ANNet: A Lightweight Neural Network for ECG Error Detection in Edge Sensors

A. Rajani[1] [1]ASST PROFESSOR
ECE, Department, UCEK, JNTUK, Kakinada, Andhra Pradesh, India.
[1]rajani.alugonda@gmail.com

P.V. Hari Krishna Babu[2] [2]M. Tech, JNTUK, Kakinada, Andhra Pradesh, India
[2] hari.ece290@gmail.com

Abstract— **With the usage of Internet of Things (IoT) edge sensors, we're going to implement a less massive neural network to detect Electro-Cardiogram (ECG) anomaly. The network is an aggregate of recurrent neural networks, including both Long Short Term Memory (LSTM) cells and Multi-Layer Perceptron (MLP) cells. The MLP layer receives the characteristics produced from instants of heart rate and the LSTM is fed with a series of coefficients of the denoised signal which are denoised using moving average filter constitutes the characteristics of ECG beat. By simultaneously training the blocks, the entire network is driven to learn unique characteristics that complement one another for decision-making. Accuracy and computational complexity of the network were measured using data from the MIT-BIH arrhythmia database. With the SMOTE network training method, we were able to expand the dataset and correct the class imbalance that had plagued it. The network achieved an overall 98% accuracy in classifying data across many entries in the database. To remedy the quantization problem, we retrained the network in a bit-accurate fixed-point environment, moved it to an embedded architecture based on the ARM Cortex M4, and then retrained it. Wearable devices in the Internet of Things may benefit greatly from the proposed solution since it outperforms previous methods in terms of processing complexity and can run independently at the edge node without the need for continual wireless connectivity.**

Key words-- **Anomaly detection, edge computing, IoT sensors, LSTM, MLP, neural networks, Moving Average Filter.**

I. INTRODUCTION

This will add sampling bias, overfitting problems, and other problems to the reported works in real-world circumstances. While most IoT devices operate under a fixed-point paradigm due to its simplicity and cost-effectiveness, few works have managed to properly create such an ecosystem. Existing works don't do a good job of addressing the quantization issues and performance reduction that can occur when converting floating point algorithms. Our study addresses all of the aforementioned research gaps as well as the problem of previous techniques' performance degrading in the presence of unknown real-world situations. The objective of this research is to tackle these issues through the creation of a machine learning technique for the binary classification of ECG signals, which can be readily integrated into an Internet of Things (IoT) sensor at a local level. Wireless communication won't be available until the classifier determines that an ECG beat is abnormal, which will limit the amount of power used by the sensors. To further tackle the problem of class irregularity inside the MIT-BIH Arrhythmia information base, we have embraced the Manufactured Minority Oversampling Technique (SMOTE)to improve the preparation information.This lessens differences between the suggested technique's real-world performance and the test data.

The proposed novel plan consolidates a repetitive block in view of Long Short Term Memory (LSTM)to distinguish the examples in normal time-series information, close by a clear Multi-Layer Perceptron (MLP) block that gets a handle on the intrinsic associations among the separated highlights. These elements incorporate initiation maps got from Principal Component Analysis (PCA) coefficients of a grouping of beats and ventricular rates. Our inventive methodology includes the simultaneous preparation of this multitude of blocks, convincing the whole design to catch different parts of the succession and collaborate in decision-making, much akin to the principles of ensemble learning. Additionally, we have developed both floating-point what's more, fixed-point

variants of pivotal AI parts to work with model improvement in a drifting point climate and its ensuing variation to a fixed-point execution. We also covered rapid approximations of functions and derivatives of such functions. The implementation loss that arises from pruning, quantization, and look-up table-based approximation losses is absent from our solution, making it distinct from the prevalent Tensor Flow approach. In this research, the suggested network demonstrates significantly reduced complexity and parameter count compared to current approaches, all the while achieving top-tier performance. The fixed-point model is surveyed and executed through the usage of an ARM Cortex M4-based Bluetooth inserted advancement unit, specifically the Nordic Semiconductor nRF52DK. By comparison to sending each sample via a Bluetooth connection, this method significantly reduces battery consumption by 50%. The proposed technique stands out as a strong candidate for deployment in IoT edge applications due to its classifier's minimal complexity, reduced sensor power consumption at the system level, and the reliability of real-world performance estimates facilitated by data augmentation.

In traditional approaches to processing electrocardiogram (ECG) signals, features are meticulously extracted using conventional methods, which encompass signal processing and feature extraction techniques like frequency domain analysis, wavelet transform (WT), and morphological features. Arrhythmia detection relies on the application of manually crafted algorithms and criteria to these extracted characteristics. A range of machine learning-based methods, including Decision Trees, Random Forests, K-Nearest Neighbours, Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), Reservoir Computing with Logistic Regression (RC), Linear Discriminants (LD), Hidden Markov Models (HMMs), Hyper Box Classifiers (HBCs), Optimal Path Forests (OPFs), Conditional Random Fields (CRFs), and Rules-Based Models, are employed in this endeavor.
The term "multi-layer perception" (or "MLP" for short) refers to this method. It is built up of thick, entirely linked layers that may convert any input dimension into the required dimension. The phrase "multi-layer perception" is employed to characterize a neural network with numerous levels of processing. Neural networks are built by connecting neurons in a manner where the results of particular neurons function as inputs for others. A multi-layer perceptron comprises an input layer equipped with a neuron (or node) for each input, an output layer featuring one node for each output, and an arbitrary count of hidden layers, each of which can potentially encompass varying quantities of neurons. An MLP, or Multi-Layer Perceptron, is shown in the simplified form below.

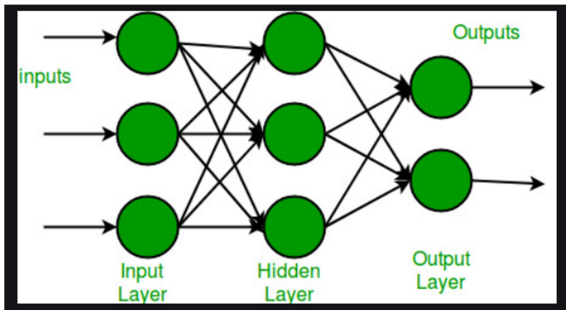

Fig 1: A Multi-Layer Perceptron (MLP) Network

The Long Short-Term Memory (LSTM) is a unique type of recurrent neural network capable of acquiring input dependencies over an extended duration. The four-tiered, interdependent module that serves as the model's recurrent component makes this feasible.

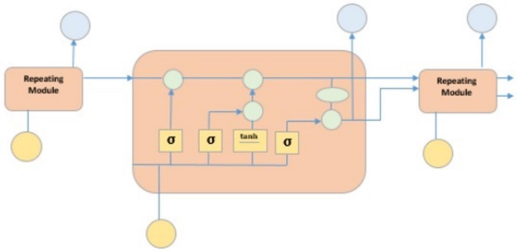

Fig 2: Long Short Term Memory (LSTM) Network

The illustration above displays four brain network layers, portrayed as yellow boxes. Pointwise administrators are represented by green circles, while input is addressed by yellow circles, and the cell state is outlined as blue circles. They can learn from certain units while ignoring others, or keep some while discarding others, all thanks to a cell state, three gates, and a long short-term memory (LSTM) module. The cell state in LSTM facilitates the seamless transfer of data between units by limiting the number of possible linear interactions to a minimum. Each part may add or delete information from the cell state through its own forget gate, input, and output. The process of determining which information from the preceding cell state should be disregarded involves the application of a sigmoid function by the forget gate.

II. RELATED WORKS

In the realm of automatically identifying and categorizing cardiac arrhythmias using ECG data, various approaches have been put forward in scholarly publications. The classification of arrhythmias is a pattern recognition job that can be carried out utilizing machine learning or syntactic techniques [17]. Features of an electrocardiogram (ECG) signal are painstakingly extracted using conventional syntactic approaches, which include signal handling and element extraction strategies, for example, recurrence area examination, wavelet change (WT), and morphological highlights. Arrhythmia detection is achieved by applying hand-crafted algorithms and criteria to the derived characteristics. Bayesian models use feature vectors that incorporate signal morphologies and characteristics to classify ECG signals [18]. Additional techniques encompass artificial neural networks (ANN), reservoir computing combined with logistic regression (RC), linear discriminants (LD), hidden Markov models (HMM), decision trees, random forests, K-Nearest Neighbors, support vector machines (SVM), optimum-path forests, conditional random fields, rule-based models, and various others.The training data's type and the learning approach chosen, however, have a significant impact on these methods' accuracy, and the data is frequently constrained by the wide range of patient morphologies.

Veeravalli and colleagues [17] conducted research in which they applied Fast Dynamic Time Warping (FDTW) with a limitation window to lay out an expense highlight grid for the underlying 30 beats in a patient's record. In this way, K-implies bunching was used to distinguish the overwhelming group, assigning a beat as the worldwide typical beat for that patient. Afterward, DTW distances were calculated for each incoming beat in relation to the chosen global normal beat. Additionally, a Hampel filter was employed to detect anomalies within the data. It should be noted that if certain classifications (such as Normal and Abnormal) were absent during the initial clustering phase, the strategy could encounter difficulties, as the most frequently occurring beats might not necessarily represent the clinical normal beat. In this study, only 15 records from the MIT-BIH arrhythmia information base were chosen for tending to the NP-difficult issue and assessing performance.

In the preprocessing approach proposed by Zadeh and colleagues [13], a bandpass filter was utilized, and a support vector machine (SVM) classifier based on characteristics extracted from a Continuous Wavelet Transform was employed. From a small collection of 8 carefully chosen patient records (118, 124, 207, 208, 209, 214, 222, and 223), the method has obtained 97% of Normal (N) vs. Abnormal (S, V, F, and Q) test accuracy across 17,784 beats. A similar method was used by Jiang et al. [14] to identify aberrant beats with 95.6% accuracy using a block-based neural network and Hermite transform characteristics over 49,600 chosen beats. However, the criteria for selecting beats in this research were not explicitly articulated.

Dan Li and colleagues [15] utilized a balanced downsampling approach when classifying ECG signals with a 1D Convolutional Neural Network (CNN). They used this method to create a dataset with an equal distribution of AAMI classes, and their results demonstrated a test accuracy exceeding 98% on a selected set of 13,200 beats.Preprocessing the ECG data using wavelet decomposition and feeding them into a neural network equipped with a SoftMax classifier.

III. METHODOLOGY

In this study, we make use of the MIT-BIH arrhythmia database [21] to assess the performance of the algorithms. This database comprises 48 ECG recordings, but does not include the paced1 beat records. These next 25 recordings showcase a variety of complicated ventricular, junctional, and supraventricular arrhythmias, while the next 23 are supposed to be a representative cross-section of regular clinical recordings. Bandpass filtering at frequencies between 0.1 and 100 Hz was used to record at 360 Hz. The moving average filter stage will then be used to further filter the bandpass filtered data. There are 15 different types of heartbeats, totaling approximately 100,000 labelled beats. There are two ECG leads on each file. The V1 or V2 or V4 or V5 modified leads are alternatives for the second wire. The initial lead accessible is the altered limb lead II, often referred to as ML II. at least two cardiologists Each 30-minute tape that was chosen from the 24-hour recordings was individually annotated [6], [14]. The database lists 15 different beat kinds.

For the purposes of this study, the rhythms labelled as "Abnormal" are those defined as "Supra Ventricular Ectopic Beats" (SVEB - S), "Ventricular Ectopic Beats" (VEB - V), "Fusion Beat" (F), and "Unclassified Beat" (Q), whereas the other rhythms are labelled as "Normal." All AAMI guidelines have been followed in creating this category. The original ECG data samples are used to generate two feature vectors:

X: Input to the LSTM_X Layer, and,
RR: Input to the MLP_R Layer

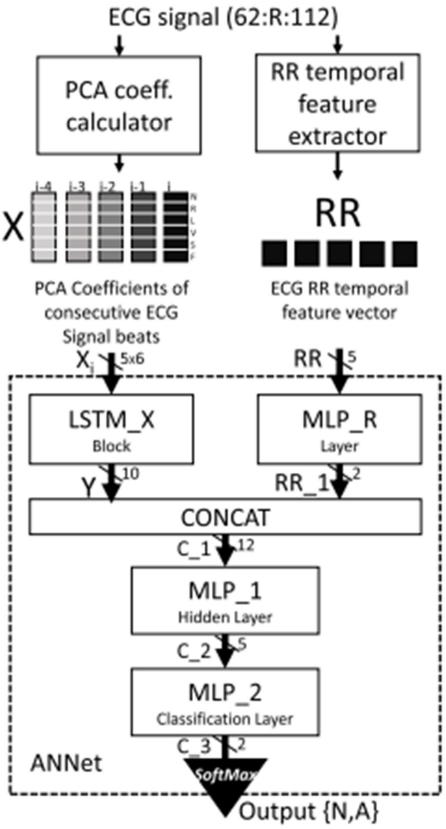

Fig 3: Block Diagram of our Proposed Method

Feature Vectors:

X: With respect to the principal Normal=N, RBBB=R, LBBB=L, Ventricular=V, Supra-Ventricular=S, and Fusion=F beats, Next, it moves on to compute Xi, where 'i' denotes the current beat index, and Xi represents a 6-component vector containing Principal Component Analysis (PCA) coefficients for every beat. For two unrelated

beats, Fig. demonstrates the process of creating a PCA feature vector. The 5 beat window should be used in the same way.

RR: The second feature vector is a 5-element array of RR values: [RRi, RRi+1, RRi, RRwSDNNi, RRIndexi]. The RR intervals from an electrocardiogram serve as the basis for this formula. The underlying two components inside this vector compare to the RR spans that happen preceding and after the ongoing ECG beat. Containing the third part is the mean worth of 11 RR-stretches traversing from RRi=9 to RRi=+1. Heart Rate Variability (HRV) metrics2, such as RRwSDNN and RRIndex, are based on [26 and are defined in Fig.

$$RR_{wSDNN_i} = \sqrt{\frac{1}{10}\sum_{j=-9}^{+1} W_j (RR_{i+j} - \overline{RR}_i)^2}$$

$$\text{where:} W_j = \begin{cases} 10 & j = 0 \\ 1 & \text{otherwise} \end{cases}$$

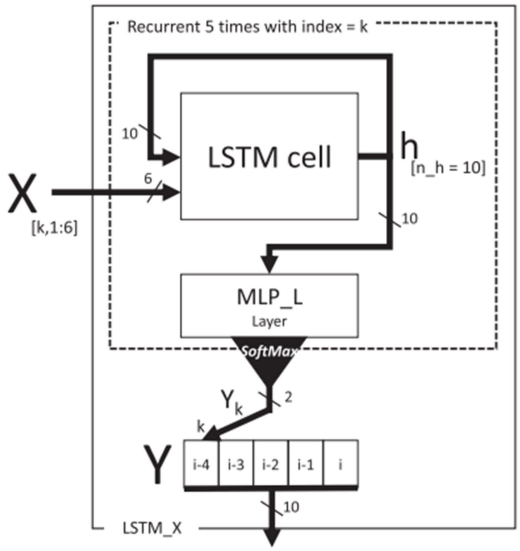

Fig 4: LSTM cell pipeline as part of LSTM_X

While a LSTM-based intermittent block is decided to survey the consistency qualities of normal time-series flags, the excess crucial separated highlights are handled through MLP layers to better understand their inherent relationships and predict anomalous patterns.Complementary learning, as opposed to ensemble learning of models, is the result of simultaneously training the blocks. This means that the blocks will generally learn distinct characteristics of the sequence and will assist one another when making a choice. In Figure 3, the LSTM_X module, akin to the approach by Zhang et al. [27], generates an attention map at its output denoted as Y, which retains information about the characteristics of prior beats. A more detailed representation of the LSTM_X module is illustrated in Figure 3. For every ECG beat, the LSTM cell is run five times with the inputs Xi4 through Xi coming one after the other. At each cycle, the LSTM cell updates both its output h (of length 10) and its internal state vectors. Additionally, the MLP_L layer receives each output vector h, which creates a vector (Yk) of length two as illustrated in Fig. 3. These two outputs are combined over the course of five execution cycles to create the vector Y, which has a length of 10.

The RR feature vector is sent into the MLP_R layer (Fig. 2), which operates in parallel with the LSTM_X block and produces an RR_1 of length 2. The result is formed by combining the vector Y generated by the LSTM_X block with the output (C_3) produced by an MLP network consisting of MLP_1 and MLP_2. This MLP network has a sole hidden layer comprising five neurons. Following this, the two outputs are blended together by passing through a SoftMax layer, which assigns the beat a classification of either "Normal (N)" or "Abnormal (A)."

(a) Utilizing the Sigmoid Activation Function (σ):

We utilized the Sigmoid cellular activation function to adjust the neuron weights in ANNet. The following formulas illustrate the conventional sigmoid (3) and its derivative (4).

$$\sigma(x) = \frac{\exp(x)}{\exp(x) + 1}$$

$$\frac{\partial}{\partial x}\sigma(x) = \sigma(x)(1 - \sigma(x))$$

(a) Tanh Activation Function: For the LSTM cell, it performs essentially as the activation function of the candidate gate Ct (Fig. 3). Anguita et al. [29] inspired a fast approximation fixed point implementation of the original (9) and its derivative (10), which we show here. The original formula (9) is used here, but its coefficients have been tweaked for efficiency to accommodate for bit-level manipulation.

$$\tanh(x) = \frac{(\exp(2x) - 1)}{(\exp(2x) + 1)}$$

(b) SoftMax Function: The classification layers are where this is typically employed. Instead of being proportional to the output of a single neuron, it is linked to the output of the last fully connected layer, from which a normalised vector, s (15), is calculated:

$$s_i(\boldsymbol{x}) = \frac{\exp(x_i)}{\sum_k \exp(x_k)} \forall 0 \leq i < |\boldsymbol{x}|$$

Despite all earlier efforts using greater sampling rates (360 Hz) and floating point implementations, the suggested method outperformed them all while keeping lower complexity. The training and testing procedures in [13], [17], and [15] claim performance that is on par with or slightly (1%) better than ours; however, these methods do not follow AAMI guidelines [12] and comprise about half as many test beats as we do. However, it employs LSTM cells inconsistently and depends on two lead ECG data, which is difficult to get in a wearable device, making it similar to [11], which delivers somewhat greater accuracy. Both [6] and [16] use a CNN-based architecture and depend entirely on the nearby morphology of a solitary beat ECG section. Notwithstanding not having the option to remove and use the RR span data, these approaches are not recommended to be used because of the wide variation in patient morphologies, which might lead to poor performance in an untested setting. The mediocre performance of [11], [16], [35], and [6] is achieved at the cost of over a million instruction cycles each beat classified, and the 1D-CNN architecture of [35] is seldom employed for the extraction of temporal and morphological data. Most of the above methods utilise the imbalanced MIT-BIH dataset for assessment purposes.

IV. RESULTS AND DISCUSSIONS

The results of ANNet's experimental runs are shown in the figures below:

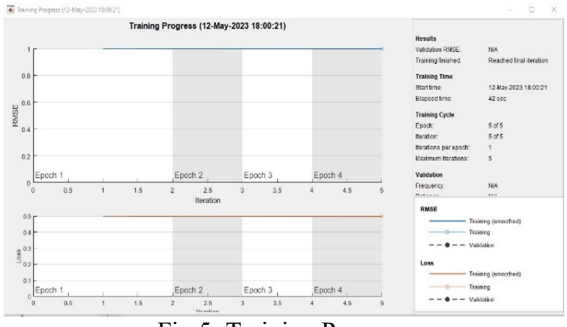

Fig 5: Training Progress

The above figure is considered as the training progress of the constructed ANNet architecture where it's been trained on the both types of signals i.e., normal as well as abnormal. The training progress will take some time depending on the amount of data that we've fed to the classifier. So, it is better to choose the features in less amount that

constitutes the exact resemblance of the type of the signal easily is a challenge here, but, we've used PCA for feature extraction which extracts better and less features that almost resembles the type of signal easily.

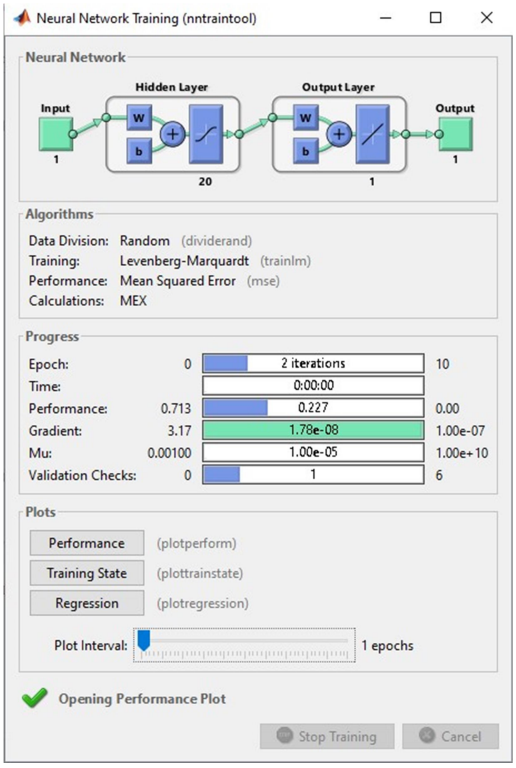

Fig 6: Training

The time, date, number of epochs, number of iterations, validation information, and more may all be tracked here as the training proceeds.

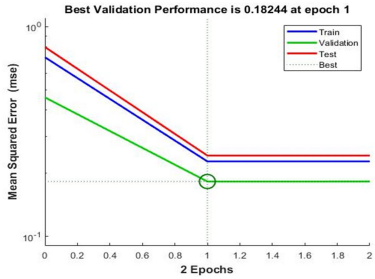

Fig 7: Training Validation

In this scenario, the input signal could be just OK. Therefore, it fits the criteria for a normal. In the event of anything becoming odd, it will show up as abnormal and halt processing. Validation accuracy is more than 98% as shown in below figure.

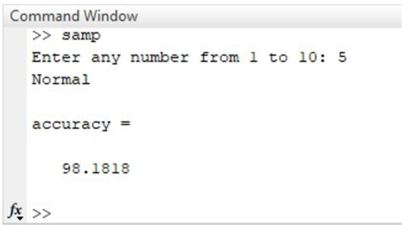

Fig 8: Classification Output

S.NO	Parameters	ANNet method	Proposed Method
1	MSE	>1	<1
2	Average Classification Accuracy	97.72	98.18

Fig 9: Comparison of Classification accuracy

The classified output is displayed in the command window along with the accuracy of prediction. The accuracy of prediction and detection of signals became so better compared to existing results.

V. CONCLUSION

In this study, our approach involves the utilization of a basic neural network for the detection of irregular heart rhythms within ECG recordings. Our network is designed to receive input comprising two components: a feature vector that consists of PCA coefficients and a temporal feature vector derived from the ventricular R-R interval rate, along with five consecutive beats.For a greater reduction of noise from the input signal, we have employed the moving average filter in this case. The suggested method can achieve minimal complexity in normal clinical recordings while having better anomalous signal detection accuracy and acceptable accuracy in challenging records. After retraining, the approach was converted to an embedded platform with the least amount of implementation loss and the least amount of implementation cost by replacing many activation functions with approximations and mapping to fixed point. Compared to the state of the art, a computationally complicated design.

By using a binary classifier to gate the wireless transmission such that only irregular beats are broadcast, we demonstrated a significant reduction in the overall system power usage as compared to continuous data transfer.

VI. REFERENCES

[1] W. Jiang and S. G. Kong, “Block-based neural networks for personalized ECG signal classification,” IEEE Trans. Neural Netw., vol. 18, no. 6, pp. 1750–1761, Nov. 2007.

[2] J. Leal et al., “Economic burden of cardiovascular diseases in the enlarged European union,” Eur. Heart J., vol. 27, no. 13, pp. 1610–1619, Feb. 2006.

[3] D. L. T. Wong et al., “An integrated wearable wireless vital signs biosensor for continuous inpatient monitoring,” IEEE Sensors J., vol. 20, no. 1, pp. 448–462, Jan. 2020.

[4] Y. Wei et al., “A review of algorithm & hardware design for AI-based biomedical applications,” IEEE Trans. Biomed. Circuits Syst., vol. 14,no. 2, pp. 145–163, Apr. 2020.

[5] E. J. da et al., “ECG-based heartbeat classification for arrhythmia detection: A survey,” Comput. Methods Programs Biomed., vol. 127, pp. 144–164, 2016.

[6] W. A. Zoghbi et al., “Sustainable development goals and the future of cardiovascular health: A statement from the global cardiovascular disease taskforce,” J. Amer. Heart Assoc., vol. 3, no. 5, 2014, Art. no. e000504.