Multiclass Heartbeat Classification using ECG Signals and Convolutional Neural Networks

Muhammad Deedahwar Mazhar Qureshi
School of Electrical Engineering
and Computer Sciences
National University of Sciences
and Technology, Islamabad, Pakistan
Email: mqureshi.bese18seecs@seecs.edu.pk

Daniel Peralta Cámara
Department of Information
Technology at
Ghent University, imec
Ghent, Belgium
Email: daniel.peralta@imec.be

Eli De Poorter
Department of Information
Technology at
Ghent University, imec
Ghent, Belgium
Email: eli.depoorter@imec.be

Rafia Mumtaz
School of Electrical Engineering
and Computer Sciences
National University of Sciences
and Technology, Islamabad, Pakistan
Email: rafia.mumtaz@seecs.edu.pk

Adnan Shahid
Department of Information
Technology at
Ghent University, imec
Ghent, Belgium
Email: adnan.shahid@imec.be

Ingrid Moerman
Department of Information
Technology at
Ghent University, imec
Ghent, Belgium
Email: ingrid.moerman@imec.be

Timo de Waele
Department of Information
Technology at
Ghent University, imec
Ghent, Belgium
Email: timo.dewaele@imec.be

***Abstract*—Given a large enough time series signal from an ECG signal, it is possible to identify and classify heartbeats not only into normal and abnormal classes but into multiple classes including but not limited to Normal beat, Paced beat, Atrial Premature beat and Ventricular flutter as originally suggested by benchmark electrocardiogram (ECG) datasets like the MIT-BIH Arrhythmia Dataset. There are multiple approaches that target ECG classifications using Machine and Deep Learning like One Class SVM, ELM, Anogan etc. These approaches require either very high computational resources, fail to classify classes apart from normal/abnormal classes or fail to classify all classes with an equivalent or near-equivalent accuracy. With these limitations in mind, this paper proposes a deep learning approach using Convolutional Neural Networks (CNNs) to classify multiple classes of heartbeats in an efficient, effective, and generalized manner. By using the MIT-BIH Arrhythmia dataset to filter and segment individual correctly structured heartbeats, we have designed a network which can be trained on different classes of heartbeats and present robust, accurate and efficient results. The class imbalance prevalent in the MIT-BIH dataset has been dealt with using Synthetic Minority Over-sampling Technique (SMOTE). The robustness of the model is increased by adding techniques of loss minimization such as dropout and early stopping. The approach gives an accuracy of approximately 96% and an extremely short time span for class prediction(classification), i.e., less than 1 second. The results are also illustrated over multiple (10) classes to exemplify the generality of the model. We have illustrated these results over multiple (10) classes to exemplify generality of the model.**

***Index Terms*—ECG, Convolution Neural Networks, Deep Learning, Supervised Learning**

I. Introduction

Non-threatening and treatable Arrhythmia kills about 250,000 people every year according to a research by Duke University [13]. These types of arrhythmias occur suddenly and are hard to detect. An Arrhythmia can be defined as an irregular heartbeat that occurs due to faulty neural signals to the heart. Arrhythmias are tachycardia (fast), bradycardia (slow) and irregular. A heartbeat is classified as regular, or irregular based on its beat structure. The structure is composed of a P-wave, QRS complex, T and U waves [14]. The total length of a beat is 0.6 to 0.8 seconds on the time axis. The sturcture of a measured heartbeat is shown in Figure 1. According to the MIT-BIH Arrhythmia database [7], a heartbeat can be categorized in 16 classes, 6 of which constitute a broader definition of a “Regular” or “Normal” heartbeat but can still cause arrhythmias. Li, Zhi et al [15] characterizes the biggest issues faced in heartbeat classification are structural variations within classes and among patients. Additionally, many classes are so rare, they are practically unseen in training data which is naturally unbalanced.

These problems can be solved with a Neural Network that can not only learn based on the ECG signal but also store long term dependencies and information of the data. These Neural Networks can be Recurrent Neural Networks (RNNs) [5] like Long-Short Term Memory (LSTM) [3] or a Convolutional Neural Network [4]. For training these networks, the datasets must be balanced which is unlikely due to the natural rarity

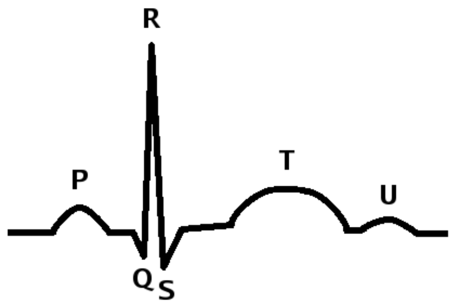

Fig. 1: Labeled Structure of a Heartbeat

of certain classes or using a reconstruction based approach for anomaly detection, the model can be trained on normal classes as a regenerative auto-encoder or a Generative Adversarial Network.

There are numerous challenges faced in the field of ECG classification by medical practitioners as well as Machine Learning models. Abnormalities present in time-series heartbeat data need to be explained. Therefore, there is a need to classify these abnormalities separately into multiple classes as not all abnormalities exhibit a similar level of risk and not all abnormalities should be grouped together. This challenge paves a way for the need to classify multiple ECG classes.

There are further challenges associated with multiclass ECG Classification. Massive ECG time-series often do not contain enough minority classes. This leads to insufficient and imbalanced data. Additionally, within a class, there might be variations of ECG representations. This intra-class variation proves to be a challenge for training a robust model that can avoid misclassifying such variations. Finally, even in healthy patients the structure of heartbeats belonging to various classes can differ. This leads to characteristic differences in beat structure like R-R intervals, P-R intervals amongst beats of the same patient.

This paper aims to tackle these issues with robust techniques of dropout for limiting the number of nodes in the fully-connected layers of the network and early stopping to prevent over-fitting on noise as the techniques of loss minimization in the network. For techniques of Data Augmentation, this paper utilizes SMOTE [9], which Synthetically Regenerates datapoints using the KNN [1] algorithm. Experiments thus show that the approach detects multiple classes of heartbeats with a high accuracy and with well established efficiency i.e Classification Time.

The main contributions of the work are as follows:

1.Effectiveness: A robust Multiclass heartbeat classification that utilizes deep learning techniques using the CNN network structure with loss minimization for robustness and a synthetic modification of the benchmark MIT-BIH dataset. This modification was made using SMOTE to improve the class balance in the dataset. The model can accurately classify multiple classes of heartbeats (96%).

2.Efficiency: The model can make fast inferences and predictions with less than 1 second required to successfully classify a heartbeat into its respective class.

3.Generality: This paper uses the definition of generality used by [19]. This defines generality as the ability of a model to provide a balanced or generalized approach towards classification within the scope of the application domain. The model can classify 10 classes of ECG data accurately as opposed to two-class classification models.

The organization of this paper is given as: Literature Review, Methodology, Analysis and Results, Conclusion, References.

II. Related work

The literature of this subject is divided into two categories:

A. Classification Models

These models require labeled data for training. With enough training data, these models are often able to obtain high scores in terms of testing accuracy and in some cases have proven to be efficient as well in terms of class prediction time. However, these models often fail to catch edge-case heartbeats. These heartbeats lie on the boundaries of class divisions and are often misclassified by such models. Moreover, such models also require an extensively balanced dataset to be able to classify minority classes accurately. Due to this reason, many models opt to classify heartbeats into normal/abnormal classes to achieve an improved balance in the data they gather from personal sources as well as benchmark datasets.

One-Class SVM (Scholkopf et al 2000) [8] is a model which aims to learn/train on normal heartbeat data to be able to identify anomalies found in abnormal heartbeats. OCSVM is one of the initial classification works on ECG data that gained identification as an unsupervised approach to ECG classification using a modified application of Support Vector Machines, which are typically considered Supervised Learning Models.

Xu et al. 2021 [18], proposed multiclass heartbeat classification using ELM (Extreme Learning Machine). ELM is an artificial neural network having a single hidden layer. This model reduces the training time requires with the least-square algorithm due to its high training accuracy. The heartbeats acquired in this model are first decomposed using discrete wavelet transform (DWT) to acquire time-frequency features. Additionally, this model uses R-R peak intervals as features in the time domain, essentially not utilizing a raw signal but using feature extraction to improve train time at the expense of test time efficiency.

B. Reconstruction Models

Many models aim to detect abnormalities in any dataset by computing a synthetic reconstruction of the data presented. These models are often only used for single-class classification and can only classify heartbeats as normal or abnormal. There are numerous techniques that are used for this purpose.

Principal Component Analysis (PCA) [2] can be used for this reconstruction, but it exhibits a limitation. PCA can only reconstruct data in a linear space. This low representation of

Model/Property	Efficient	Non-Linear	Robust	Generalized
OCSVM [8]	✓	✓		
PCA [11]	✓		✓	✓
AnoGAN [12]		✓	✓	
ELM [18]		✓	✓	✓
Proposed Model	✓	✓	✓	✓

TABLE I: Comparison of Related Literature

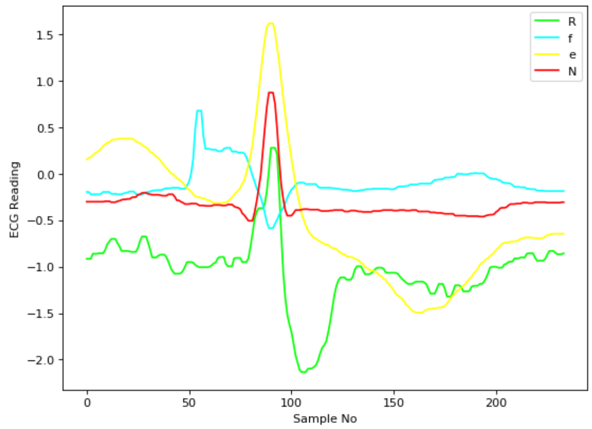

Fig. 2: Variation amongst different heartbeat types.

data such as in FBOX (Shah et al 2014) [11] can easily lead to overfitting without precautions for robustness. This ultimately results in a low accuracy and generality in experimental setups. Recently, with growing interest in generative adversarial networks, there have been proposals for abnormality detection using adversarial training techniques. AnoGAN (Schlegl et al, 2017) [12] is such a model which aims to learn abnormality detection using visual inputs. AnoGAN has a requirement to learn a singular latent vector for each input it encounters making it very slow at runtime thus limiting its applications. This paper provides an efficient, effective, and generalized approach for heartbeat classification by incorporating deep learning techniques of CNNs with effective robustness. The table given at TABLE I summarizes the existing works that relate to our problem and show that this paper satisfies all the desired characteristics.

III. Methodology

A. Dataset

The dataset used for this paper is the MIT-BIH Arrhythmia Dataset [7] which was started to be compiled in 1975 at Beth Israel in Boston, Texas and was published by Physionet [6] in 1980. It is an open-source benchmark dataset consisting of 48 half hour recordings from 47 different patients using different machines. The recordings are digitized at a frequency of 360 Hz with an 11-bit resolution over a range of 10mV.
The MIT-BIH dataset has 16 annotated heartbeat classes at each R-Peak. The description for each label of classes is in Table II:

Label	Description
N	Normal
L	Left bundle branch
R	Right bundle branch
A	Atrial premature
V	Premature ventricular
/	Paced
a	Aberrated atrial premature
!	Ventricular flutter
F	Fusion of ventricular and normal
x	Blocked atrial premature
j	Nodal escape
f	Fusion of paced and normal
E	Ventricular escape
J	Nodal premature
e	Atrial Escape
Q	Unclassifiable
Total	16

TABLE II: Description of Labels in the MIT-BIH Arrhythmia Dataset

B. Heartbeat Segmentation

Each recording was read and converted into a raw signal reading. The ML-II lead was used for reading the heartbeats which is placed on the upper rib in normal cases with a few notable exceptions. The records were passed through a median filter like of window size 5 to reduce noise encountered commonly when dealing with such ECG datasets like Xe et al [18]. Each recording was then segmented into multiple 234 sample heartbeats with centered R-peaks. The structure for each heartbeat was taken from ELM with 0.25 seconds before the R-peak and 0.4 seconds after the heartbeat as given by the inference below:
For every R-peak at point y, the beginning point of the heartbeat is given as:

$$begin = y - 0.25(f) \tag{1}$$

and the ending point of the heartbeat is given as:

$$end = y + 0.4(f) \tag{2}$$

Here, f is the sampling frequency of the ECG Monitor. In case of the MIT-BIH, the sampling frequency is 360 Hz.

Each heartbeat structure is characterized by a P-wave, a QRS complex and T-wave, and U-wave. The R-peak values are pre-annotated in the dataset [7], so the structure of each beat was centered at the local R-peak value. Heartbeats with R-peaks close to either end of the recording was discarded due to insufficient data samples available for complete heartbeat construction.
A sample segmented heartbeat from a single record is given in Figure 3.

C. Synthetic Minority Oversampling Technique

Chawla et al [9]. proposed SMOTE in 2002 as an approach to classification model construction from datasets that are imbalanced in nature. The nature of real world datasets tend to contain a very minute number of abnomralities and the cost of misclassifying these can be immense. The method shows

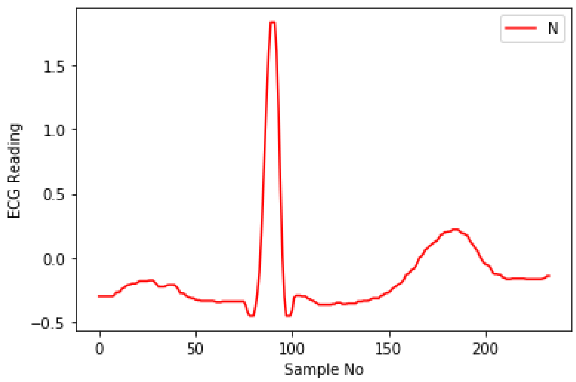

Fig. 3: Structure of a singular Heartbeat with ECG Magnitudes(Vertical axis)

over-sampling the minority and under-sampling the minority class can create favorable results for classifier performance.

The MIT-BIH dataset is a highly imbalanced dataset [7]. The major reason for this is that many cases of heartbeats do not occur commonly. The dataset prepared post segmentation had an overwhelming majority of 75,016 Normal Heartbeats. Many of these edge cases and minority classes are misclassified as the model does not encounter enough of their samples in the training set. Such misclassification can be detrimental to patient health and early myocardial infarction detection which is divisive in the medical realm. To ensure that such misclassifications do not take place, this paper utilizes SMOTE.

SMOTE is a technique commonly used in Machine Learning and Deep Learning when dealing with imbalanced datasets. Rather than simply duplicating minority class data in the training set, SMOTE works by focusing on examples that are close to each other in the feature space. A random sample is chosen first from the minority class. K nearest neighbors of that example are then found, typically we take K=5. One these neighbors is then chosen, and a line is drawn between the randomly chosen sample and the randomly chosen neighbor. Therefore, SMOTE is basically based on KNN classification algorithm [1]. Once connected, a random point is taken along this connection in the feature space and a synthetic example is created at this point. This procedure is repeated until the minority class has as many samples in its set as the majority class. A key suggestion in the paper is to first trim the majority class using random under-sampling before oversampling the dataset with SMOTE.

With SMOTE, that dataset now consists of a training set of 623818 samples and a validation set of 307255 samples with a 1/3 train-validation split. The split was done in a stratified way to prevent intermixing of instances in the training and validation sets.

Before SMOTE, the shape of the dataset in terms of class distribution is given in Table III:

Label	**Number of Samples**
N	71621
R	7149
/	6998
L	6594
V	6167
A	2547
f	1777
J	311
Q	34
e	16
Total	**103214**

TABLE III: Number of Samples per Class

The classes with the number of occurences less than 5 have been grouped into unclassifiable heartbeats due to the lower limit of samples being more than 5 present in SMOTE. With this class distribution and management, the dataset is completely balanced and prepared to train a classifier.

D. *Classification Model*

One of the most common signal processing techniques is convolution which is defined as a mathematical operation performed on two functions. It is a commutative operation that denotes how the shape of one function is changed or affected by the other. The convolution operation is represented by the * symbol. Convolution of two functions f and g is given as:

$$(f * g)(t) := \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau \tag{3}$$

LeCunn et al [4]. proposed Convolutional Neural Networks, a technique of shape based recognition and classification using minimally processed images (and signals) rather than feature annotation in 1999. This technique has since proven to be advantageous for classification of images and signals. These networks, when fed with raw signals or images are able to extract the right set of useful features and use them for object-recognition as well as image/signal classification.

The model used for the classification is a CNN model with 3 1-D Convolution units, followed by 1D-MaxPooling and a Dropout to avoid overfitting the dataset as well as make the model robust. Each Convolution layer is given a leaky-ReLU activation with a leak of 0.001. The mathematical expression for the activation function leaky-ReLU is given as f(x):

$$f(x) = \max(ax, x) \tag{4}$$

Here, a is defined as the leak. In the case of this model, the leak is set to 0.001.

The 1-D Convolution Kernal Size and the number of each layer in the convolution network is given as 32(3/1)-64(3/1)-128(3/1) where 32 is the number of filters, 3 in the size of the filter and 1 is the step-size or stride. Padding for each layer was set to ‘same’. The Convolution Layers are followed by a 1-D MaxPool layer of pool size 3 and stride of 2 with the same padding followed by a dropout of 0.5.

The model is flattened followed the 2 fully connected dense layers of size 256 and 512 respectively with leaky ReLU activation of leak 0.001. The final output/classification layer is then added with softmax activation. The size of this layer is 10 which signifies the 10 classes the model aims to predict. The structure of the Multi-Layer-Perceptron (MLP) is given as:

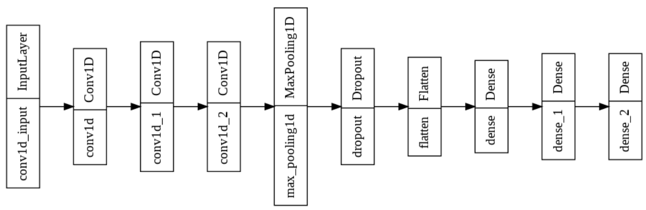

Fig. 4: Structure of the Proposed CNN Model

The model is optimized using the Adam optimization algorithm. This is a replacement algorithm for the traditional SGD and is used for training deep models. Adam aims to combine the best properties from two algorithms namely the AdaGrad and RMSProp to enable an algorithm to handle noisy problems which can cause sparse gradients. The learning for the optimizer was set to 0.001. Adam was chosen as the optimizer for this model due to its relatively user-friendly initialization and faster convergence as compared to other algorithms like SGD (Stochastic Gradient Descent).
Loss minimization of the model is carried out by minimizing the Sparse Categorical Cross Entropy Loss. The mathematical expression for Cross-Entropy loss is given as:

$$L_{CE} = -\sum_{i=1} T_i log(S_i) \tag{5}$$

Here, T is the indicator which indicates whether the class label is correct, its value is 1 if the label is correct and 0 if the label is incorrect while calculating loss. S gives the predicted probability that the member lies under the class label T.
The model was trained for 7 epochs with a batch size of 64 giving each epoch 9748 batches of the training data. Cross Validation technique was adopted to ensure the best-fit solution for the problem.
The loss encountered whilst training the model for 10 epochs is given in Figure 5 and the accuracy is given in Figure 6.

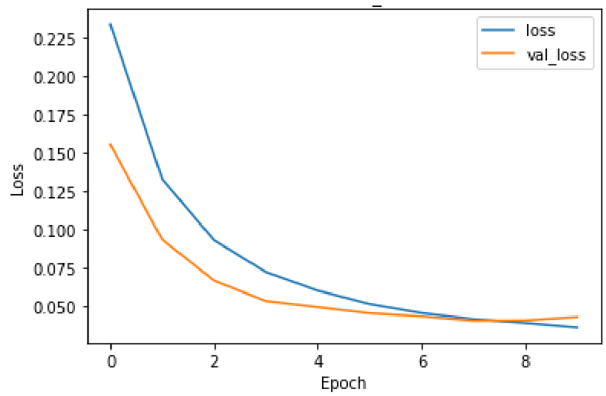

Fig. 5: Plot of Training vs Validation Loss of the model.

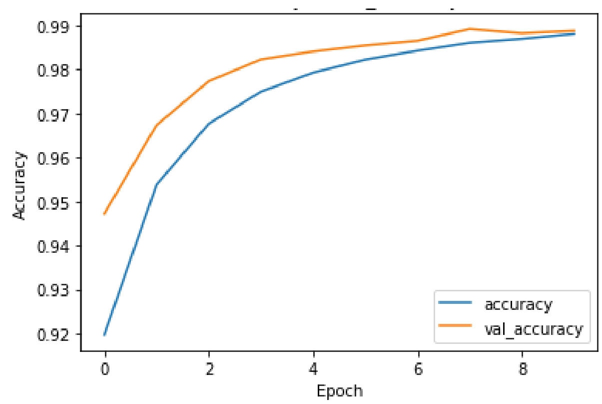

Fig. 6: Plot of Training vs Validation Accuracy of the model.

IV. Analysis and Results

A. Accuracy

The model is tested on the test set of the MIT-BIH. The test set in a unbalanced set of heartbeats that is similar to data encountered in experimental settings in terms of class disparity. The original MIT-BIH dataset was given a train-test split of 1/3 in 5 different iterations. The model was then tested on each test set and an accuracy score was retrieved. The accuracy score obtained by testing the model on these iterations is 95.7% with a standard deviation of 0.03%. In comparison with other bi-class and multiclass ECG classification models, the performance of this model is given in Table IV:

Model	Accuracy
OCSVM [8]	81.2%
PCA [11]	83.6%
AnoGAN [12]	90.2%
ELM [18]	93.4%
Proposed Model	95.7%

TABLE IV: Comparison of Prediction Accuracy of Discussed Models

B. Generality

The model can distinctly identify all 10 classes of the data with a significant accuracy. Generality is this paper is defined within the scope of ECG classification. This is therefore limited to the classification of the 10 classes defined below in the class-precision table. Each of these classes is taken from the MIT-BIH Arrhythmia dataset. As explained earlier, classes with insufficient samples were classified under the 'Q' class which denotes Unclassified Heartbeats. The performance of the model on each of the 10 classes is given in Table V:

Label	Accuracy in Prediction
N	97.8%
R	96.6%
/	94.2%
L	95.4%
V	94.1%
A	96.9%
f	92.8%
J	96.1%
Q	95.5%
e	95.2%

TABLE V: Prediction Accuracy per Class of Proposed Model

C. Efficiency

The model gives a training time of 4.8-5 hours with mean estimated time of 1730 seconds per epoch of training. This training was carried out using the standard Google Colab CPU and RAM without using GPU. The Google Colab system dedicates an Intel(R) Xeon(R) 2.3 GHz Haswell Dual-Core CPU with 12 GB RAM for a notebook. The prediction time for each heartbeat is very low compared to other systems of predictions present in the literature. The proposed model is capable of classifying a heartbeat in 0.6 seconds due to the light CNN structure as compared to other state-of-the-art models like AnoGan, ELM etc, while it is not an oversimplified approach like PCA, OCSVM which would lead to misclassifications.
The comparison of prediction times of other models discussed in the literature is given in Table VI:

Model	Prediction Time(seconds)
OCSVM [8]	0.87
PCA [11]	0.34
AnoGAN [12]	3.6
ELM [18]	8.4
Proposed Model	0.6

TABLE VI: Prediction Time of Discussed Models

V. Conclusion

To conclude, this paper proposes a deep learning model for multiclass ECG classification with additive robustness. The proposed model aims to address the shortcomings of current approaches by focusing on two major problems. i.e:
1) Class Imbalance.
2) Variation amongst Binary Classes (Normal and Abnormal classes contain a wide distribution of classes).
The paper tackles these issues by using various techniques that have been previously implemented. This includes using Minority Oversampling to tackle imbalanced classes as explained earlier. This enabled the model to be trained on sufficient samples to properly classify heartbeats.
In light of the literature discussed in the paper, the proposed model aims to make the following contributions:
1) 10 Classes can be inferred through this model. These classes cover most of the major as well as minor classes of heartbeats encountered by medical practitioners and especially ones which have proven fatal in most heart attacks.
2) The model outperforms baseline models with an approximate 96% accuracy and very fast inference of less than 1 second.
3) The model is made robust with dropout and early stopping. This makes the model efficient, generalized, robust as well as effective.

VI. Acknowledgements

The research material of this paper is based on and supported by the Arrhythmia Detection using ECG classification Project at SEECS-NUST. The work is also supported by the Erasmus+ programme of the European Union through the “Capacity building for Digital Health Monitoring and Care Systems in Asia (DigiHealth-Asia), 2021-2024” project under grant agreement no 619193-EPP-1-2020-1-BE-EPPKA2-CBHE-JP.

References

[1] Cover, Thomas, and Peter Hart. ”Nearest neighbor pattern classification.” IEEE transactions on information theory 13.1 (1967): 21-27.
[2] Wold, Svante, Kim Esbensen, and Paul Geladi. ”Principal component analysis.” Chemometrics and intelligent laboratory systems 2.1-3 (1987): 37-52.
[3] Hochreiter, Sepp, and Jürgen Schmidhuber. ”Long short-term memory.” Neural computation 9.8 (1997): 1735-1780.
[4] LeCun, Yann, et al. ”Object recognition with gradient-based learning.” Shape, contour and grouping in computer vision. Springer, Berlin, Heidelberg, 1999. 319-345.
[5] Medsker, Larry R., and L. C. Jain. ”Recurrent neural networks.” Design and Applications 5 (2001): 64-67.
[6] Goldberger, A., et al. ”PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220.” (2000).
[7] G.B. Moody and R.G. Mark, ”The impact of the MIT-BIH arrhythmia database”, IEEE Eng. Med. Biol. Mag., vol. 20, no. 3, pp. 45-50, 2001
[8] Schölkopf, Bernhard, et al. ”New support vector algorithms.” Neural computation 12.5 (2000): 1207-1245.
[9] Chawla, Nitesh V., et al. ”SMOTE: synthetic minority over-sampling technique.” Journal of artificial intelligence research 16 (2002): 321-357.
[10] Srivastava, Nitish. ”Improving neural networks with dropout.” University of Toronto 182.566 (2013): 7.
[11] Shah, Neil, et al. ”Spotting suspicious link behavior with fbox: An adversarial perspective.” 2014 IEEE International conference on data mining. IEEE, 2014.
[12] Schlegl, Thomas, et al. ”Unsupervised anomaly detection with generative adversarial networks to guide marker discovery.” International conference on information processing in medical imaging. Springer, Cham, 2017.
[13] Jackson, Larry R., et al. ”Sinus node dysfunction and atrial fibrillation: a reversible phenomenon?.” Pacing and clinical electrophysiology 40.4 (2017): 442-450.
[14] Rosli, Khairul Affendi, Omar, Hafizi, Hasan, Ahmad Fariz, Musa, Khairil Syahmi, Muhamad Fadzil, Mohd, Neu, Shu. (2018). Development of Electrocardiograph Monitoring System. MATEC Web of Conferences.
[15] Li, Zhi, et al. ”Heartbeat classification using deep residual convolutional neural network from 2-lead electrocardiogram.” Journal of Electrocardiology 58 (2020): 105-112.
[16] Black-Maier, Eric, et al. ”Catheter ablation of atrial fibrillation in cardiac amyloidosis.” Pacing and Clinical Electrophysiology 43.9 (2020): 913-921.
[17] Li, Mingchen, Mahdi Soltanolkotabi, and Samet Oymak. ”Gradient descent with early stopping is provably robust to label noise for overparameterized neural networks.” International conference on artificial intelligence and statistics. PMLR, 2020.
[18] Xu, Yuefan, et al. ”Extreme Learning Machine for Heartbeat Classification with Hybrid Time-Domain and Wavelet Time-Frequency Features.” Journal of Healthcare Engineering 2021 (2021).
[19] Raji, Inioluwa Deborah, et al. ”AI and the everything in the whole wide world benchmark.” arXiv preprint arXiv:2111.15366 (2021).
[20] Kumar, Abhishek, et al. ”IoT-based ECG monitoring for arrhythmia classification using Coyote Grey Wolf optimization-based deep learning CNN classifier.” Biomedical Signal Processing and Control 76 (2022).
[21] Chen, Shan Wei, et al. ”Review of ECG detection and classification based on deep learning: Coherent taxonomy, motivation, open challenges and recommendations.” Biomedical Signal Processing and Control 74 (2022).