Heart arrhythmia detection on ECG time series data using RNN and CNN

Norhen Abdennadher[†], Achraf Sahnoun[‡]

Abstract—An electrocardiograph (ECG)being one of the most extensively used signals to monitor cardiovascular diseases (CVDs), captures the heart's arrhythmias [1]. The acquired ECG time series (TS) signal is often manually evaluated by qualified medical doctors in order to discover any arrhythmia that the patient may have experienced. This is especially inefficient when indicative patterns in the biological signals are infrequent, requiring more analysis, and entailing a difficult visual search task for the diagnosis. Much effort has been put into automatizing the process of evaluating ECG readings to learn the various forms of arrhythmia.

In this paper, we propose hybrid deep learning models based on recurrent neural networks (RNN) more precisely long short term memory (LSTM) cells, and convolutional neural network (CNN) which can diagnose irregular heart rhythms, also known as arrhythmias, from single-lead ECG signals. This approach is used to classify the heart's pathological conditions into two normal or abnormal heartbeat. Comparative research was conducted to identify differences between deep architectures in terms of performances and to compare between personalized learning and generalized learning. The detection accuracy has been evaluated on the publicly available MIT-BIH arrhythmia dataset, also being processed and modified to denoise data and prepare sequences for the learning and validation processes. It achieves 85% in accuracy and reaches out to 98% as maximum in the personalized learning.

ECG, Arrhythmia, time series data, recurrent network, anomaly detection, wavelet transform, convolutional neural network, LSTM, sliding window.

I. INTRODUCTION

Recently AI techniques have sent vast waves across healthcare, thus reshaping how medical services are deployed and becoming a reality in many medical specialties. AI automated systems enable healthcare stakeholders and medical professionals to identify needs and develop solutions faster with higher accuracy to make medical decisions efficiently [2]. The power of AI can be leveraged to augment physicians in their abilities to diagnose and treat patients and enhance the entire spectrum of the patient journey. Specifically, in cardiovascular medicine, AI-based systems have found new applications to predict more accurately cardiovascular risks at early stages, as they are a leading cause of human death worldwide [3], [4], [5]. However, to detect such early indicators, continuous heart monitoring needs to be performed over long periods of time. To date, this has been impractical for several reasons. First,

[†]ICT for internet and multimedia at the university of padova, email: norhen.abdennadher.1@studenti.unipd.it

[‡]ICT for internet and multimedia at the university of padova, email: achraf.sahnoun@studenti.unipd.it

long-term surveillance generates a large amount of data, most of which are non-pathological. Moreover, searching visually such anomalies for indicative segments is a tedious and error-prone task, which also must be performed by only qualified medical professionals. Therefore, an accurate arrhythmia detection model, to identify personal CVDs at their early stage, could effectively help to reduce the mortality rate by providing an instantly fast notification to provide better support. To automatically detect heart arrhythmia in an ECG stream, an algorithm must implicitly recognize the distinct wave types and discern the complex relationships between them over time. One of the major challenges to detecting cardiovascular diseases is that a normal heartbeat varies with age, body size, activity, and emotions [6]. Therefore, personalized learning is required.

In this paper, we leverage the benefits of a recurrent neural networks (citation) and convolutional neural network (citation) to create an automatic solution for detecting heart's arrhythmia from the ECG signal. This is achieved using different deep learning architectures and deep neural networks. Different models were implemented to leverage the power of RNN and CNN and perform a comparative study between them according to the achieved performance. To simplify the problem, we will assume that a QRS detector is capable of automatically identifying the peak of each heart beat. Our suggestions were analytically formulated and illustrated with examples and results from experiments on the publicly available MIT BIH dataset after being modified to mimic temporal features as it will be described in section II. Predict if a heart beat from the first ECG signal has an arrhythmia for each 6 second window centered on the peak of the heart beat.

II. MIT BIH DATASET

The MIT-BIH Database [7] was the first set of standard test material that was publicly available to evaluate any arrhythmia detectors and it has been used for that purpose as well as for basic research into cardiac dynamics. This database consists of 48 half-hour excerpts of two-channel ambulatory ECG recordings, obtained from 47 subjects studied by the BIH Arrhythmia Laboratory. Twenty-three recordings were chosen at random from a set of 4000 24-hour ambulatory ECG recordings collected from a mixed population. The remaining 24 recordings were selected from the same set to include less common but clinically significant arrhythmia that would not be well-represented in a small random sample. The recordings were digitized at 360 Hz for each channel with 11-bit resolution over a 10-mV range. For the labels of this dataset, two

or more cardiologists independently annotated each record; disagreements were resolved to obtain the computer-readable reference annotations for each beat included with the database. Ten seconds from record 205 of the MIT-BIH Arrhythmia Database are shown in figure 1.

Fig. 1: ECG signals in record 205 for both channels (above:MLII, below: V1)

A. dataset exploration

The source of the ECGs included in the MIT-BIH Arrhythmia Database is a set of over 4000 long-term Holter recordings that were obtained by the Beth Hospital Arrhythmia Laboratory. Approximately 60% of these recordings were obtained from inpatients. The database contains 23 records chosen at random from this set, and 25 records selected from the same set to include a variety of rare but clinically important phenomena that would not be well-represented by a small random sample of Holter recordings. Each of the 48 records is slightly over 30 minutes long. The first group is intended to serve as a representative sample of the variety of waveforms and artifact that an arrhythmia detector might encounter in routine clinical use. Records in the second group were chosen to include complex ventricular, junctional, and supra-ventricular arrhythmias and conduction abnormalities. Several of these records were selected because features of the rhythm, QRS morphology variation, or signal quality may be expected to present significant difficulty to arrhythmia detectors; these records have gained considerable notoriety among database users. The subjects were 25 men aged 32 to 89 years, and 22 women aged 23 to 89 years.
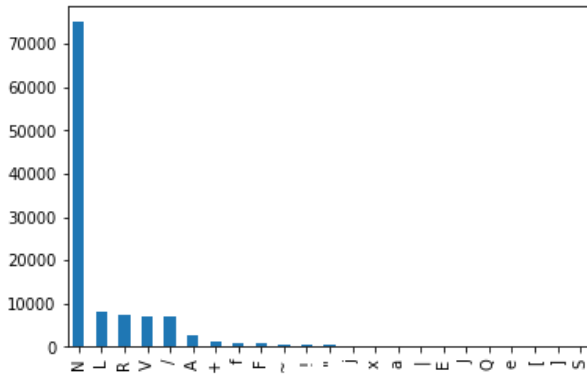

Fig. 2: Labels distribution in the MIT BIH dataset

Loading all annotation files and looking through the distribution of the heartbeat types across all files, we can see the number of occurrence of each label in the whole dataset as shown in figure 2. The distribution of the labels can be seen as well in each record separately, in order to compute the percentage of abnormality included in each record.

TABLE 1: Labels distribution on different MIT BIH records

Record	Labels count
100	+1 A33 N2239 V1
101	+1 A3 N1860 Q2 —4 4
200	+148 A30 F2 N1743 V826 43
201	+35 A30 F2 J1 N1625 V198 a97 j10 x37 4

A sequence from the MIT BIH dataset after merging all records together is displayed in figure 3 showing a 6 seconds sequence of normal and abnormal heartbeats.
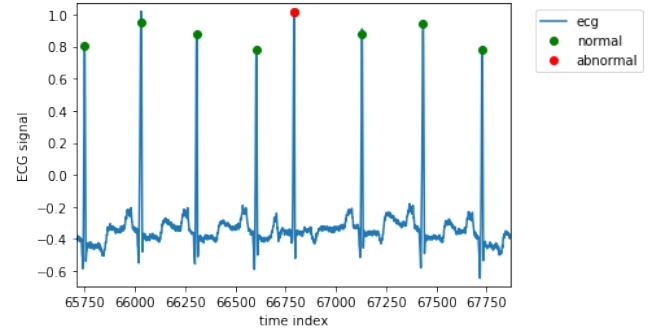

Fig. 3: Single channel, 6 seconds from the MIT BIH dataset

B. dataset preprocessing

ECG is susceptible to different types of noises, which might distort the morphological features and the interval aspects of the ECG, leading to a false diagnosis and improper treatment of patients [8]. For that matter, ECG signal denoising is a major preprocessing step which attenuates the noises and accentuates the typical waves in ECG signals. Much interest has been given to study and compare ECG denoising techniques [9] [10] [11]. In this study, Wavelet transform has been used to denoise records within the MIT BIH dataset [12]. This method has recently become very popular when it comes to analysis, denoising and compression of signals and images. A discrete wavelet transform (DWT) is a transform function that decomposes a given signal into a number of sets, where each set is a time series of coefficients describing the time evolution of the signal in the corresponding frequency band (citation). In practical cases, noise signals usually appear as high frequency signals in signal processing, but useful signals appear as either low frequency or more smooth signals. One dimensional signal denoising process is as follows: the one dimensional signal is decomposed by wavelet decomposition, selecting threshold and threshold function to quantify the high frequency coefficients of wavelet decomposition and reconstruct the one dimensional wavelet. Because wavelets localize features in your data to different scales, you can preserve

important signal or image features while removing noise. What this means is that the wavelet transform concentrates signal and image features in a few large-magnitude wavelet coefficients. Wavelet coefficients which are small in value are typically noise and you can "shrink" those coefficients or remove them without affecting the signal or image quality. After you threshold the coefficients, you reconstruct the data using the inverse wavelet transform.

In practical cases, noise signals usually appear as high frequency signals in signal processing, but useful signals appear as either low frequency or more smooth signals. One dimensional signal denoising process is as follows: the one dimensional signal is decomposed by wavelet decomposition, selecting threshold and threshold function to quantify the high frequency coefficients of wavelet decomposition and reconstruct the one dimensional wavelet. Because wavelets localize features in your data to different scales, you can preserve important signal or image features while removing noise. What this means is that the wavelet transform concentrates signal and image features in a few large-magnitude wavelet coefficients. Wavelet coefficients which are small in value are typically noise and you can "shrink" those coefficients or remove them without affecting the signal or image quality. After you threshold the coefficients, you reconstruct the data using the inverse wavelet transform.

After denoising the ECG signal, we move to the preparation of the dataset to the learning and validation steps. Looking at the annotation of the MIT BIH dataset, different symbols were used to annotate the beats, each leads to different interpretations. For the purpose of this experiment, we considered only two classes: normal and abnormal beats. Invalid beats were discarded from the labels set, so let's assume having a filter in the system able to discard invalid beats before the learning process. Thus, the dataset was relabelled in order to form a bi-class dataset composed of normal and anomalous heartbeats.

Since time series streaming-like data are being considered, the location of the anomaly cannot be accurately determined unless using a segmentation algorithm. Heartbeat's segmentation implies a priory knowledge in time of the beat's position in order to split the ECG recording into individual heartbeats. Generally, most of the foregoing methods exhibited good performance in QRS detection [13] [14]; however, in this project we assumed having an accurate QRS detector. Therefore, we modified the dataset and introduced a sliding window over the time-series ECG records, that generated multiple input shapes to be classified as normal or abnormal by the proposed model. These were used for both learning and evaluation processes. Two scenarios were studied: first, We applied a sliding window of size 300 samples, which corresponds approximately to the number of samples per heartbeat as shown in figure 18. Assuming a healthy heart and a typical rate of 70 to 75 beats per minute, each cardiac cycle, or heartbeat, takes about 0.8 seconds to complete the cycle [15]. The sampling frequency for the MIT BIH dataset is 360Hz, so that the number of samples per heartbeat is

equal to 360*0.8 samples. This explains the chosen size of the sliding window for the MIT BIH dataset. This fixed-size window will slide over the recordings with a stride equal to 300 samples, which is a kind of a segmentation technique to separate heartbeats and work on them individually.
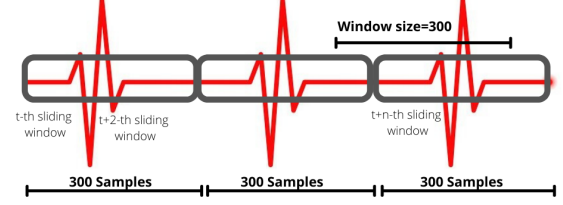

Fig. 4: Sliding window on ECG signal

In a second experiment we divided the dataset into sequences of 6 minutes each centered around a heartbeat with 3 seconds before and 3 seconds after, so we can compare the current beat to beats just before and after. This decision was based after talking to a physician who said it is easier to identify if you have something to compare it to (citation https://stanfordmlgroup.github.io/projects/ecg/). Thus, all MIT BIH dataset records were divided into a succession of 6 seconds sequences and were merged together into a single list while noting the start and the end of each patient record and the percentage of anomalies within it as shown in figure 5.

	subject	percentage	num_seq	start	end
0	100	0.015018	2264	0	2264
1	101	0.002691	1858	2264	4122
2	102	0.954566	2179	4122	6301
3	103	0.000962	2078	6301	8379
4	104	0.926610	2221	8379	10600
5	105	0.017941	2564	10600	13164
6	106	0.256309	2021	13164	15185

Fig. 5: Data preparation

As a last step, the resulted data ready-to-use will be splitted into training 80% and validation 20% sets both stratified based on the rate of abnormal beats, so that the proportion of values in the produced sets contains the same proportion of normality and abnormality.

III. PROPOSED ARCHITECTURES

There are three crucial steps in the whole process, i.e., signal pre-processing, extraction of features from the data, and its classification. The overall process is shown in Figure 6.
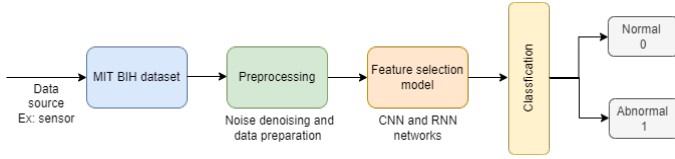
Fig. 6: Process of ECG signal classification

In this project, to do make our models learn, we have two different approaches:

1) First option is the generalized learning: Training the model on the whole MIT BIH dataset, and generalize the inference on every new subject not belonging to the 47 existing subjects in the dataset. This is the most common option nowadays since training a neural network is time and resource consuming.

2) Second option is the personalized learning: One of the major challenges for an ECG learning is the variety of the signal's morphology from one person to another depending on their gender, age, health status, type of medicines... A personalized learning can be advantageous in this scenario. The system should be occupied by low-power sensors able to save the ECG signal from the device user and process learning and inference at the same time [16]. In case of changing the user, restarting the device will be required to repeat the process from scratch.

In this paper, an efficient deep convolutional neural network (CNN) [17] and recurrent neural network (RNN) [18] based architectures were proposed to extract depthwise temporal features and consider time dependencies along with a robust end-to-end scheme to automatically detect and classify arrhythmia from denoised electrocardiogram (ECG) signal.

A. Parameters and tools selection

1) The number of epochs is the cycle of browsing forward and backward across the complete dataset, and each epoch has a set of iterations dependent on the batch size and dataset size. We won't be able to get the most out of the model if the number of epochs is too short, because only a few cycles will not disclose the true performance and optimize the training. However, because we are working with high-level models and enormous datasets, training on a large number of epochs takes a long time and can lead to overfitting.As a result, we should choose a reasonable number of epochs.

2) Batch size: The ability to train the model with less memory is a benefit of employing batches in a neural network.Additionally, because the weights are updated frequently, employing micro batches allows the model to be trained faster. We chose to set the batch size to 64 samples after some manual tuning.

3) Learning rate: The learning rate is a hyperparameter that controls the speed of learning and the stages of updating weight values when attempting to minimize the loss function, and it ranges from 0 to 1.After experimenting with various learning rates,We decided to choose 0.001 as the final value since we are using a stable optimizer (Adam).

4) The activation function [19]: also known as the transfer function, is an important component of the neural network that allows it to break linearity and direct the learning process. It specifies how the input weighted sum is translated into an output. For the hidden layers, we will use the rectified linear activation function (ReLu) With the convolutional layers and The hyperbolic tangent activation function with recurrent layers. According to the output layers we used the logistic function.

5) Loss function: Also known as the cost function , is minimized by the model during optimization. It converts the complex components and values of the model into a single real number that can be used to rank and compare results depending on the model and task. In this work the binary crossentropy is used which is shown in the following equation:

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i.log(p(y_i)) + (1 - y_i).log(1 - p(y_i)) \quad (1)$$

6) The optimizer: It is a crucial part of the training process. The optimizer's job is to update the weights based on the value of the loss function used, which in this case is the cross entropy. The optimization strategy will be guided by this, and the goal will be to produce the best accurate model possible. In this work Adaptive Moment Estimation also known as Adam optimizer is used.

7) Dropout: The dropout layer is a last regularization approach that we used. Dropout seeks to minimize the neural network's complexity and avoid overfitting. Based on a dropout rate that governs the fraction of dropped neurons, it works by deactivating a portion of the neurons and changing their values to 0.

8) Batch Normalization: Normalization is the process of converting data into a common balanced scale without losing the data's shape or information. It's done as a layer, and it adds two new weights to reflect the mean and deviation. Batch normalization is regarded as a regularization method. It can also make the network more stable and speed up the learning process

9) Early stopping: We applied early stopping as a technique to perform regularization and avoid overfitting. The approach is very simple: using the validation set, we can track the loss function values. At a certain point, the value shows no improvement on the validation set so we can stop the training and keep the best version of the model that corresponds to the lowest validation loss.

B. Deep Neural Networks Architectures for generalized learning

Once the pre-processed ECG beats are extracted, the next objective is to develop efficient deep neural network architectures to reach our goal and in our case CNN and a combination between CNN and RNN are experienced since they showed their efficiency for the arrhythmia classification task on the whole MIT-BIH dataset.

1) CNN + Dense: In fact The investigated CNN consists of two 1D Convolutional layers with 16 and 8 units

respectively, and a kernel size of 4x4. And to achieve regularization, Batch Normalization and dropout (40 percent) are applied after each convolutional layer.Next the Rectified Linear unit (ReLu) activation function, the flattened output feeds a fully connected layer, with 32 neurons followed by a dropout layer to reduce overfitting. The last layer consists of only one node, and produces the classification output.
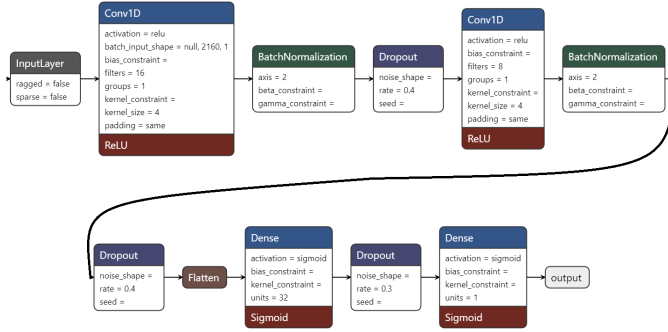

Fig. 7: CNN + Dense architecture

2) CNN + LSTM model: The investigated CNN+LSTM consists of two 1D Convolutional layers with 16 and 8 units respectively, and a kernel size of 4x4. And to achieve regularization, Batch Normalization and dropout (40 percent) are applied after each convolutional layer. Next the Rectified Linear unit (ReLu) activation function, the flattened output feeds a LSTM layer with 8 neurons followed by a dropout layer to reduce overfitting. The last layer consists of only one node, and produces the classification output.
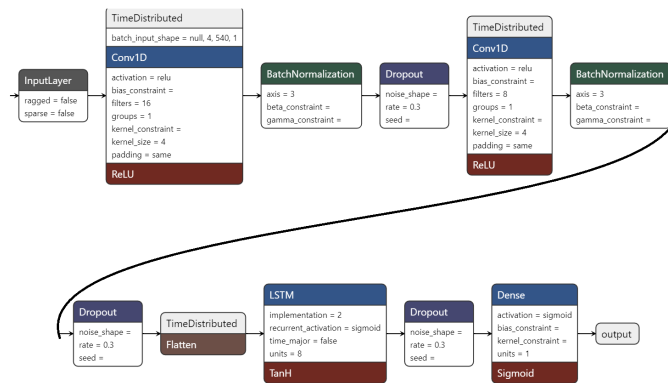
Fig. 8: CNN + LSTM architecture

C. *Deep Neural Networks Architectures for personalized learning*

In this approach we need to develop an efficient deep convolutional neural network and recurrent neural network architectures for arrhythmia classification on individual recordings separately.

1) CNN based architecture:

A single layer convolutional neural network composed of 8 filters and 2 kernels was used in a first step to extract features within the ECG signal. An illustrated design of the topology and input shape is shown in Figure 9.
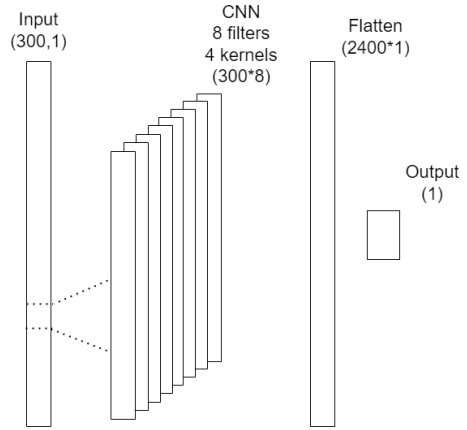

Fig. 9: CNN architecture

2) Merged CNN and RNN based model for a personalized learning: Another common architecture used for sequential models is CNN with LSTM. The idea is to have a CNN layer working as a feature extraction layer and an LSTM to interpret the sequence of subsequences as either normal or abnormal. Using recurrent neural network while dealing with time series data is recommended as this kind of network captures long time dependencies. An illustrated design of the topology and the input shape is shown in Figure 10.
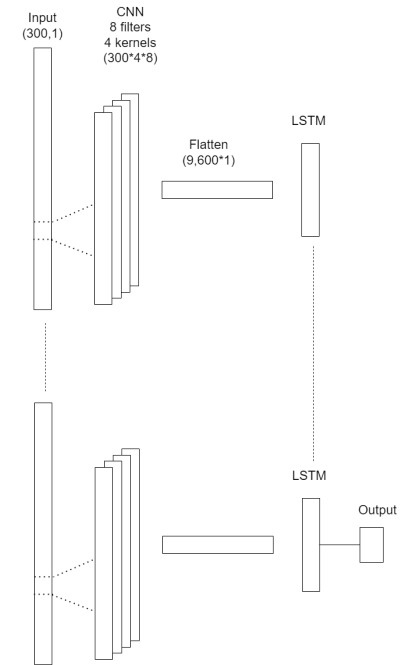

Fig. 10: CNN+LSTM architecture

The proposed CNN and RNN classifiers are implemented in one of the most essential and famous machine-learning languages, Python, with an open-source library Tensor Flow developed by Google for deep learning.

IV. RESULTS

In this section, for the purpose of demonstrating the performance of the proposed methods experimental results are presented along with performance comparison and detail analysis on the effects of various options on the performance. The performances of the different proposed networks are reported in terms of training and validation accuracy. Validation of data will determine any model which gains maximum accuracy with the training data. If we did not use the validation data technique, then the model will fall into the problem of overfitting without noticing.

1) Generalized learning results: The investigated CNN+Dense, and the CNN+LSTM topologies were trained to classify arrhythmia from ECG signal following the first generalised approach. Various type of layers were used as well as different regularization techniques. The models took a long time to finish the training and a small gap between training and validation results is noticed. The following figures represents the history of the accuracy and loss during the training and validation phases. Finally the table 2 contains the final accuracy of training and validation of the two models during the first approach. Later on a comparison will be presented between the two approaches.

TABLE 2: Results of the first approach

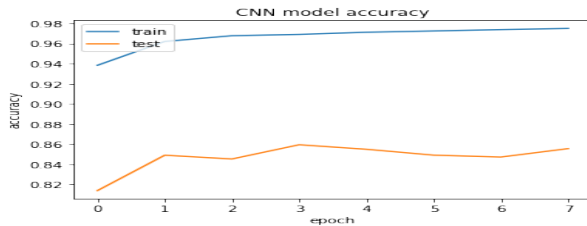| NN Architecture	Train accuracy	Val accuracy	Train loss	Val loss
CNN+Dense	97,12	85,94	0.074	0.51
CNN+LSTM	96,77	82.99	0.077	0.56

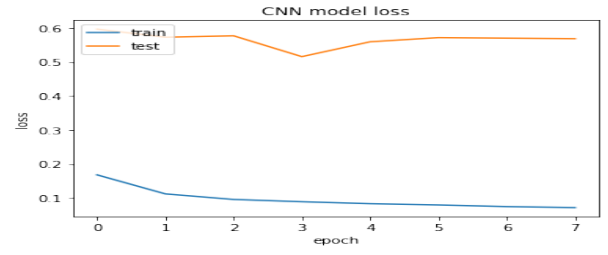Fig. 11: Training and validation accuracy using CNN+Dense architecture

Fig. 12: Training and validation loss using CNN+Dense architecture

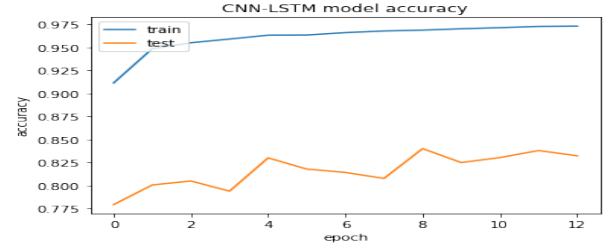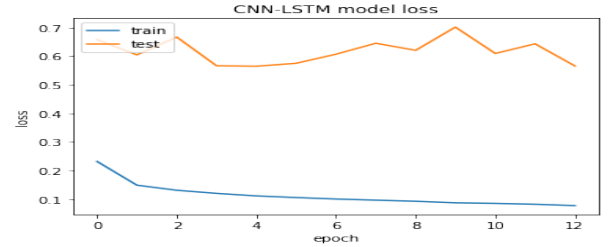Fig. 13: Training and validation accuracy using CNN+LSTM architecture

Fig. 14: Training and validation loss using CNN+LSTM architecture

2) Personalized learning results: The main aim that leads us to think abou the personalized learning is to improve those previously studied models by reducing overfitting and improving the accuracy.
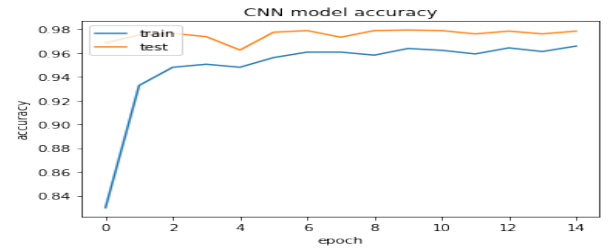

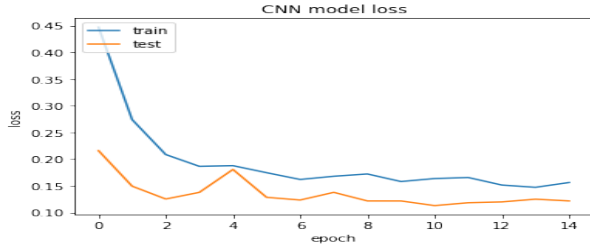Fig. 15: CNN model accuracy on training and validation

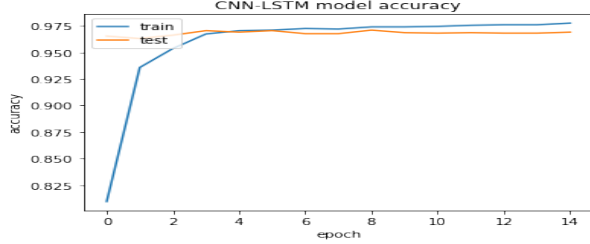Fig. 16: CNN model loss on training and validation

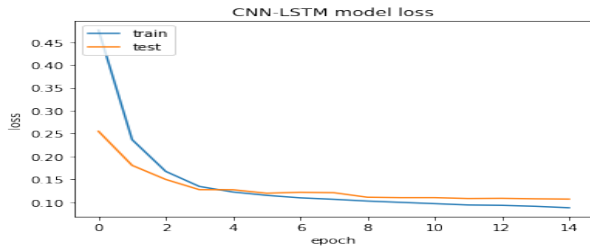Fig. 17: CNN+LSTM model accuracy on training and validation

Fig. 18: CNN+LSTM model loss on training and validation

TABLE 3: Results of the second approach

NN topology	Train accuracy	Val accuracy	Train loss	Val loss
CNN	96.59	97.85	0.1561	0.1216
CNN+LSTM	97.76	96.91	0.0878	0.1067

V. CONCLUDING REMARKS

In this paper, we proposed an automatic heart arrhythmia detection solution able to classify the extracted and processed ECG signal into normal and abnormal heartbeats. Two approaches: generalized and personalized learning based on CNN and RNN networks were presented and studied. The achieved reported results with the applied techniques and data denoising are considered as outstanding results that outperform many other approaches. We can clearly notice that the personalized learning performs better than the generalized learning in terms of accuracy and loss even while using much simpler architectures. This difference in performances can be interpreted by the uniqueness of the beat's morphology in a single subject record, which allows the network to easily learn it discarding all kind of confusions.

REFERENCES

[1] S. C. A. S. H. Gottlieb, "Silent ischemia on holter monitoring predicts mortality in high-risk postinfarction patients," *JAMA*, vol. 259, pp. 1030–1035, 1988.

[2] M. Rangasamy, "Role of telemedicine in health care system: a review," *Int J Recent Adv Pharma Res*, vol. 2, pp. 1–10, 2011.

[3] W. V. Lerberghe, "The world health report 2008: primary health care: now more than ever," *World Health Organization,*, 2008.

[4] A. A. M. Naghavi, C. Abbafati, "Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the global burden of disease study 2016," *The Lancet*, vol. 390, pp. 1151–1210, 2017.

[5] A. C.-S. K. Dickstein, "Esc guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The task force for the diagnosis and treatment of acute and chronic heart failure 2012 of the european society of cardiology. developed in collaboration with the heart," *European journal of heart failure*, vol. 14, pp. 803–869, 2012.

[6] A. C.-S. K. Dickstein, "Emotional response and changes in heart rate variability following art-making with three different art materials," *New England Journal of Medicine*, p. 968, 2018.

[7] R. G. M. Moody, "The impact of the mit-bih arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, pp. 45–50, 2001.

[8] A. N. Uwaechia and D. A. Ramli, "A comprehensive survey on ecg signals as new biometric modality for human authentication: Recent advances and future challenges," *IEEE Access*, 2021.

[9] S. Poungponsri and X.-H. Yu, "An adaptive filtering approach for electrocardiogram (ecg) signal noise reduction using neural networks," *Neurocomputing*, vol. 117, pp. 206–213, 2013.

[10] S. L. Joshi, R. A. Vatti, and R. V. Tornekar, "A survey on ecg signal denoising techniques," in *2013 International Conference on Communication Systems and Network Technologies*, pp. 60–64, IEEE, 2013.

[11] B. Singh, P. Singh, and S. Budhiraja, "Various approaches to minimise noises in ecg signal: A survey," in *2015 Fifth International Conference on Advanced Computing & Communication Technologies*, pp. 131–137, IEEE, 2015.

[12] M. Alfaouri and K. Daqrouq, "Ecg signal denoising by wavelet transform thresholding," *American Journal of applied sciences*, vol. 5, no. 3, pp. 276–281, 2008.

[13] T. W. J. Pan, "Real time algorithm detection for qrs," *IEEE Trans. Eng. Biomed Eng*, vol. 32, pp. 230–236, 1985.

[14] S. M. Manikandan, "A novel method for detecting r-peaks in electrocardiogram (ecg) signal," *Biomedical Signal Processing and Control*, vol. 7, pp. 118–128, 2012.

[15] B. J. Gersh, "A novel method for detecting r-peaks in electrocardiogram (ecg) signal," *Mayo Clinic Heart Book*, pp. 6–8, 2000.

[16] A. Norhen, P. Danilo, and B. Arcangelo, "Fixed complexity tiny reservoir heterogeneous network for on-line ecg learning of anomalies," in *2021 IEEE 10th Global Conference on Consumer Electronics (GCCE)*, pp. 233–237, IEEE, 2021.

[17] P. Kim, "Convolutional neural network," in *MATLAB deep learning*, pp. 121–147, Springer, 2017.

[18] L. R. Medsker and L. Jain, "Recurrent neural networks," *Design and Applications*, vol. 5, pp. 64–67, 2001.

[19] D. L. Elliott, "A better activation function for artificial neural networks," tech. rep., 1993.