iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.44095

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

Robust Classification of Cardiac Arrhythmia Using Machine Learning

Yashaswini T G[1], Dr. Ravi Kumar G K[2], Ms. Sindhu D[3]
[1]*Student, Professor,* [2]*Head(R&D),* [3]*Assistant Professor, Dept. of CSE, BGS Institute of Technology, Adichunchanagiri University, BG Nagar, Karnataka, India-571418*

Abstract: ***Machine Learning (ML) is developing progressively in the healthcare industry. The classifying of electrocardiogram (ECG) results depending on Cardiac Arrhythmia is one such instance, where ECG record's the heart's rhythm & electric activity. Two ML techniques, Random Forest (RF) and Convolution Neural Network (CNN) are employed in the presented work to yield rapid & effective categorization of heartbeats. The heartbeats are divided into five classes applying the PTB Diagnostic ECG Database plus the MIT-BIH Arrhythmia Collection. Both datasets are massive, with data being unbalanced for all five classes. The balancing of data is done using both oversampling & under sampling approaches to bring the data equal for each class. Based on performance indicators like f1-score, precision, recall, & accuracy, a comparison will be done among RF and CNN methods.***

Keywords: ECG, Cardiac Arrhythmia, RF, CNN

I. INTRODUCTION

Cardiac arrhythmia is among the more common cardio conditions with potentially fatal effects. Conventional electrocardiography (ECG) equipment, on the other hand, has such a challenging time capturing arrhythmia signs throughout hospitalization appointments since they happen infrequently. To fix the issue, one group of experts has proposed continual surveillance devices. Unfortunately, certain practical challenges, such as poor capacity, offline information collecting & analysis, and excessive power usage, could protect this method from becoming extensively adopted.

Including an approximately 300 million ECG signal generated yearly, different cardiac disorders such as Myocardial Infarction, AV Block, Ventricular Tachycardia, and Atrial Fibrillation can already be identified via ECG readings. Let's look into the ability to diagnose arrhythmias utilizing an ECG report. This is considered a difficult assignment to systems, though a professional could typically figure it out with a single, well-placed lead.

Due to the significant failure levels of computerized interpretations, arrhythmia identification using ECG records is normally done by skilled technicians and cardiologists. Only around half of all computerized forecasts for non-sinus beats were right in one study, and only one out of every seven displays of second-degree AV block were properly acknowledged by the program in some other. A computer should absolutely decide the different signal patterns and identify the intricate interactions among throughout the period designed to check heart arrhythmias in an ECG. Because of the heterogeneity in signal morphology across sufferer and the incidence of noise, this is problematic.

Mostly in the United States, heart arrhythmias remain some of several top reasons for cardiovascular disease (CVD) in both men and women. [1] Electrocardiogram (ECG) equipment including that kind of Holter monitors, loop recorders, as well as surgically implanted cardioverter defibrillators (ICD) is commonly used to diagnose and treat these diseases. The invention of a simple, standard, real-time ECG gadget for the recognition and monitoring of ventricular tachycardia was discussed in the article (VT). Utilizing Laboratory VIEW's biomedical toolset, a controlling application was developed and tested utilizing simulated ECG signals. The sensitivity and accuracy of numerical simulations are 97.6% and 98.0 percent, accordingly.

Variations throughout the regular rhythm of a person's heartbeat can produce a variety of ventricular arrhythmias, which can also be fatal right away or cause irreversible heart problems across years. The capacity to detect arrhythmias via ECG data autonomously is critical for early identification and treatments. Applying conventional 12 lead ECG signal measurements information, we suggested an Artificial Neural Network (ANN) driven cardiac arrhythmia illness diagnosing method in this research.

The pulse rate of the individual cardiac is represented by an electrocardiogram (ECG). It is made up of five signals: P, Q, R, S, and T. The P single wave relates to the stimulant of the atria (Atrial Contraction), whereas the T single wave refers to the depolarization of the ventricular (Ventricular relaxation). The stimulation of the ventricles is referred by the QRS complex (ventricular contraction). The QRS complex, ST segment, PR interval, RR interval, PR segment, and Intervals are more critical parts of an ECG signal for diagnosing many cardiac disorders, including arrhythmia. An arrhythmia is a change in the heartbeat's normal speed or rhythm.

©IJRASET: All Rights are Reserved

Computerized ECG categorization is being more widely used by cardiologists in healthcare identification and treatment. Pre-processing, identification of features, and categorization procedures are all used in conventional ECG signal classifying techniques. After removing numerous types of noise and artifacts, the signals are segmented to produce information vectors across time, maybe with the use of a function reduction algorithm method to minimize dimensionally.

The remainder of the article is laid out as follows. In presentations, we give complete explanations of our suggested deep model, including experiments and evaluating criteria, as well as analyze actual outcomes.

II. LITERATURE SURVEY

A superintendent wearing system method is presented enabling longer high blood pressure rates tracking apps. The Electrocardiogram and PPG detectors are almost all incorporated within a single band to improve wearability. A non-standard separate design is given enabling the arm-ECG collection, which is much more accessible and pleasant than traditional wireless ECG, also includes placing the sensors on the heart (ECG and PPG) or 2 different wrists + 1 fingers (ECG) (PPG). This recently created device sample effectively obtains a poor arm-ECG signal having an intensity of just about 10% of such heart ECG signals. A supervised ml system was therefore designed to examine cardiac place & estimate heartbeats using this poor arm-ECG signal. The evaluation contains a mean absolute error (MAE) of 0.21 beats per minute (BPM) and a root mean square error (RMSE) of 1.20 BPM. After that, the pulse transit time (PTT) data is gathered & just utilized with calculate systolic blood pressure (SBP) using combined hand and arm-PPG digital waveform [2].

Using such a simple single point, we demonstrated a unique, compact, and small price wearable circle sensors device that can detect EDA, Heart Rate, activity, and temps. The fingers as just a collecting point for particular bio-data signals, the downsizing of the overall system, and the architecture of the planned platform that provides for unnoticed examination of physiologically parameters too are innovative aspects of our approach. The recommended program's experiment outputs were significantly associated that the reference modules. To the extent of their researchers' information, it was the first study to use a circle-based device to continuously record EDA and temp [3].

The planned technology can be employed in both the house and in the hospitals. To distinguish pre- and post-opioid medical problems in users, we used a learning algorithm and retrieved 23 characteristics using duration and amplitude domain investigation. To decrease the number of characteristics and processing time, attribute filtering approaches are applied. We analyzed 3 classifiers for supervised methods and chose the one with the best responsiveness and performance: decision tree, k-nearest neighbors (KNN), and Extreme Gradient Boosting with customized parameters. Findings suggest the offered approach could help identify opioid consumption in the real moment. Furthermore, this approach can be used to detect the usage of medicines apart from opioids. The statistical analysis has been done using data obtained over four weeks from 30 individuals [3].

A technique for ECG pulse categorization based on a portable representation is provided here. We specially developed a deep CNN with feedback links for such ECG categorization challenges and demonstrated that perhaps the information acquired for this job will be used as a foundation for training effective MI classifications. As per the findings, the proposed technique of predicting things along with both challenges with precision is similar to those found in the literature's state-of-the-art methodologies. We also used the t-SNE analysis to understand the learned form and demonstrate the efficiency of the suggested strategy. In this, we suggest a Deep Neural Network (DNN)-based framework that automatically classifies aberrant ECG waves and distinguishes them among standard pulses. DNN was created utilizing the Tensor Flow framework, which is a DL toolkit, as well as it includes just 7 hidden layers, each containing 5, 10, 30, 50, 30, 10, and 5 neurons. They used WEKA tools to automate a thorough evaluation of 11 different well-known classifications to illustrate the efficacy of the recommended DNN in respect of accurate categorization. The approach's performance has been demonstrated numerically, particularly in concerns of precision [7].

This paper provides actuation artefact identification and elimination technique for PPG data signals. The approach of choosing thresholds is used to identify and delete MA sections by collecting the increased confidence sections after evaluating the parameters of separated data signals. Experiments in minimizing the MA across various motion states show that this methodology is feasible. The interference of MA could be insulated while using the recovered increased sections instead of the entire data to determine SpO2 measurements, thus enhancing the precision of the SpO2 estimate [9].

Since approximately 440 bradycardia occurrences, we provide the results of a predictive method based on immediate linear measurements (mean AUC = 0.790.018). The system predicts bradycardia starts 116 seconds ahead of time on averages (FPR = 0.15). Increasing variation in the pulse rate signals is a predictor of serious bradycardia, according to our findings. Previous to bradycardia, such a rise in variability is linked to greater energy from minimal rates in the LF range (0.04-0.2 Hz) and decreasing multiscale entropy rates [10].

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

III. DATASET DETAILS

This resource is comprised of two groups of cardiac measurements generated from both the MIT-BIH Arrhythmia Set of information and the PTB Diagnosis ECG Data system, both well-known databases for cardiac plus classification. Those datasets have a sufficient quantity of instances to train a DNN. This collection of the dataset was used to investigate cardiac pulse categorization utilizing DNN framework, to test certain transferable learn skills. For such normal situations & instances impacted by diverse arrhythmias and myocardial infarction, these data signal waves correspond to ECG forms of pulse rate. The above data signals were subdivided and normalized, for each part representing a cardiac pulse.

In this dataset, we have the following five classes

1. Normal Beat

2. Supraventricular premature beat

3. Premature ventricular contraction

4. Fusion of ventricular and normal beat

5. Unclassifiable beat

A. Random Forest (RF)

RF is a popular ML technique that mixes the results of several decision trees to build a particular outcome. Its popularity is due to its simple usage and adaptability since it can solve both categorization and regression challenges.

The 3 major variables of RF methods must be established during training. The overall size of the nodes, the quantity of trees, as well as the number of characteristics collected is all factors to consider. The RF classifier could then be used to tackle challenges involving regression or classification.

The RF algorithm is composed of a group of decision trees, and every tree inside the ensembles is made up of a random subset, which is an information subset obtained from the trained examples with replacing. One of the training datasets is allocated as a testing dataset, referred to as the out-of-bag (OOB) sample, which we'll discuss later. Further examples of randomness were fed to the information via features bagged, improving the dataset's breadth while reducing dependencies among decisions tree. According upon their level of challenge, your forecasts would be confirmed in various methods.

Several decision trees will be aggregated in a prediction work and a major vote. most often classifiers produce the forecasting class in a categorization problem. Finally, the OOB dataset can be used for cross-validation, bringing the predictions to a conclusion.

B. Data preprocessing

All adjustments on the original information before it has been delivered to the ML or DL method are mentioned to as data pre-processing. Using actual information to teach a CNN, for example, would virtually likely result to poor categorization outcomes.

MIT Outcome = {0: 'Normal Beat',

1, 'Supraventricular premature beat',

2: 'Premature ventricular contraction',

3: 'Fusion of ventricular and normal beat',

4: 'Unclassifiable beat'}

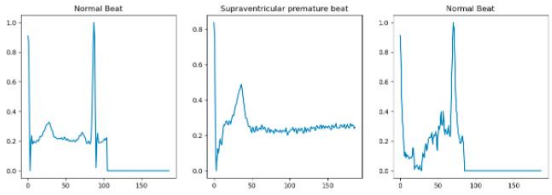


Fig -1: ECG Data Visualization

Displayed in Fig. 1 ECG the graphically display of data and information is known as data visualization. The representation of information in graphical representations to aid in the identification of features and linkages in your information is known as data visualizing.

©IJRASET: All Rights are Reserved

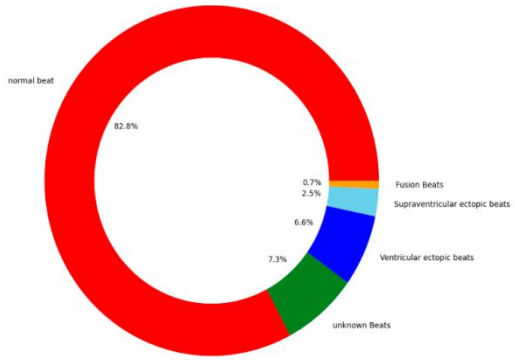


Fig -2: ECG data before balanced

Imbalanced datasets one in that some category has greater information than someone for the goal variables. Let's pretend we have such a collection that can be utilized to identify a fraudulent action in Fig -2: depicted.

1. **Normal Beat - 82.8%**
2. **Supraventricular premature beat - 2.5%**
3. **Premature ventricular contraction - 6.6%**
4. **Fusion of ventricular and normal beat - 0.7%**
5. **Unclassifiable beat – 7.3%**

Every outcome category (or goal class) is provided by some other corresponding inputs in a balancing dataset. Oversampling, under sampling, and class weighting are all strategies that can be used to achieve balancing. As shown in

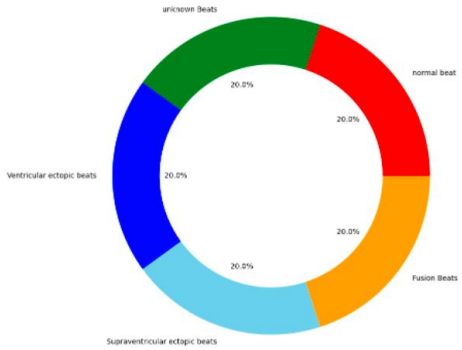


Fig -3: ECG data after balanced

1. **Normal Beat – 20.0%**
2. **Supraventricular premature beat - 20.0%**
3. **Premature ventricular contraction - 20.0%**
4. **Fusion of ventricular and normal beat - 20.0%**
5. **Unclassifiable beat – 20.0%**

IV. RESULT AND ANALYSIS

A. CNN

A CNN is a form of ANN used for picture and number identification and detection in DL. To use a CNN, DL recognizes items in a picture.

Deeply is a business phrase that is used to make anything sound more respectable than it used to. There are several different forms of DNN, including CNN. CNNs are useful in that networks can be used for picture identification as well as statistical datasets.

This CNN Classifier analysis is being utilized to assess the accuracy of the CNN categorization model prediction. That number all guesses were correct, exactly numerous of those were incorrect. True Positives, False Positives, True Negatives, and False Negatives were utilized & calculate classification analytic metrics, such seen here.

	precision	recall	f1-score	support
0	0.99	0.89	0.94	18118
1	0.26	0.85	0.40	556
2	0.86	0.92	0.89	1448
3	0.28	0.90	0.43	162
4	0.92	0.98	0.95	1608
accuracy			0.90	21892
macro avg	0.66	0.91	0.72	21892
weighted avg	0.95	0.90	0.92	21892

Fig -4: CNN classification report

Accuracy would be the number which calculates whether a program would perform across all categories. It comes in handy if each of your classes are equally essential. as seen in fig. -4 It really is calculated using the various in the number of correct estimates and the total number of predictions.

Depending on the confusion matrix designed, CNN shows how to estimate efficiency utilizing Scikit-learn. The outcome of dividing the total of True Positives and True Negatives so over the total of all values in the matrix is stored in the variables acc.

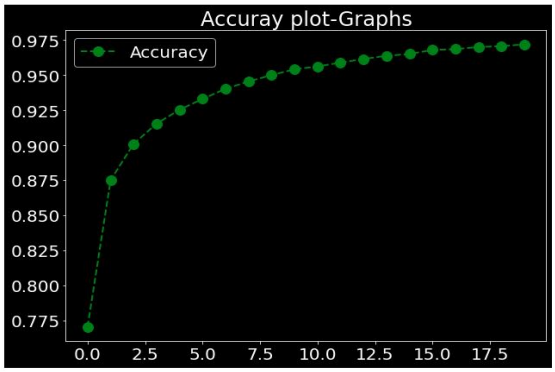


Fig -5: CNN Accuracy Graph

The CNN outcome in Fig. -5 is Training Accuracy of 89 percent and Test Accuracy of 90 percent, indicating that the system is 90 percent accurate in producing accurate predictions.

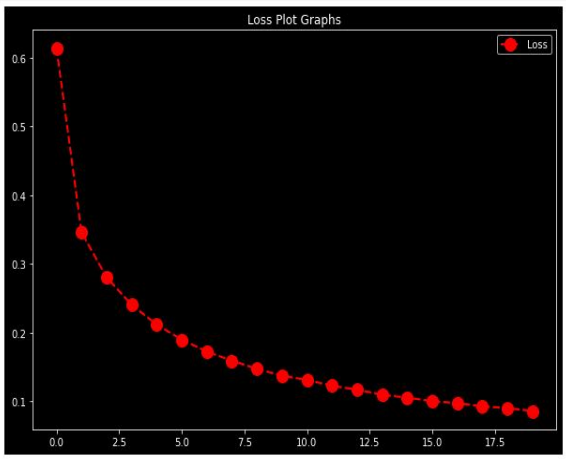


Fig -6: CNN Loss Graph

Figure -6 is displayed. Losses are nothing more than a Neural Net prediction error. The loss Factor is the way of calculating the loss. To say in another perspective, these rates were determined based on the Loss. Differences are indeed implemented to improve the values of the Artificial Network. Its way an Artificial Network can be trained.

B. Random Forest

The RF Classifications summaries are performed to evaluate the accuracy of the results of RF classification. How percentages of your predictions are correct, how percentage are incorrect? The variables of a categorizing analysis are evaluated using True Positives, False Positives, True Negatives, and False Negatives, see here.

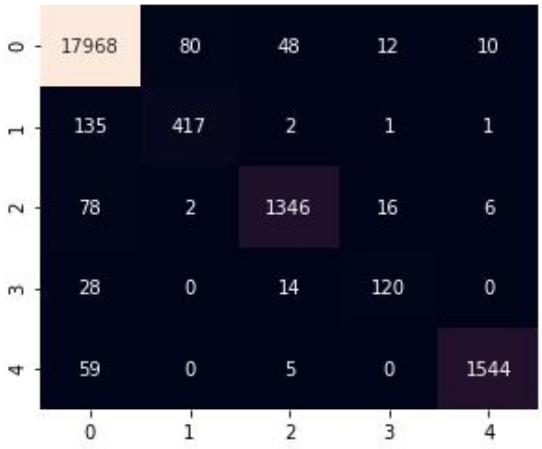


Fig -7: Confusion Matrix for RF

The RF outcome in Fig. -7 is Training Accuracy of 97.7% and Test Accuracy of 98.7%, which shows that the model is 98 percent accurate in producing accurate predictions.

V. CONCLUSION

The usage of the ECG database is a great way for classifying heartbeats. The proposed method used two methods such as RF and CNN for detecting and classifying the heartbeats based on the five classes applied to the two ECG datasets. These ML-based models allow for reliable, accurate, but most significantly for quickly classifying heartbeats from the ECG set of data. Balanced, concise input data additional increases the accuracy and prevents general ML problems, especially the issue of over-fitting. Furthermore, data preprocessing is done for the ECG information to be correctly formatted and balanced for the data information loaded. The comparison was done for RF and CNN, where CNN gave 92% accuracy and RF gave almost 98% accuracy based on the classification reports obtained for both. This implies that RF is the best fit for classifying the ECG data-based Cardiac Arrhythmia.

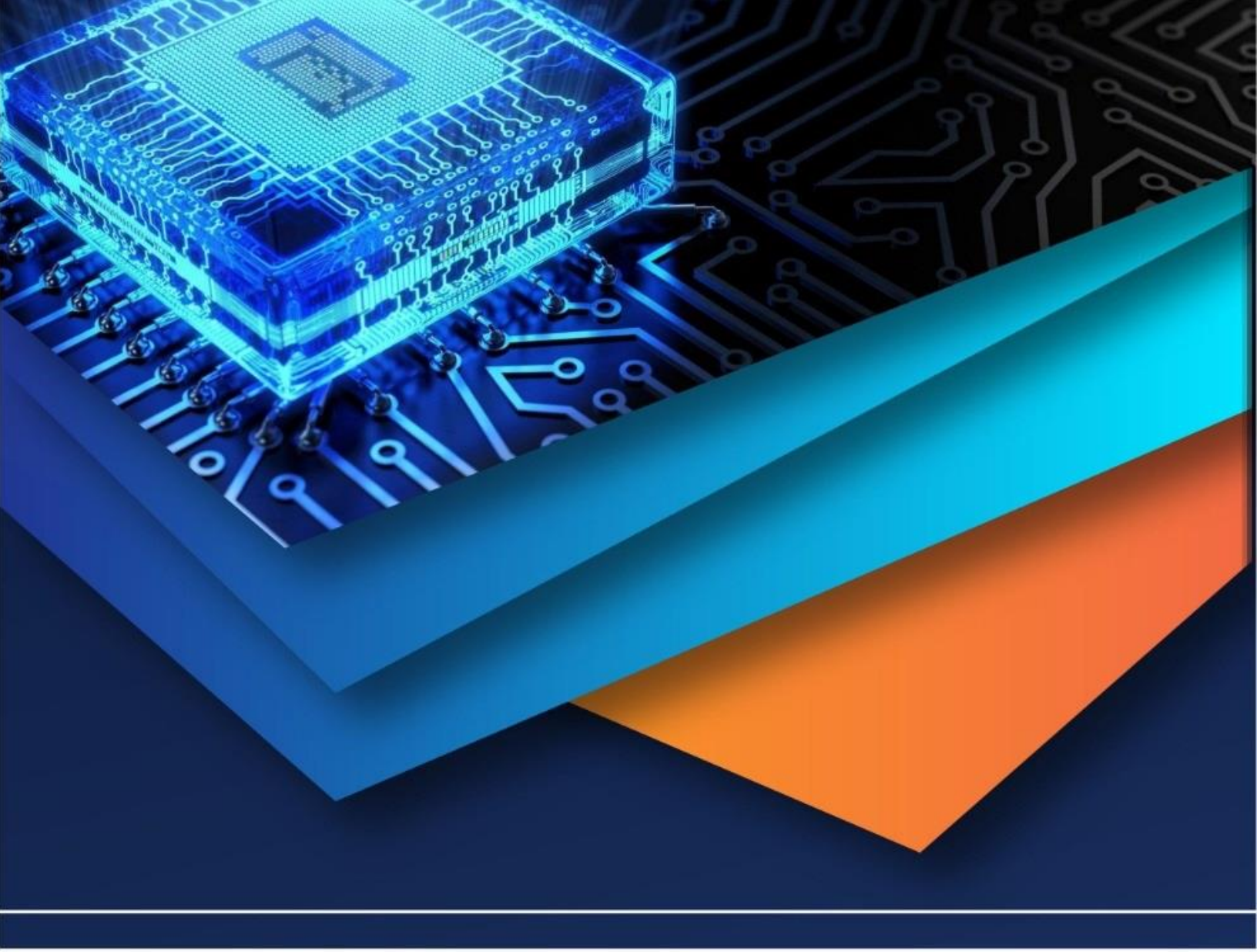
REFERENCES

[1] Lennox, C. and Mahmud, M.S., 2020, July. Robust classification of cardiac arrhythmia using a deep neural network. In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 288-291). IEEE.
[2] Zhang, Q., Zhou, D. and Zeng, X., 2017. Highly wearable cuff-less blood pressure and heart rate monitoring with single-arm electrocardiogram and photoplethysmogram signals. Biomedical engineering online, 16(1), pp.1-20.
[3] Mahmud, M.S., Wang, H. and Fang, H., 2018, May. SensoRing: An integrated wearable system for continuous measurement of physiological biomarkers. In 2018 IEEE International Conference on Communications (ICC) (pp. 1-7). IEEE.
[4] Mahmud, M.S., Fang, H., Wang, H., Carreiro, S. and Boyer, E., 2018, March. Automatic detection of opioid intake using wearable biosensor. In 2018 International Conference on Computing, Networking and Communications (ICNC) (pp. 784-788). IEEE.
[5] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M. and Ghemawat, S., 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467.
[6] Kachuee, M., Fazeli, S. and Sarrafzadeh, M., 2018, June. Ecg heartbeat classification: A deep transferable representation. In 2018 IEEE international conference on healthcare informatics (ICHI) (pp. 443-444). IEEE.
[7] Sannino, G. and De Pietro, G., 2018. A deep learning approach for ECG-based heartbeat classification for arrhythmia detection. Future Generation Computer Systems, 86, pp.446-455.
[8] Fandango, A., 2018. Mastering TensorFlow 1. x: Advanced machine learning and deep learning concepts using TensorFlow 1. x and Keras. Packt Publishing Ltd.
[9] Hanyu, S. and Xiaohui, C., 2017, May. Motion artifact detection and reduction in PPG signals based on statistics analysis. In 2017 29th Chinese control and decision conference (CCDC) (pp. 3114-3119). IEEE.
[10] Gee, A.H., Barbieri, R., Paydarfar, D. and Indic, P., 2016. Predicting bradycardia in preterm infants using point process analysis of heart rate. IEEE Transactions on Biomedical Engineering, 64(9), pp.2300-2308.

[11] Verma, A., Cabrera, S., Mayorga, A. and Nazeran, H., 2013, May. A robust algorithm for derivation of heart rate variability spectra from ECG and PPG signals. In 2013 29th Southern Biomedical Engineering Conference (pp. 35-36). IEEE.

[12] Zafeiris, D., Rutella, S. and Ball, G.R., 2018. An artificial neural network integrated pipeline for biomarker discovery using Alzheimer's disease as a case study. Computational and structural biotechnology journal, 16, pp.77-87.

[13] Parvaneh, S., Rubin, J., Babaeizadeh, S. and Xu-Wilson, M., 2019. Cardiac arrhythmia detection using deep learning: A review. Journal of electrocardiology, 57, pp.S70-S74.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)