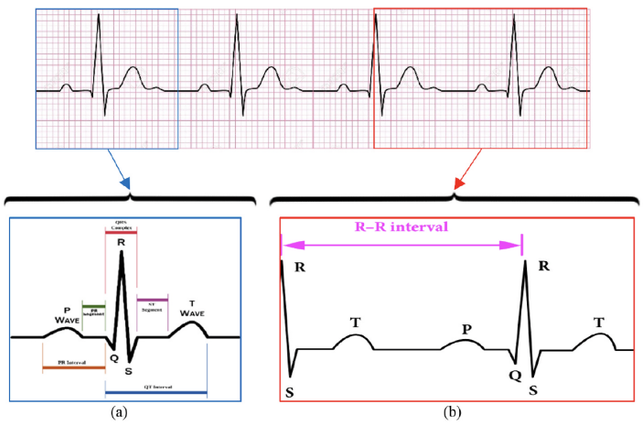

Figure 2: Comparison of ECG Segmentation Algorithms: a) Pan Tompkins algorithm; identifies the R-peak and centers the heartbeat around it b) Algorithm proposed by [9] and used in this study; heart beat starts at an R-peak and last 1.2 times the median R-R interval in a 10s window

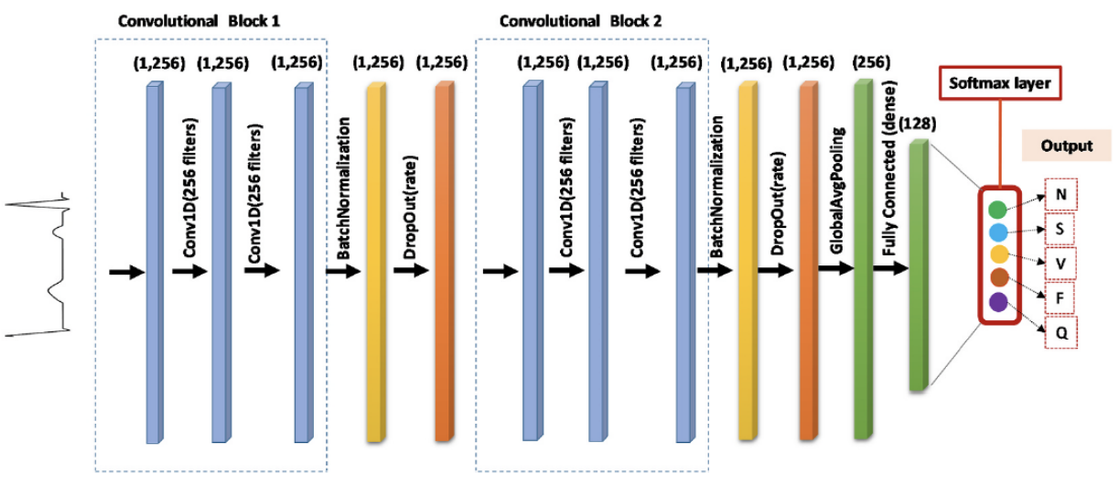

Figure 3: Architecture of proposed convolutional neural network model

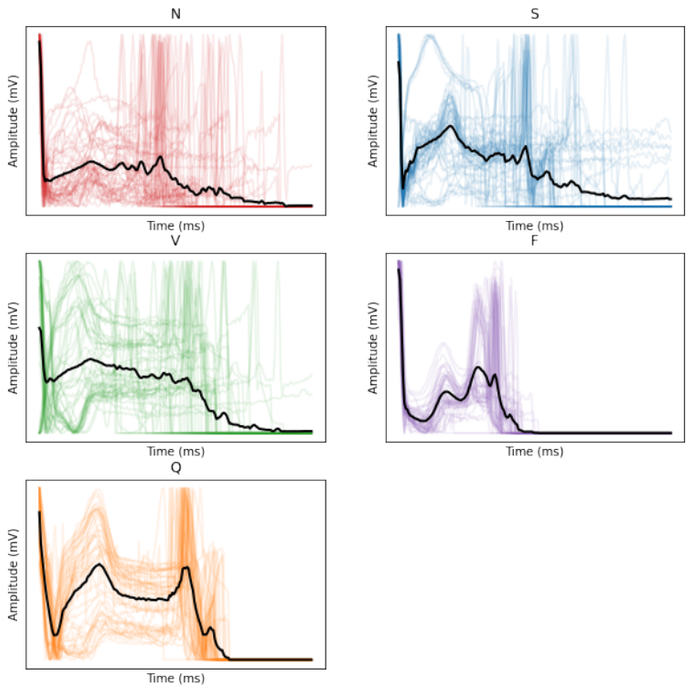

Figure 4: MIT-BIH Data-set: 50 sample beats from each AAMI class (N, S, V, F, Q) are plotted along with the average recording. This pre-processed and segmented version of the MIT-BIH data-set was provided by [10]

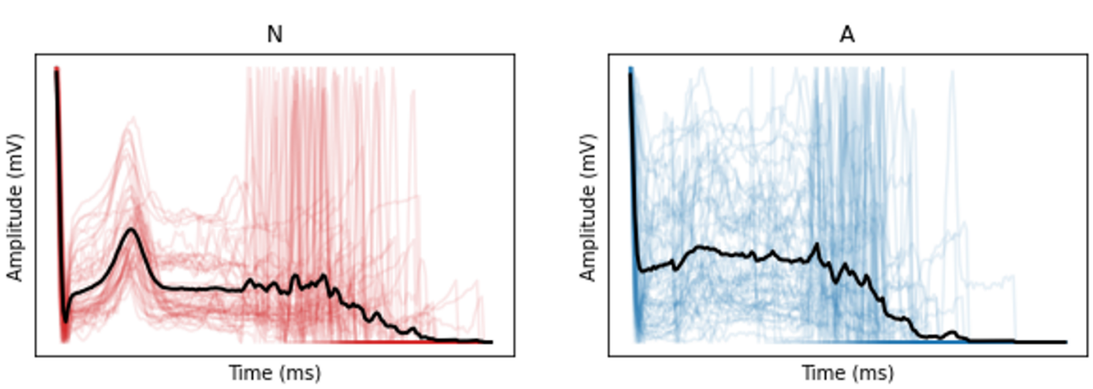

Figure 5: PTB diagnostic data-set: 50 sample beats from each class (N, A) are plotted along with the average recording. This pre-processed and segmented version of the PTB diagnostic data-set was provided by [10]

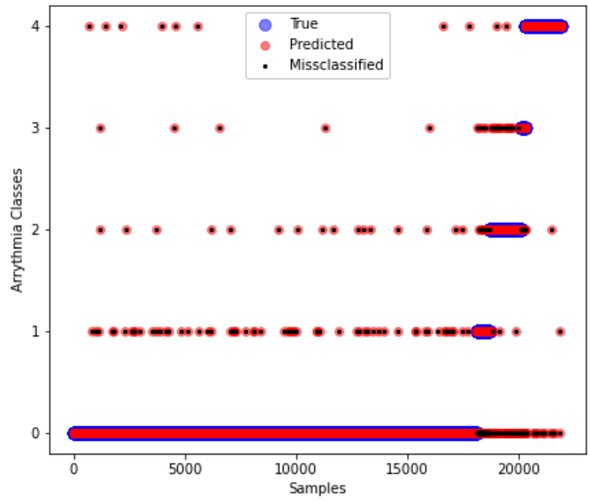

Figure 6: Classification of Test Data: the true (blue) and predicted (red) class labels for the MIT-BIH test data-set (DS2) available from [10] are plotted. The miss-classified points are indicated by a small black dot.

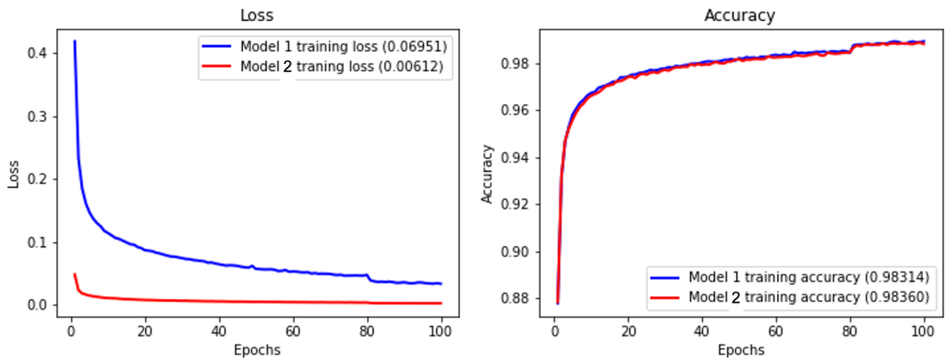

Figure 7: Training loss and accuracy of the proposed model with and without focal loss on the MIT-BIH data-set: loss drops and accuracy increases after 80 epochs due to fine tuning of the learning rate

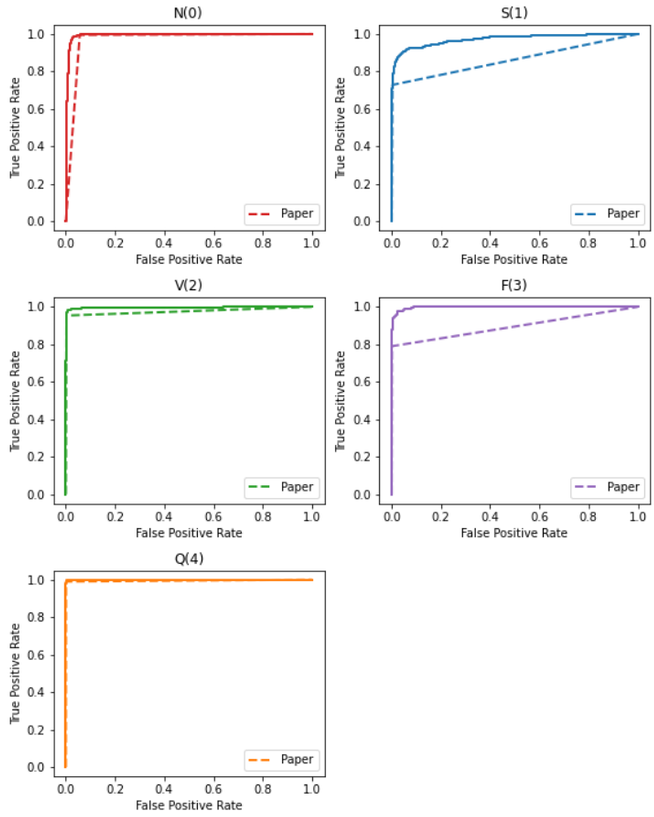

Figure 8: One vs. Rest Receiver-Operator Characteristic (ROC) Curves: The paper (dashed curves) significantly underestimates the one-vs-rest ROC curves (solid curves) by computing the false positive and true positive rate from the confusion matrix instead of training binary one-vs-rest models.

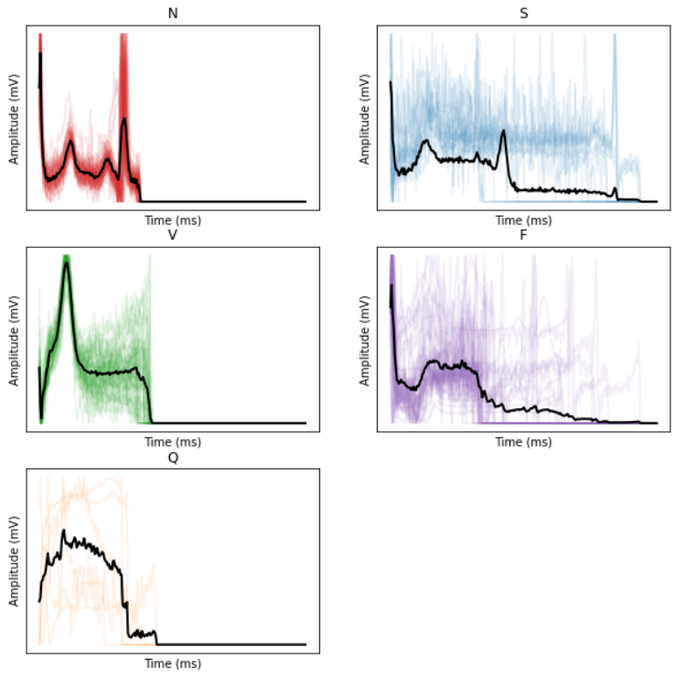

Figure 9: Segmented INCART Data: 50 sample beats from each AAMI class (N, S, V, F, Q) are plotted along with the average recording. Beats were generated using our own implementation of the alogirhtm proposed in [9] with down-sampling.

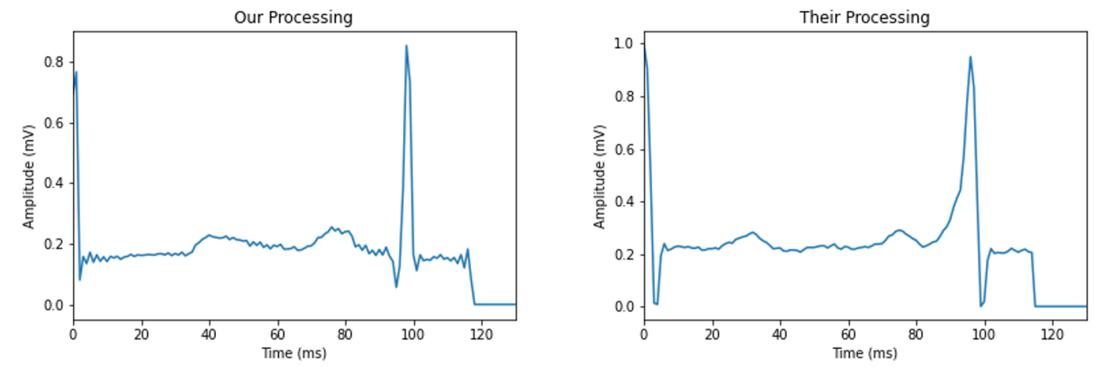

Figure 10: Noise comparison between MIT-BIH heartbeats pre-processed using our own implementation of the algorithm proposed in [9] with down-sampling (left) and in [10] (right)

6.2 Tables

Variable/ Hyper-parameter	**Value**
Filters	2
Filter Size	2
Stride	1
Batch size	400
Learning Rate	0.001
Epochs	100
γ	2
α	0.25
Dropout	0.5

Table 1: Model hyper-parameters and variables

		Model 1						**Model 2**						
		0	**1**	**2**	**3**	**4**	**FN**		**0**	**1**	**2**	**3**	**4**	**FN**
Paper	**0**	18051	35	24	4	4	**67**	**0**	18025	49	34	6	4	**93**
	1	124	425	6	0	1	**131**	**1**	110	433	11	0	2	**123**
	2	51	6	1377	11	3	**71**	**2**	44	11	1369	19	5	**79**
	3	21	0	19	122	0	**27**	**3**	17	0	12	133	0	**29**
	4	23	1	3	0	1581	**40**	**4**	18	1	4	1	1584	**26**
	FP	**219**	**49**	**52**	**15**	**8**		**FP**	**189**	**61**	**61**	**21**	**11**	
		0	**1**	**2**	**3**	**4**	**FN**		**0**	**1**	**2**	**3**	**4**	**FN**
Reproduction	**0**	18084	13	14	3	4	**34**	**0**	18027	58	16	5	12	**91**
	1	178	367	8	3	0	**189**	**1**	135	405	12	4	0	**151**
	2	54	1	1372	19	2	**76**	**2**	46	3	1381	16	2	**67**
	3	17	0	17	128	0	**34**	**3**	16	0	18	128	0	**34**
	4	29	1	5	1	1572	**36**	**4**	13	1	2	0	1592	**16**
	FP	**278**	**15**	**44**	**26**	**6**		**FP**	**210**	**62**	**48**	**25**	**14**	

Table 2: Confusion matrices with (model 2) and without (model 1) focal loss: false positives (FP) and false negatives (FN) for each class are added. The class labels (0,1,2,3,4) along the Y-axis denote the true labels, while the class labels along the X-axis denote the predicted labels

	Class	Model 1			Model 2		
		Precision	Recall	F1 score	Precision	Recall	F1 score
Paper							
	N	98.81 ±0.10	99.30 ±0.27	99.05 ±0.11	98.96 ±0.03	99.49 ±0.05	99.22 ±0.03
	S	85.39 ±4.88	77.09 ±1.73	80.89 ±1.40	87.65 ±1.51	77.88 ±0.42	82.48 ±0.72
	V	95.81 ±1.31	93.49 ±0.66	94.63 ±0.46	95.73 ±0.30	94.54 ±0.29	95.14 ±0.12
	F	81.11 ±3.44	79.63 ±3.68	80.22 ±0.92	83.65 ±0.89	82.10 ±1.21	82.87 ±0.79
	Q	98.52 ±0.76	98.33 ±0.23	98.43 ±0.36	99.31 ±0.10	98.51 ±0.15	98.91 ±0.12
	Avg	91.93 ±1.22	89.57 ±0.60	90.64 ±0.60	93.06 ±0.30	90.50 ±0.32	91.72 ±0.24
	Avg$_w$	98.11 ±0.17	98.14 ±0.19	98.11 ±0.18	98.37 ±0.06	98.41 ±0.06	98.38 ±0.05
	AUC$_w$			99.02 ±0.07			98.86 ±0.10
	Acc			98.14 ±0.19			98.41 ±0.06
Reproduction							
	N	98.49	99.81	99.14	98.85	99.50	99.17
	S	96.07	66.01	78.25	86.72	72.84	79.18
	V	96.89	94.75	95.81	96.64	95.37	96.00
	F	83.12	79.01	81.01	83.66	79.01	81.27
	Q	99.62	97.76	98.68	99.13	99.00	99.07
	Avg	94.84	87.47	90.58	93.00	89.15	90.94
	Avg$_w$	98.29	98.31	98.23	98.30	98.36	98.31
	AUC$_w$			99.43			99.41
	Acc			98.31			98.36

Table 3: Sample results on pre-processed MIT-BIH data-set from [10] with (model 2) and without (model 1) focal loss

	N	**A**
N	1037	6
A	16	2579

Table 4: Confusion matrix on PTB Data-set: the class labels (N, A) along the Y-axis denote the true labels, while the class labels along the X-axis denote the predicted labels

Class	Precision	Recall	F1 score
N	98.48	99.42	98.95
A	99.11	99.38	99.58
Avg	99.12	99.40	99.26
Avg$_w$	99.40	99.40	99.40
AUC			99.95
Acc			99.40

Table 5: Sample results on PTB data-set

References

[1] U Rajendra Acharya et al. "A deep convolutional neural network model to classify heartbeats". In: *Computers in biology and medicine* 89 (2017), pp. 389–396.

[2] Moussa Amrani et al. "Very deep feature extraction and fusion for arrhythmias detection". In: *Neural Computing and Applications* 30.7 (2018), pp. 2047–2057.

[3] R Bousseljot, D Kreiseler, and A Schnabel. "Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet". In: *Biomedizinische Technik/Biomedical Engineering* 40.s1 (1995), pp. 317–318.

[4] Thomas Cascino. *Electrocardiography*. https://www.msdmanuals.com/en-kr/professional/cardiovascular-disorders/cardiovascular-tests-and-procedures/electrocardiography. Accessed: 2020-12-14. 2019.

[5] Aiyun Chen et al. "Multi-information Fusion Neural Networks for Arrhythmia Automatic Detection". In: *Computer Methods and Programs in Biomedicine* (2020), p. 105479.

[6] Susan Cheng et al. "Long-term outcomes in individuals with prolonged PR interval or first-degree atrioventricular block". In: *Jama* 301.24 (2009), pp. 2571–2577.

[7] Robert Fagard. "Athlete's heart". In: *Heart* 89.12 (2003), pp. 1455–1461.

[8] AL Goldberger et al. "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals, June 13 2000http". In: *circ. ahajournals. org/cgi/content/full/101/23/e215* ().

[9] Mohammad Kachuee, Shayan Fazeli, and Majid Sarrafzadeh. "ECG Heartbeat Classification: A Deep Transferable Representation". In: *2018 IEEE International Conference on Healthcare Informatics (ICHI)* (June 2018). DOI: 10.1109/ichi.2018.00092. URL: http://dx.doi.org/10.1109/ICHI.2018.00092.

[10] Mohammad Kachuee, Shayan Fazeli, and Majid Sarrafzadeh. "ECG Heartbeat Classification: A Deep Transferable Representation". In: *2018 IEEE International Conference on Healthcare Informatics (ICHI)* (June 2018). DOI: 10.1109/ichi.2018.00092. URL: https://www.kaggle.com/shayanfazeli/heartbeat.

[11] Carlos Lastre-Dominguez et al. "ECG signal denoising and features extraction using unbiased FIR smoothing". In: *BioMed research international* 2019 (2019).

[12] Zhi Li et al. "Heartbeat classification using deep residual convolutional neural network from 2-lead electrocardiogram". In: *Journal of Electrocardiology* 58 (2020), pp. 105–112.

[13] T. Lin et al. "Focal Loss for Dense Object Detection". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42.2 (2020), pp. 318–327. DOI: 10.1109/TPAMI.2018.2858826.

[14] Andrew Long. *Arrythmia Project*. https://github.com/andrewwlong/deep_arrhythmias/blob/master/Arrythmia%20Project.ipynb. 2020.

[15] G. B. Moody and R. G. Mark. "The MIT-BIH Arrhythmia Database on CD-ROM and software for use with it". In: *[1990] Proceedings Computers in Cardiology*. 1990, pp. 185–188. DOI: 10.1109/CIC.1990.144205.

[16] Sajad Mousavi and Fatemeh Afghah. "Inter-and intra-patient ecg heartbeat classification for arrhythmia detection: a sequence to sequence deep learning approach". In: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2019, pp. 1308–1312.

[17] Saroj Kumar Pandey and Rekh Ram Janghel. "Automatic arrhythmia recognition from electrocardiogram signals using different feature methods with long short-term memory network model". In: *Signal, Image and Video Processing* (2020), pp. 1–9.

[18] Taissir Fekih Romdhane and Mohamed Atri Pr. "Electrocardiogram heartbeat classification based on a deep convolutional neural network and focal loss". In: *Computers in Biology and Medicine* 123 (2020), p. 103866.

[19] Giovanna Sannino and Giuseppe De Pietro. "A deep learning approach for ECG-based heartbeat classification for arrhythmia detection". In: *Future Generation Computer Systems* 86 (2018), pp. 446–455.

[20] Ronald W Schafer. "What is a Savitzky-Golay filter?[lecture notes]". In: *IEEE Signal processing magazine* 28.4 (2011), pp. 111–117.

[21] Ozal Yildirim et al. "A new approach for arrhythmia classification using deep coded features and LSTM networks". In: *Computer methods and programs in biomedicine* 176 (2019), pp. 121–133.