Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

Classifying ECG Signals with an Effective Machine Learning Method

Fatimah M. Raheem[1*] , Heyam A. Marzog[1]

[1]Department of Communication Engineering, Engineering Technical College of Najaf, Al-Furat Al-Awsat Technical University, Najaf 31001, Iraq. E-mail:fatimah.albadiri.ms9@student.atu.edu.iq, heyam.marzog@atu.edu.iq
*Corresponding author: fatimah.albadiri.ms9@student.atu.edu.iq

https://doi.org/10.46649/fjiece.v4.2.11a.24.9.2025

Abstract. *The goal is to compare deep learning and classical methods for automated ECG beat classification in a rigorous inter-patient evaluation. Band-pass filtering, baseline correction, Kalman denoising, and QRS detection using the Pan-Tompkins algorithm were applied to 48 half-hour, two-lead recordings from 47 subjects from the MIT-BIH Arrhythmia Database. Discrete wavelet transforms features fed a tree-based baseline (random forest, RF) following R-peak segmentation were compared to a convolutional neural network (CNN). Per-class precision/recall/F1, macro-F1, and ROC-AUC/PR-AUC with 95% bootstrap CIs were reported. In a four-class setup (atrial fibrillation, normal, other rhythms, and noise), RF obtained macro-F1 ≈ 0.70 and CNN macro-F1 ≈ 0.59. AF↔Other, Noise↔Other, and Noise↔AF were the most commonly confused terms. In inter-patient evaluation, RF serves as a competitive baseline. Comparability across studies is improved by PR-centric evaluation, standardized reporting, and clear preprocessing, especially when there is a class imbalance.*

Keywords: *classification; Machine learning; Arrhythmia detection; Deep learning; Signal processing.*

1. INTRODUCTION

Text An electrocardiography (ECG) is a recording of electrical activity of the heart and a central part of cardiovascular disease diagnosis and management. Irregularities in ECG waveforms are good predictors of rhythm disorders (arrhythmias). An average cardiac cycle has the following waves; the P wave (atrial depolarization), the QRS complex (ventricular depolarization), and the T wave (ventricular repolarization) [1].

Although most early methods regarded ECG signals as being stationary and primarily used frequency-domain analysis as their approach, recent research highlights the fact that they are actually non-stationary, i.e. spectral contents change with time [2]. Simultaneously, technological improvements in digital sensing and computing have increased the size of ECG devices and data-acquisition systems and boosted the creation of computer-aided diagnostic pipelines. The presence of substantial and well-structured ECG data has also allowed more progress in arrhythmia detectors in design and assessment [3].

The ECG beat classification generic workflow presented in Figure 1 comprises four stages, namely signal preprocessing, heartbeat (beat) detection, feature extraction, and classifier development [4]. A four-class beat taxonomy, including atrial fibrillation (AF), normal sinus rhythm (N), remaining rhythms (O), and noise, is used in this work. This taxonomy is used continuously in data partitioning, training and evaluation. Also to put it into perspective, previous literature usually lists five-beat labelling (e.g., atrial premature contraction (A), right bundle block (R), left bundle block (L), and premature ventricular contraction (V) in addition to normal); in this case we map such beat types into the other category of rhythms unless stated otherwise.

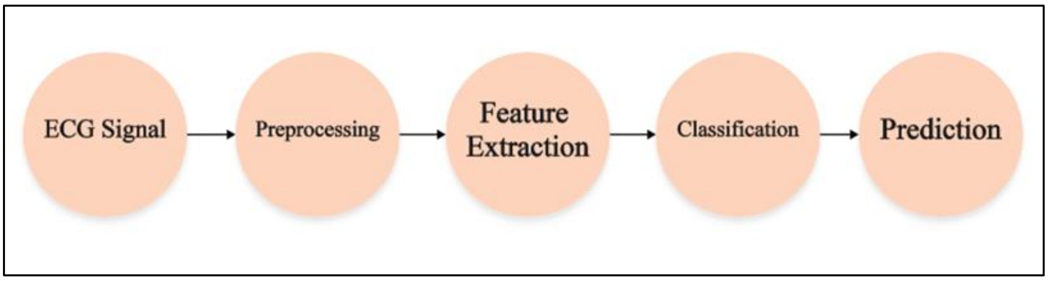

Fig. 1. General scheme of classification of electrocardiogram.

The effective deployment of the ECG analysis methodology relies on the correct amplitude and timing-indicators to support the expert and automated pipelines [5]. Beat discrimination can use tree-based featuring classifiers (e.g., random forests), fed by RR-interval and transform-domain descriptors (e.g., DCT/DWT) [6,7], and modern neural networks (e.g., CNNs) use features directly in the waveform. The reported headline accuracies differ broadly among studies as they are based on data set composition and, most importantly, whether it is intra patient testing or the more realistic inter-patient (patient-exclusive) testing (patient-only) [8]. In order to have fair comparison in the imbalance of classes, we highlight macro-F1 and PR-AUC with per-class precision/recall/F1.

The principal findings of the research will be as follows:

1. Kalman filtering and DWT are combined to perform the problem of ECG denoising.
2. the performance of the random forest and CNN classifiers are compared.
3. a knowledge distillation framework is implemented in order to achieve better results in ECG classification.

2. DATABASE

The construction of the ECG classification model usually incorporates the division of the set of data into two rather different parts: a training set (D1 or DS1) and a testing one (D2 or DS2). This division gives equal ratio of heartbeat classes and patient groups. The information division is based on an between-patient design, in which the subjects used in training and testing are totally separate, and hence, data leakage and bias in the assessment of model performance are avoided [9].

The 48 half-hour ECG records in MIT-BIH Arrhythmia Database are categorized into two mutually exclusive subsets DS1: training, and DS2: testing [10]. This design allows model training and tuning to be done on a cohort of subjects and testing on a separate cohort which enhances the ability of the model to generalize to unknown data.

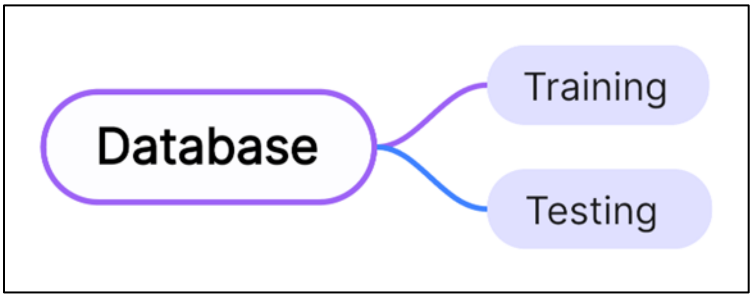

Fig. 2. Database.

3. MACHINE LEARNING

As a subfield of artificial intelligence machine learning (ML) develops mathematical programs which analyze substantial datasets to produce logical decisions. Through statistical analysis and computational modeling machines can recognize data patterns as they learn performance improvement from practical experience rather than manual programming [11]. The fundamental process in ML begins with model training because algorithms utilize labeled or unlabeled data to learn from. (entity learning method uses labeled datasets to predict sample outputs but unsupervised algorithms detect hidden patterns in unlabeled information. The decision optimization mechanism in reinforcement learning operates based on reward-based feedback [12].

The different sectors employ machine learning technology to benefit from natural language processing (NLP) along with financial market forecasting medical diagnostics and autonomous systems implementation. The use of deep learning approaches has led to major improvements in automated language translation systems which now offer better contextualized translations [13]. Predictive analytics powered by ML plays a fundamental role in stock market analysis because it uses historical data to recognize patterns and generate future price predictions according to [14].

The healthcare field uses ML algorithms to diagnose medical issues through the examination of images and the identification of anomalies while helping predict diseases. The diagnostic accuracy of cancer and diabetic retinopathy detection relies on convolutional neural networks (CNNs) according to research presented in [15].

The diverse strengths of ML come with several obstacles which consist of data biases together with complexities in model interpretation and computation processes. Scientific teams work to find solutions which enhance the strength and ethical aspects and operational efficiency of ML models when applied in practical settings [16].

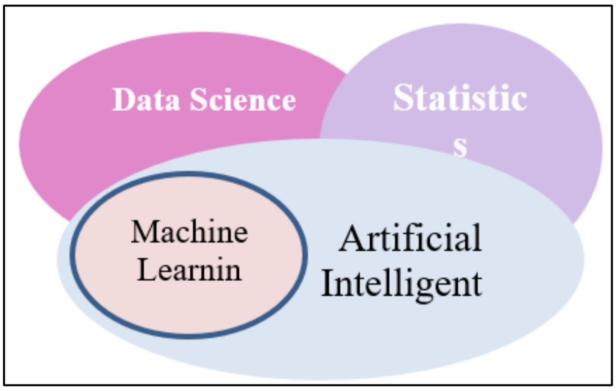

Fig. 3. Different between data science, artificial intelligent, statistics and machine learning.

Machine learning model can be categorized into three general types as shown in “Fig. 4”.

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

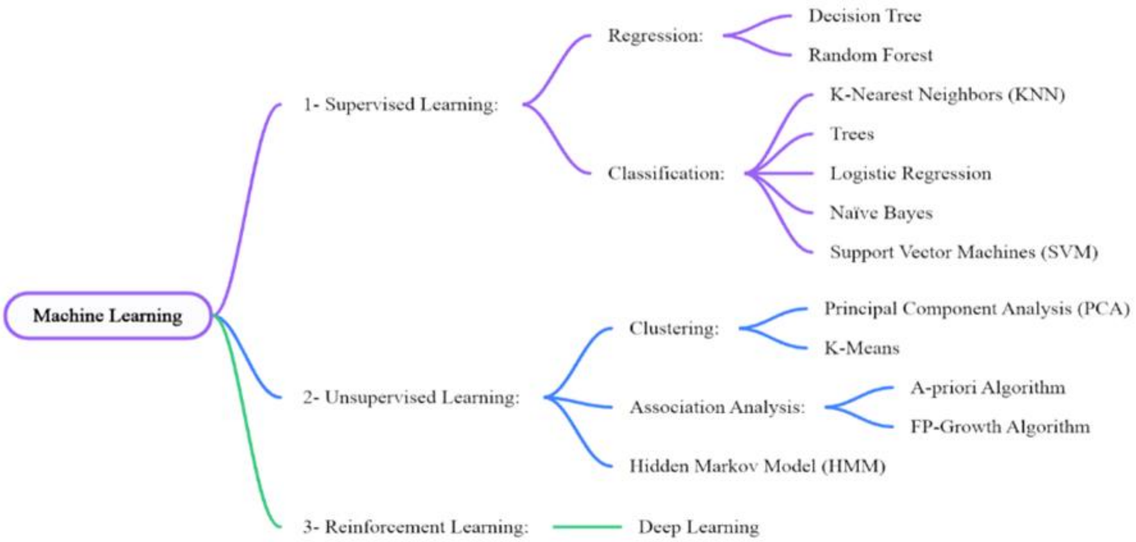

Fig. 4. Family of Machine learning.

3.1 Supervised Learning

1. The machine learning process of supervised learning trains algorithms through the use of datasets that have labeled inputs alongside their corresponding outputs. Such training allows the model to study historical datasets which helps achieve correct future predictions. The system acquires classification features for pictures by processing images that human annotators have labeled as "girl" [17]. Supervised learning algorithms execute their processes as demonstrated in Figure 3. Multiple supervised learning techniques which are commonly used will be examined in this section.
2. Decision Tree: The classification method known as decision trees uses hierarchical structures to arrange data attributes according to their values. Naïve Bayes exists primarily to categorize different information types. Each node in the tree stands for an attribute while branches depict potential values that attribute can achieve. Both medical diagnosis together with credit risk assessment and customer segmentation make extensive use of decision trees which are actively applied in various real-world domains [18].
3. Naïve Bayes: The Naïve Bayes classifier works as a probabilistic system which specifically addresses text classification needs and serves for clustering operations and spam detection tasks. The classifier operates using Bayes' theorem to determine probabilistic values when each feature maintains a separate relationship to other features. The basic nature of Naïve Bayes allows it to deliver outstanding results in 4) Support Vector Machine (SVM) applications.
4. The Support Vector Machine (SVM) represents one of the leading advanced classification methods which machine learning practitioners use extensively. SVM finds the best possible hyperplane to create the largest distance space between different classification areas. The method uses margin-based principles to reduce classification mistakes as well as improve model ability to generalize. SVM exhibits superior performance in various fields such as image recognition as well as bioinformatics and handwriting recognition [23]. such as sentiment analysis and document classification [19]. A Bayesian Network representation exists for the model while preserving the probabilistic variable relationships [20][21][22].
5. Logistic Regression: Logistic Regression represents a supervised learning method dedicated to creating binary classifications. The model generates output probabilities which

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

range from 0 to 1 for determining the membership of an input to a particular class. A prediction system identifies how many images have men and women along with babies. The logistic (sigmoid) function transforms prediction outcomes into probabilistic categories because of its suitability for categorical outputs [24].

6. k-Nearest Neighbors Algorithm (k-NN): The k-Nearest Neighbors (k-NN) serves as a basic non-parametric approach for supervised learning that enables classifications together with regressions through its data management tasks. Unlabeled data points find their categories through k-NN by selecting the nearest 'k' labeled data points in the feature space that use distance measures such as Euclidean distance. The performance of the model depends directly on the value selection of 'k' which represents any positive integer. The k-NN seeks the 'k' nearest neighbors from each unlabeled instance to determine the most frequent class assignment [25].

3.2 Unsupervised Learning

Without supervised control algorithms search for patterns within datasets that contain no predefined class specifications. These algorithms can perform several tasks which include clustering applications and association rule learning followed by dimensionality reduction. Unsupervised learning analyzes customer buying patterns to discover which products frequently accompany other items through its examination of data in e-commerce recommendation systems [26].

1. K-Means Clustering: This unsupervised learning method divides data points between 'k' clusters through feature similarity assessments. A part of the iteration process takes each data point to its correct cluster while reducing the variance within groups. Each product placed inside its cluster becomes similar to all cluster members but less comparable to items in different clusters. Prospective users find K-Means valuable for market segmentation together with image compression and pattern recognition applications [27].
2. Principal Component Analysis (PCA): By using Principal Component Analysis (PCA) we can transform data of many dimensions into reduced dimensions while maintaining most dataset variability. PCA analyzes data to identify principal components that represent maximal variance which subsequently simplifies datasets and speeds up computations and removes noise effects. PCA has strong applications in visualizing data structures as well as removing noise when preparing data for subsequent machine learning modeling processes [27].

3.3 Reinforcement Learning

The learning process of Reinforcement Learning (RL) allows agents to build optimal decision solutions through environmental interactions. The system receives a mark-up or a penalty from the environment which helps direct its behavior adjustment. An RL agent operating in autonomous driving learns its safety protocols through reward feedback that gives it points for staying in its lanes yet deducts points for deviating from the path.

4. CLASSIFICATION OF ECG SIGNALS

Detecting and diagnosing different cardiac arrhythmias requires immediate classification of ECG signals. Alternative analysis methods for ECG signals enable doctors to detect irregular heart rhythms and perform correct classifications.

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

Experts in cardiology fields currently use their eyes to study ECG waveforms for classifying multiple arrhythmia types. Computational technique development has improved both the classification speed and accuracy rates leading to earlier detection abilities in myocardial infarction (MI) and ischemic heart disease [29].

Quantitative analysis of ECG abnormalities is now achieved through artificial neural networks (ANNs) and mathematical models developed with artificial intelligence (AI) and machine learning (ML) advances. The detection of fiducial points starts with identifying the QRS complex because this essential point helps determining cardiac conditions during analysis. For the past four decades researchers have continuously worked on QRS detection because it stands as the essential step toward successful ECG classification [28]. The acquired and preprocessed ECG signal enters the classification algorithms which identify and separate the signal into various arrhythmia groupings. Two categories of ECG classification algorithms are discussed in this section: traditional and machine-learning-based approaches.

4.1 Traditional ECG Classification Approaches

The initial classification methods used threshold-based analysis to classify ECG signals as either normal or abnormal. Researchers improved adaptive threshold approaches by using the Pan-Tompkins algorithm that stands as a strong tool for QRS detection [30].

The use of wavelet transform methods involves DWT with PCA and ICA integration to extract significant features from ECG signals [31]. The Multi-Model Decision Learning (MDL) algorithm achieves high sensitivity values up to 100% for distinguishing between normal and abnormal ECG signals when analyzing data from the MIT-BIH Arrhythmia Database thus becoming a reliable traditional method for classification [32].

The model learns to map input features to predefined output classes. Unsupervised Learning: This technique focuses on detecting patterns and structures in unlabeled data without predefined classifications.

The ECG classification has benefited from supervised learning algorithms including k-Nearest Neighbors (k-NN) and Naïve Bayes and Decision Trees as well as Support Vector Machines (SVM) and Artificial Neural Networks (ANNs). The algorithms need precise feature selection methods to obtain important information from ECG signals according to [34].

4.1.1 Decision Tree-Based Classification

Decision trees represent one of the primary methods used for ECG classification through tree-based algorithms to produce trained and tested models. The preprocessed ECG features serve as inputs to decision tree classifiers used for training purposes. Decision trees operate efficiently to perform classifications and predictions, cluster data correctly within ECG analysis systems.

The J48 Decision Tree Algorithm stands out among different decision tree solutions as it excels with high-dimensional data analysis while offering straightforward implementation capabilities. The J48 algorithm creates optimized decision trees through its pruning method which enhances the classification results [35].

The Classification and Regression Tree (CART) algorithm stands as a standard tool in ECG classification as it develops binary decision trees which separate between normal and abnormal heart rhythms [36]. The research of Iryna Mykoliuk et al. incorporated a decision tree algorithm which operated as a part of their machine learning-based ECG classification system [37]. The initial implementation started with the random forest algorithm which proved effective due to its divide-and-conquer rule that supports parallel processing and fast training at an average of a few seconds. Random forests provide efficient optimization of tree configurations because of their distinctive property. The classification required a random forest model containing 60 decision trees. The implementation of the model and evaluation used the Scikit-learn library. Table 1 shows random forest algorithm performance by measuring different class precision, recall and F1-score metrics.

Table 1. Performance Metrics of the Random Forest Model.

Class Identifier	Precision	Recall	F1-Score
1	0.77	0.67	0.72
2	0.73	0.75	0.74
3	0.64	0.68	0.66
4	0.68	0.61	0.65
Average/Total	0.7	0.7	0.7

The classification system used a convolutional neural network (CNN) design with the network architecture performing feature extraction before the fully connected layer made final classification choices. The results of CNN model classification performance can be found in Table 2.

Table 2: Performance Metrics of the Neural Network Model

Class Identifier	Precision	Recall	F1-Score
1	0.58	0.26	0.36
2	0.65	0.94	0.77
3	0.6	0.28	0.38
4	0.47	0.03	0.06
Average/Total	0.62	0.64	0.59

According to the design the classification system has four distinct ECG signal categories serving as classification indicators. These categories include Atrial fibrillation and Normal sinus rhythm as well as Other cardiac rhythms and Noise signals.

- Class 1: Atrial fibrillation
- Class 2: Normal sinus rhythm
- Class 3: Other cardiac rhythms
- Class 4: Noise signals

- *Evaluation Metrics*

Performance assessments of the model relied on an evaluation metrics framework.

1. The classification accuracy is measured through Precision by determining true positive predictions against total predicted positives.
2. Recall represents the proportion of real positive cases that the model correctly identifies out of all actual positive conditions.
3. The F1-Score combines precision and recall by using this calculation:

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (1)$$

The metric allows impartial assessment of precision together with completeness across all ECG classes. The F1-score computation applied to the three principal diagnostic groups including atrial fibrillation, normal sinus rhythm and other cardiac rhythms alone.

The random forest algorithm obtained superior results than the CNN model by delivering a F1-score value of 0.70 while CNN provided a score of 0.59. ECG signal classification abilities are demonstrated by decision trees which perform especially well while maintaining computational efficiency standards. Thus, the final result of the pattern is a rate of F1 degrees for 3 cases.

$$F1 = \frac{F1(AF)+F1(N)+F1(0)}{3} \quad (2)$$

4.1.2 Naïve Bayes Classifier

Dr. S. Padmavathia and E. Ramanujam proposed the implementation of the Naïve Bayes algorithm to identify ECG signals through normal or pathological categories because of its specific approach.

Naïve Bayes classifiers function as statistical classifiers to calculate data point category membership probabilities by determining which category each point belongs to [38]. Large datasets experience exceptional accuracy and fast computational performance when Bayes classifiers are applied to them. The Naïve Bayes classifier analyzes which features matter most for classification outcomes through an evaluation that ignores other present features.

The simplification algorithm depends upon this assumption that is known as category conditional independence to become "naïve."S. T. Aarthy and J. L. Mazher Iqbal also proposed using the Naïve Bayes method for ECG signal classification. Their approach was effective in real-time ECG classification using Naïve Bayes combined with fuzzy logic for disease prediction. Initially, the method reads the available ECG signals and removes noise. Then, the relevant features are extracted from the graphical representation of the signals. Any faulty or missing signals are discarded from the dataset before proceeding to feature extraction.

A fuzzy-based disease classification system is built using extracted features after successful preprocessing termination. The proposed calculation method derives conditional probabilities through the association of features inside the fuzzy system. The process of disease classification or prediction relies on MFSS to evaluate signal similarities for diagnosis purposes. The obtained MFSS values help calculating cardiac disease procumbent weight (CDPW) for distinct categories before performing classification through CDPW values [39][40] .

The artificial neural network (ANN) functions by connecting interlinked neurons which use activation functions to calculate their outputs based on differences among input attributes. The adjustment of output depends on modifying both the weight and activation function after assigning a weight to an input. Output adjustments from the learning process get established according to original input values. The essential components of an ANN include:

- Neurons, where each neuron (j) receives an input signal $P_j(t)$ from previous neurons at a discrete time step t.
- Activation function $a_j(t)$, which determines the neuron's response.
- Threshold value, which remains unchanged unless the learning rate is modified.
- A function f that computes the updated activation state $a_j(t+1)$based on previous activation and input values, expressed as:

$$a_j(t+1) = f\left(a_j(t), P_j(t)\right)\theta \quad (3)$$

- Finally, the output function O_j (t), is derived accordingly.

$$O_j(t) = f_{out}\left(a_j(t)\right) \quad (4)$$

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

- Weight and bias All neurons hold a one-to-one mapping relation between their elements and single weight values. The subscripts i and j in wij represent previous neuron output and input neuron respectively.
- The subscripts indicate input neurons through (i) while the subscripts (j) indicate to the output neuron. Through change of the weight adjustment helps us to enhance performance results.
- Bais expression is the outcome from adding all input weights translates into an activation function which functions as the step serves as threshold that results in transfer processing through the activation function [1][2].
- Propagation Function The main purpose of propagation function rests in calculating the differences between predicted and actual output values. A crucial part of weight adjustment in error reduction and prediction accuracy enhancement occurs through this function. The propagation function establishes its mathematical definition as follows:

$$P_j(t) = \sum o_j(t)w_{ij} + w_{oj} \qquad (5)$$

where $w_{ij} =$ weight of the input neuron; $w_{oj} =$ bias term

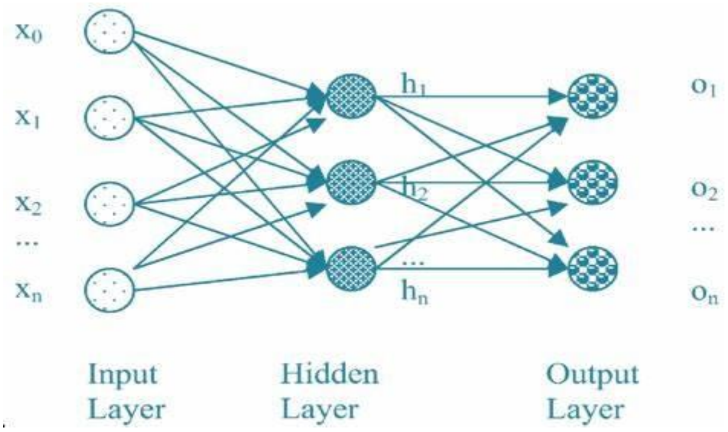

Fig. 5. Neural Network System.

4.1.3 Deep Learning (DL)

The process of developing complex models through large datasets known as training data constitutes Deep Learning (DL). Deep Learning includes the analysis of information extraction together with intelligent decision-making and predictive modeling. Deep Neural Networks (DNNs) provide better scalability and flexibility than traditional machine learning methods by deepening networks or adding training dataset points to enhance precision [12].

Most conventional learning systems encounter challenges with advanced applications as they need large amounts of labeled data for efficient generalization. The broad adoption of deep learning occurs across multiple sector applications which include object detection alongside biometric authentication and image classification and computer vision systems. ECG arrhythmia detection proves most effective in cardiology when Deep Convolutional Neural Networks (Deep-CNNs) are applied according to research [42].

- **Multilayer Perceptron (MLP):** Multi-layer Perceptron (MLP) functions as a prevalent supervised neural network which handles advanced learning operations. Multiple neuron layers composed of interconnected components form the basic structure of MLPs that work along a feed-forward system. The processing of weighted inputs by each neuron produces an output through a nonlinear activation function. Researchers use MLPs as a standard tool for solving classification together with regression problems [14].

- **Convolutional Neural Network (CNN):** The Convolutional Neural Network (CNN) works as a popular deep learning architecture through training based on gradient-based optimization methods. A Convolutional Neural Network being built by multiple sequential layers functions with a feed-forward arrangement. Its main layers include:
    - Convolutional Layer – Extracts spatial and hierarchical features from input data.
    - The Pooling Layer maintains key features by combining data while it decreases the dimensionality of input data.
    - Normalization Layer – Enhances learning stability.
    - Fully Connected Layer – Responsible for the final classification.

Feature extraction functions at the first three layers but the fully connected layer executes the classification step. CNN stands as an extension of MLP which includes convolutional operations to learn spatial object hierarchies in data. Medical images and ECG signals benefit from CNN performance enhancement because the networks can extract vital features by themselves [43].

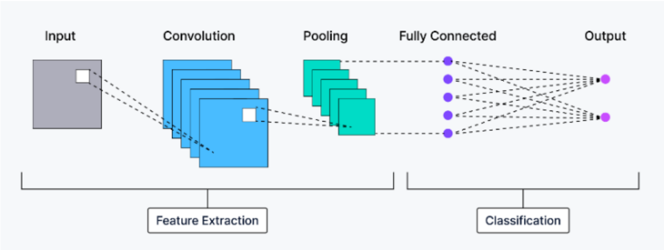

Fig.6. Architecture of convolution Neural network.

- **Deep Belief Network (DBN):** A Deep Belief Network (DBN) integrates numerous levels of Restricted Boltzmann Machines (RBMs). DBNs function as advanced learning platforms which implement temporal random variable transformations. Multiple RBM layers build these networks to transmit input data from one layer to the next [44]. DBN training adopts a stratified strategy which trains Restricted Boltzmann Machines beginning from the base layers up toward the superior layers. The RBM model functions as an undirected probabilistic graphical model which processes binary random variables exceptionally well in the analysis of binary valuated data [45].

In an RBM the visible layer functions as an input data representation while the hidden layer performs feature extraction together with pattern detection. The network displays weighted link connections between consecutive layers that are labeled as w_{ij} in Figure 8. There are no relationships between individual nodes positioned in the identical layer. [46]. The basic unit of RBM called Boltzmann Machine forms Markov random fields through symmetrical networks made of binary stochastic units.

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

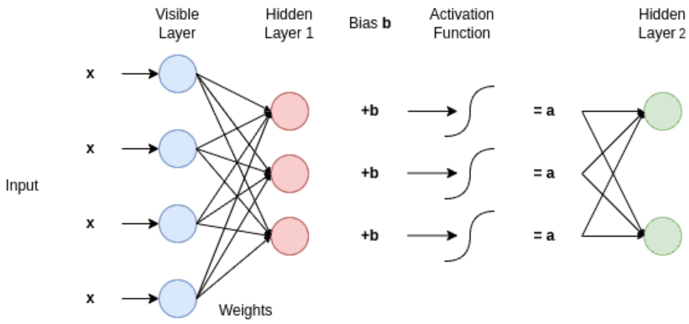

Fig. 7. Restricted Boltzmann Machine architecture.

4.1.4 Echo State Network (ESN)

Echo State Networks function as the dominant method of implementing Reservoir Computing (RC) yielding successful application across many tasks [47].

Miquel Al Faras et al. and other researchers have enhanced the ESN model through ring connection modifications that improve its performance [48]. The classification application of ESNs utilizes a methodology which combines ensemble-based learning with raw ECG waveforms and heartbeats time intervals as input information [49].

5. STRUCTURE AND FUNCTIONALITY OF ESNS

ESNs consist of three layers:

1. The input layer accepts the delivered input signals.
2. The reservoir layer exists as a random connection and weight-based recurrent neural network.
3. The reservoir utilizes the transformed features to produce predictive outputs through its output layer.

ESN performs context-specific calculations because the reservoir behaves as a dynamic system which uses previous states to influence current computations [50].

The main advantage of ESNs occurs from their ability to perform high-dimensional nonlinear input data mapping. The transformed reservoir responses become easier to classify by using simple linear regression models according to [51].

The initialization process of the reservoir during ECG signal processing requires ECG data input through a random input matrix $W^{in}_{N\times d}$ the range from -1 to 1 functions as the interval for uniform distribution values. Transformation occurs to the original ECG feature vector $u\times Hb$ through a specified mathematical expression.

$$X_{N\times Hb} = W^{in}_{N\times d} \times u_{d\times Hb} \qquad (6)$$

The reservoir-system processes real-time signals through its continuous state updates which rely on the stored previous responses [52].

6. METHODOLOGY

The proposed ECG signal classification method is composed of a few primary steps that comprise data pre-processing, feature extraction, dataset balancing, model training and evaluation.

As an experimental dataset, the MIT-BIH Arrhythmia Database was utilized. The frequencies of the classes in this database are very disproportionate and therefore, as a method of balancing the sampled sizes of the classes during training, random oversampling was used. This was done to make sure that the

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

minority heartbeat groups were equally represented and would make the models to learn the discriminative features better. The data were stratified split into inter-patient (DS1/DS2), preserving the equal proportions of the labels. Such an arrangement enabled an effective assessment and reduction of information loss between training and testing data.

The proposed framework adopts a knowledge distillation framework that involves two convolutional neural networks (CNNs) which consist of a teacher model and a student model. The teacher model is a more complex 1D-CNN that consists of several convolutional layers that contain batch normalization and max-pooling, and there are additional dense layers and a softmax output layer with four heartbeat classes. The student model is a light CNN consisting of fewer convolutional layers and a dense compressed head. The student model is trained with softened output probabilities by the teacher, as well as on the true labels. The overall loss is a summation of the KullbackLiebner divergence between the teacher and student outputs and the regular cross-entropy loss. Validation has been done to tune the temperature and weighting parameters to optimize knowledge transfer.

The Adam optimizer was used to train the models with a learning rate of 0.001, a batch size of 128, and 400 epochs. To avoid overfitting, early stopping was used to track the validation macro-F1 score.

Accuracy, precipacity, recall, and F1-score on each of the classes, macro- and micro-averaged scores were used to assess model performance to control the imbalance in the classes. A confusion matrix was created as well to visualise the misclassification between the categories of heartbeat. The procedure of work are explained in block diagram as shown in figure (8). the steps of the system when entering data, processing it, and then classifying it. After obtaining the results, the system will be applied with real data as shown in Figure (8).

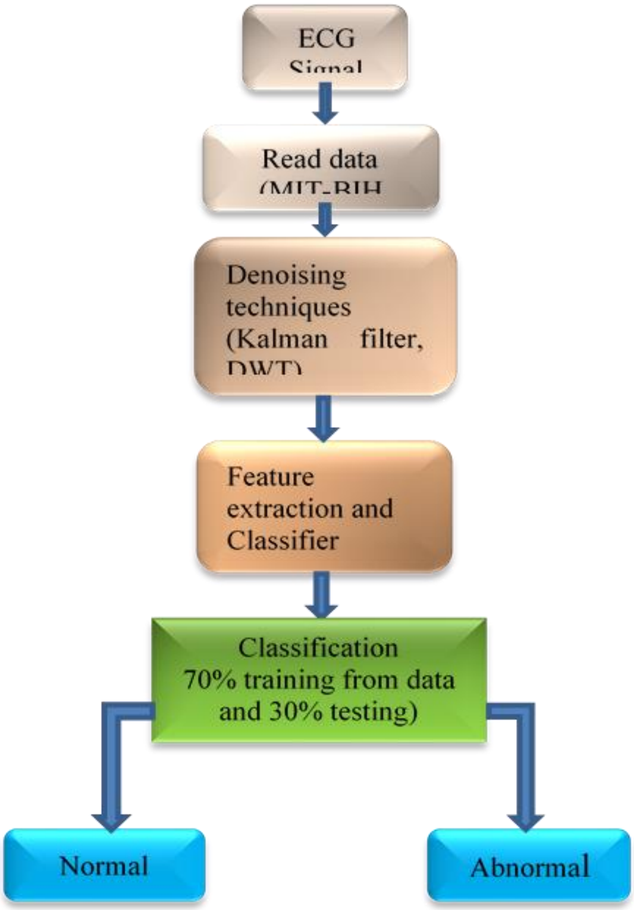

Fig. 8. The procedure of our work

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

7. RESULTS

Overall test performance. The model performed Loss = 0.3296, Accuracy = 0.9470, Precision = 0.9490 and Recall = 0.9458 on the held-out test set. These scores represent balanced behavior: low false positive (high precision) and high sensitivity to true beats (high recall).

Learning curves (Figure 9). Left panel (Training and Test Loss): The training loss gradually drops over 400 epochs and nearly goes to zero. The loss to test ratio decreases rapidly in the early 100 epochs, followed by irregular fluctuations in the range of 0.30-0.35 including a gradual upward trend, which indicates the beginning of overfitting in the subsequent epochs. Right panel (Training Metrics): Training accuracy, precision, and recall are increasing very fast and approaching 1.0. Accuracy of the test increases to approximately 0.90 at approximately epoch 60, after which, accuracy increases gradually and gets steady to some level 0.94-0.95, similar to the final accuracy (0.9470). Most of the training period has precision being slightly greater than recall, which happens when classes are imbalanced.

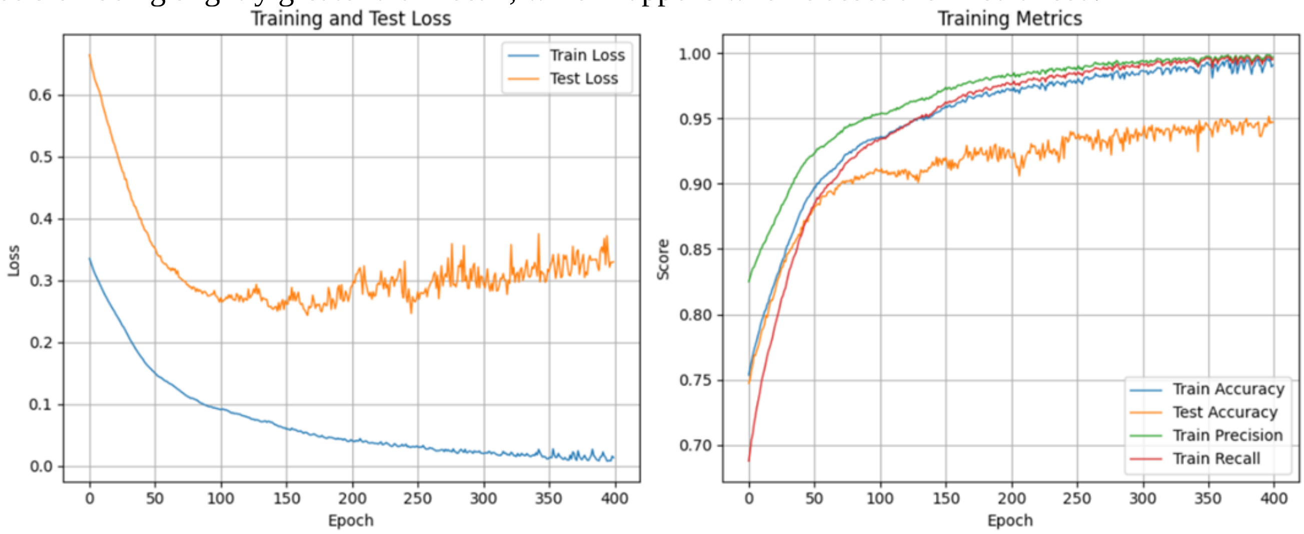

Fig. 9. Left=Training vs. test loss in 400 epochs. Right-Training accuracy/precision/recall and test accuracy of 400 or more epochs.

8. *DISCUSSION*

Generalization. The change in the divergence between the learning and the objective loss after 100 epochs (Figure 9, left) shows that the model begins to be sensitive to training-specifics. Practically, the inclusion of either early stopping on a validation set, learning-rate scheduling, or a touch more powerful regularization (e.g. dropout/weight decay) would probably identify the best generalization point and prevent the test loss decreases occurring in late epochs.

Class imbalance behavior. Figure 9 (right) metric trajectories indicate that the classifier can be described as slightly conservative, where precision is always high than recall. In order to recall minority classes without degrading the precision of the overall performance, we can use ablations using a combination of these options: **synthetic oversampling (e.g. SMOTE) or class weights or a cost-sensitive objective.

Recommended reporting.For completeness, include:

- A table with precision, recall, and F1 of the per-class (as well as macro and micro averages).
- a confusion matrix of final model.

PrecisionRecall curves per class (Optional) To show how it behaves on rarer categories.

Summary. The model provides good test results (Accuracy =0.947, Precision =0.949, Recall =0.946). The Figure 1 curves indicate simple fixes, less drastic regularization, and imbalance specificity, which can also be improved further to obtain better generalization and sensitivity to the minority classes.

9. CONCLUSIONS

We compared a 4-class ECG beat classifier based on feature-based Random Forest (RF) and a 1D-CNN in the presence of an inter-patient test with shared preprocessing. A balanced model (Accuracy = 0.947, Precision = 0.949, Recall = 0.946) provided the best model. Macro-F1 (per-class) of RF 0.70 and CNN 0.59 with most of the confusions in the form of AF addingOther and Noise adding (AF/Other). Comprehensively, an adjusted tree-based baseline is very competitive with a lightweight CNN in case of class imbalance.

Following up on the existing inter-patient evaluation system with four classes, a number of research hypotheses could be suggested in the further work:

1. External validation: In order to test the robustness and generalizability of the developed pipeline, the pipeline will be tested on other ECG databases whose noise properties are different.

2. Prospective: Clinician-in-the-loop studies will be performed to assess clinical usefulness, integration in workflow, and applicability in the real world.

3. Harmonizing imbalance: Various methods like class weighting, focal loss and data-level balancing will be systematically compared to enhance minority-class recall without affecting precision.

4. Model calibration and uncertainty estimation: Improved model calibration (e.g., Expected Calibration Error) and quantification of uncertainty are going to be studied to allow making risk-perceptive decisions.

5. Explainability: Signal-conscious interpretability methods (e.g., saliency visualization of ECG waveforms) will be included to make models more transparent and clinically trustworthy.

6. Ablation and optimization: Intensive ablation experiments will be carried out on preprocessing techniques, network architecture depth and regularization techniques. Reproducible hyperparameter optimization will also be put in place.

7. On-deployment: System profiling and optimization will be conducted on embedded and edge hardware environments in order to obtain efficient on-device inference with low latency.

8. Reproducibility: It will be publicly available to achieve any level of reproducibility by making all code, configurations, and fixed DS1/DS2 subject lists publicly available in order to enable fair benchmarking in future studies.

Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE)
ISSN -2708-3985

REFERENCES

[1] N. Sharma and R. K. Sunkaria, “Improved T-wave detection in electrocardiogram signals based non-stationary wavelet transform and QRS complex cancellation with kurtosis analysis.,” Physiol. Meas., vol. 44, no., doi: 10.1088/1361-6579/ad0b3e. 12, Dec. 2023.

[2] A. K. Singh and S. Krishnan, ECG signal feature extraction trends in methods and applications, vol. 22, no. 1. BioMed Central, doi: 10.1186/s12938-023-01075-1. 2023.

[3] A. Darmawahyuni et al., “An improved electrocardiogram arrhythmia classification performance with feature optimization,” BMC Med. Inform. Decis. Mak., vol. 24, no. 1, doi: 10.1186/s12911-024-02822-7. 2024.

[4] H. A. Marzog and H. J. Abd, “ECG-signal Classification Using efficient Machine Learning Approach,” HORA 2022 - 4th Int. Congr. Human-Computer Interact. Optim. Robot. Appl. Proc., doi: 10.1109/HORA55278.2022.9800092. 2022.

[5] W.-S. Baek, D.-M. Kim, F. Bashir, and J.-Y. Pyun, “Real life applicable fall detection system based on wireless body area network,” in 2013 IEEE 10th Consumer Communications and Networking Conference (CCNC), pp. 62–67. doi: 10.1109/CCNC.2013.6488426,2013.

[6] S. Karpagachelvi, M. Arthanari, and M. Sivakumar, “ECG Feature Extraction Techniques - A Survey Approach,” Int. J. Comput. Sci. Inf. Secur., vol. 8, no. 1, pp. 76–80, [Online]. Available: http://arxiv.org/abs/1005.0957, 2010.

[7] R. Ganeshkumar and D. YSKumaraswamy, “Investigating Cardiac Arrhythmia in ECG using Random Forest Classification,” Int. J. Comput. Appl., vol. 37, no. 4, pp. 31–34, doi: 10.5120/4599-6557, 2012.

[8] V. Chawla, P. Kharmale, S. Kanbargi, S. Dalvi, and V. Gutte, “Classification of Cardiac Arrhythmia Using Machine Learning Algorithms BT - Intelligent Systems Design and Applications,” A. Abraham, A. Bajaj, and T. Hanne, Eds., Cham: Springer Nature Switzerland, pp. 40–52, 2024.

[9] P. Rashidi and A. Mihailidis, “A survey on ambient-assisted living tools for older adults.,” IEEE J. Biomed. Heal. informatics, vol. 17, no. 3, pp. 579–590, doi: 10.1109/jbhi.2012.2234129, May 2013.

[10] M. Alfaras, M. C. Soriano, and S. Ortín, “A Fast Machine Learning Model for ECG-Based Heartbeat Classification and Arrhythmia Detection,” Front. Phys., vol. 7, no. July, pp. 1–11, 2019, doi: 10.3389/fphy.2019.00103.

[11] J. Schmidhuber, “Deep Learning in neural networks: An overview,” Neural Networks, vol. 61, pp. 85–117, 2015, doi: 10.1016/j.neunet.2014.09.003.

[12] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[13] T. Hastie, R. Tibshirani, and J. Friedman, “The elements of statistical learning,” Citeseer. 2009.

[14] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, Deep learning, vol. 1, no. 2. MIT press Cambridge, 2016.

[15] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[16] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” in Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13, Springer, pp. 818–833. 2014.

[17] E. J. da S. Luz, D. Menotti, and W. R. Schwartz, “Evaluating the use of ECG signal in low frequencies as a biometry,” Expert Syst. Appl., vol. 41, no. 5, pp. 2309–2315, 2014.

[18] I. A. Zriqat, A. M. Altamimi, and M. Azzeh, “A Comparative Study for Predicting Heart Diseases Using Data Mining Classification Methods,” vol. 14, no. 12, pp. 868–879, 2017, [Online]. Available: http://arxiv.org/abs/1704.02799

[19] R. W. M. H. J. Abd, "An Efficient Technique of QRS Complex Detection of Electrocardiography Signal Based on Optimized Median Filter and Efficient Signal Reshaping," Int. J. Syst. Signal Control Eng. Appl, vol. 12, no. 1, pp. 13–20, 2019.
[20] R. Shankari, S. A. Safana, M. Kaviya, and L. Charles, "FPGA-based electrocardiography (ECG) signal analysis system using infinite impulse response (IIR) filter," Int. J. Appl. Eng. Res., vol. 14, no. 6, 2019.
[21] D. Koller and N. Friedman, Probabilistic graphical models: principles and techniques. MIT press, 2009.
[22] J. Miah, D. M. Ca, M. A. Sayed, E. R. Lipu, F. Mahmud, and S. M. Y. Arafat, "Improving Cardiovascular Disease Prediction Through Comparative Analysis of Machine Learning Models: A Case Study on Myocardial Infarction," 2023 15th Int. Conf. Innov. Inf. Technol. IIT. 2023, pp. 49–54, 2023, doi: 10.1109/IIT59782.2023.10366476.
[23] C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, pp. 273–297, 1995.
[24] A. Oza and A. Bokhare, "Diabetes prediction using logistic regression and K-nearest neighbor," in Congress on Intelligent Systems: Proceedings of CIS 2021, Volume 2, Springer, pp. 407–418. 2022.
[25] S. Thiyagarajan and T. Chakravarthy, "A Survey on ECG Data Classification Using Machine Learning Approaches," Adalya J., vol. 8, no. 8, 2019.
[26] [26] S. K. Addagarla and A. Amalanathan, "Probabilistic unsupervised machine learning approach for a similar image recommender system for E-commerce," Symmetry (Basel)., vol. 12, no. 11, p. 1783, 2020.
[27] C. Ding and X. He, "K-means clustering via principal component analysis," in Proceedings of the twenty-first international conference on Machine learning, p. 29. 2004.
[28] P. S. Hamilton and W. J. Tompkins, "Quantitative Investigation of QRS Detection Rules Using the MIT/BIH Arrhythmia Database," IEEE Trans. Biomed. Eng., vol. BME-33, no. 12, pp. 1157–1165, 1986, doi: 10.1109/TBME.1986.325695.
[29] Q. Xiao et al., "Deep learning-based ECG arrhythmia classification: A systematic review," Appl. Sci., vol. 13, no. 8, p. 4964, 2023.
[30] J. Pan and W. Tompkins, "A real-time QRS detection algorithm. IEEE Transaction on biomedical Engineering, vol," BME, vol. 32.
[31] U. R. Acharya et al., "Automated identification of shockable and non-shockable life-threatening ventricular arrhythmias using convolutional neural network," Futur. Gener. Comput. Syst., vol. 79, pp. 952–959, 2018.
[32] H. A. Deepak and T. Vijayakumar, "Review of ECG Signal Classification Using Deep Learning and Traditional Methods," Int. J. Sci. Technol. Res, vol. 9, no. 3, pp. 5683–5690, 2020.
[33] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," Science (80-.)., vol. 349, no. 6245, pp. 255–260, 2015.
[34] T. Ince*, S. Kiranyaz, and M. Gabbouj, "A Generic and Robust System for Automated Patient-Specific Classification of ECG Signals," IEEE Trans. Biomed. Eng., vol. 56, no. 5, pp. 1415–1426, 2009, doi: 10.1109/TBME.2009.2013934.
[35] S. L. Salzberg, "C4. 5: Programs for machine learning by j. ross quinlan. morgan kaufmann publishers, inc, Springer. 1993, 1994.
[36] L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone, Classification and regression trees. Routledge, 2017.
[37] I. Mykoliuk, D. Jancarczyk, M. Karpinski, and V. Kifer, "Machine learning methods in ECG classification," 2018.
[38] S. Padmavathi and E. Ramanujam, "Naïve Bayes classifier for ECG abnormalities using multivariate maximal time series motif," Procedia Comput. Sci., vol. 47, pp. 222–228, 2015.

[39] S. T. Aarthy and J. L. M. Iqbal, “RETRACTED ARTICLE: Time series real time naive bayes electrocardiogram signal classification for efficient disease prediction using fuzzy rules,” J. Ambient Intell. Humaniz. Comput., vol. 12, no. 5, pp. 5257–5267, 2021.
[40] Ö. Yildirim, “ECG beat detection and classification system using wavelet transform and online sequential ELM,” J. Mech. Med. Biol., vol. 19, no. 01, p. 1940008, 2019.
[41] S. Syama, G. S. Sweta, P. I. K. Kavyasree, and K. J. M. Reddy, “Classification of ECG signal using machine learning techniques,” in 2019 2nd International Conference on Power and Embedded Drive Control (ICPEDC), IEEE, pp. 122–128. 2019.
[42] A. Y. Hannun et al., “Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network,” Nat. Med., vol. 25, no. 1, pp. 65–69, 2019.
[43] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, “Cardiologist-level arrhythmia detection with convolutional neural networks,” arXiv Prepr. arXiv1707.01836, 2017.
[44] G. E. Hinton, S. Osindero, and Y.-W. Teh, “A fast learning algorithm for deep belief nets,” Neural Comput., vol. 18, no. 7, pp. 1527–1554, 2006.
[45] A. Adán and M. Adán, “Extracting Understandable 3D Object Groups with Multiple Similarity Metrics,” in Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 17th Iberoamerican Congress, CIARP 2012, Buenos Aires, Argentina, September 3-6, 2012. Proceedings 17, Springer, pp. 179–186, 2012.
[46] R. Salakhutdinov and G. Hinton, “Deep boltzmann machines,” in Artificial intelligence and statistics, PMLR, pp. 448–455, 2009.
[47] M. Lukoševičius and H. Jaeger, “Reservoir computing approaches to recurrent neural network training,” Comput. Sci. Rev., vol. 3,
[48] no. 3, pp. 127–149, 2009.
[49] M. Alfaras, M. C. Soriano, and S. Ortín, “A fast machine learning model for ECG-based heartbeat classification and arrhythmia detection,” Front. Phys., vol. 7, p. 103, 2019.
[50] A. Rodan and P. Tino, “Minimum complexity echo state network,” IEEE Trans. neural networks, vol. 22, no. 1, pp. 131–144, 2010.
[51] H. Jaeger, “The ‘echo state’ approach to analysing and training recurrent neural networks-with an erratum note,” Bonn, Ger. Ger. Natl. Res. Cent. Inf. Technol. gmd Tech. Rep., vol. 148, no. 34, p. 13, 2001.
[52] M. Lukoševičius, “A practical guide to applying echo state networks,” in Neural Networks: Tricks of the Trade: Second Edition, Springer, pp. 659–686, 2012.
[53] S. Scardapane and D. Wang, “Randomness in neural networks: an overview,” Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 7, no. 2, p. e1200, 2017.

