International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Analytical Progression Scale for Arrhythmia Scope prediction from Electrocardiograms

Sreedhar Jyothi [*, 1], Geethanjali Nelluri [2]

Submitted: 06/06/2022 **Accepted:** 10/09/2022

Abstract: Machine Learning (ML) techniques have exploded in popularity, especially the use of ML in automated ECG interpretation, which has been widely addressed in the literature. Other applications of machine learning in cardiac electrophysiology as well as arrhythmia are even less well recognised. Yet, the contemporary models are evincing the considerable false alarming in the process of arrhythmia prediction. In order to improve the arrhythmia prediction accuracy, this manuscript portrayed a novel analytical progression scale (APS) that learns from the given input electrocardiograms with appropriate label positive (prone to arrhythmia) or negative (not prone to arrhythmia). The experimental study has carried a 10-fold cross validation strategy on proposed and other contemporary models to scale the performance advantage of the proposed Analytical Progression Scale that compare those statistical values obtained for performance metrics such as precision, sensitivity, specificity, and accuracy. The results obtained from cross validation are evincing that the proposed model APS is outperforming the other contemporary models.

Keywords: *Analytical Progression scale, electrocardiogram, electrophysiology, cross validation, machine learning*.

1. Introduction

At the time of 20th century, the analysis of ECG has been proposed to be basic cardiovascular pathology diagnosis. In this, the heart functioning would be estimated by ECG signals. Therefore, disorders or irregularities of rhythm in the heart in waveform of ECG have been considered as evidences for underlying cardiovascular issues like arrhythmias.

The Non-invasive arrhythmia diagnosis has been dependent on 12-lead standard ECG that evaluates potentials of electricity from 10 electrodes by placing them at distinct parts in the surface of a body. Among 10 electrodes, 4 have been placed in the limbs and 6 have been placed on the chest. To provide a productive treatment to the arrhythmias, the diagnosis in the early stage is more significant. Early recognition of definite types of infrequent, transient or short-term arrhythmias needs monitoring of heart electrical activity for a long period of time.

The work [1] presents that, ECG databases are having an open access, which result to a development of several approaches and models for ECG arrhythmia classification of computer aided over former decades. All computer aided ECG-classification model incorporates 4 prominent stages called ECG signal preprocessing stage, detection of heart beat, feature selection and extraction phase and ultimately, the construction of classifier. The ECG signal preprocessing and detection of heartbeat are more prominent and both of them have been extensively

Huge amount of classifiers have been projected for discrimination of arrhythmia. The projected strategies vary from simple form of classifiers such as decision trees as in [2], [3] or Linear discriminants [4] to more sophisticated classifiers like conventional neural networks (NN) [5-8], SVM [5], [9-12], conditional random-fields as in [13] and finally deep learning strategies as in [8], [14-16]. Furthermore, several contributions have been dedicated to identify the optimal features combination, sometimes even introducing complicated signal processing approaches and to choose the optimum subset for arrhythmia categorization as in [17]. In an instance, well-known selections for the input features are extraction of morphological features from amplitudes, regions [9], [10], [18], wavelet transforms as in [5], [6], [19], HOS (higher order-statistics) [2], [3], [5], difficult representations of heartbeat [11] and features related to frequency domain as in [2], [3], [11], [12], [20]. On other dimension, feature selection models like PCA (principal component analysis) [21], PSO (particle swarm optimization) [11], ICA (independent component analysis [21], [19] or GA-BPNN (genetic algorithm-back propagation NNs) have been utilized.

[1] *Research Scholar, Department of Computer Science & Technology, S.K.University, Ananthapuramu – 515003 Email: jsreedharcs@yahoo.com*
[2] *Professor, Department of Computer Science & Technology, S.K.University, Ananthapuramu – 515003. Email: geethanjali.sku@gmail.com*

** Corresponding Author Email: jsreedharcs@yahoo.com*

Regardless of whether these techniques yield improved performance in arrhythmia classification, several of them demand more computing time for classifier optimization. The use of preprocessing procedures or challenging categorization has not been suited for online computing or requires a greater amount of computer resources. Furthermore, in this research, we provide an overall speedy and automated arrhythmias classifier that can be deployed online and effectively evaluate ECG records. With the release of feature extraction criteria, we present a fundamental or basic application based on heart rates, raw signals, and individual lead data that attempts to reduce computation time while obtaining the lowest classification mistakes.

Many of the cardiologists use raw ECG for the purpose of diagnosis. The fastest and easiest model of feature extraction is extracting the sampled points from signal curve of ECG. Nevertheless, one needs to be aware of reality that, quantity of extracted features, which have been utilized for heartbeat characterization might be a load to classification-algorithm. Many of the contributions, for this reason utilize raw signal for performing waveform down sampling some selection of feature to lower the computation time as in [10]. To evade this problem, the simple ML (machine learning) model has been elected for arrhythmias classification. One of the benefits of projected model is that amount of features affects classification speed as the aspects of classifier associated to input remain random and are not optimized. Hence, raw waveform of heart beat might be utilized for classification without negotiating in terms of speed. Moreover, this simple ML model enables rapid retraining of classifier when novel data from ECG become available.

The organization of this paper contains an Introduction related to analytical progression scale for arrhythmia scope prediction from electrocardiograms has been discussed. In section 2, related work and several models for arrhythmia scope prediction from electrocardiograms have been explored. In section 3, methods and materials related to the proposed model have been discussed. In section 4, an experimental study has been carried out, and the proposed model has been compared with other contemporary models. In section 5, the conclusion of this article has been explained, followed by references

2. Related Research

Several researchers utilized ML algorithms, incorporating recurrent NN and multi-task Gaussian procedure approaches for estimating the variability of patient data, onset heart failure, and mortality of patient and further prescriptions. Moreover, their contribution has covered the way through (a) exhibiting the value of ML algorithms in medical domain (b) by showing that definite ML algorithms might be suitable for definite datasets & (c) by recommending that incorporating trend and temporal data enhances the approaches. Their outcomes strongly impacted the design of this contribution and also motivated us to utilize random forest a ML algorithm, which has been mainly applicable to our dataset. Also, work [22] utilized RNN (recurrent NN) approaches for identifying the commencement of heart failure. They utilized data from 3884 cases of heart failure among 28903 cases from 16th may 2000 to 23rd may 2013. Furthermore, RNN approach has been adapted for identifying relations among prescriptions, process, and time-stamped diagnosis for overall instances. They finalized that utilization of time-stamped data enhanced the deep-learning approaches performance for early identification of heart-failure at the time of observation from 12-18 months.

Later, the work [23] proposed Doctor AI. The doctor AI utilized time stamped and RNN electronic health records from 2, 60,000 and among them there are 2128 physicians over 8 years of time for predicting the diagnoses and medications for the following visit relying on administered treatments at the time of earlier visits.

Also, data utilized for Doctor AI has been extracted from MIMIC (medical information mart for intensive care) database. Moreover, they attained a recall of 79.58 and also exhibited adaptability of Doctor AI s by testing the approach on other company database without losing required accuracy. They have mentioned that Doctor AI drawback is that their false positives or inexact diagnoses might be damaging severely towards healthy patients when acted upon and hence Doctor AI must not be utilized without supervision of human.

The contribution [24] utilized MIMIC database for gathering incomplete, heterogeneous, sparse, and noisy and irregularly sampled clinical data, by incorporating both clinical notes and physiological signals. They utilized multi-task Gaussian procedure approaches for estimating and evaluating the acuity of patient that is measuring the nursing care intensity needed by a patient. Most of acuity ranks depends on patient's information and could not include clinical developed data such as lab values, events of chart or doctors' notes. Furthermore, they endeavored for refining the acuity patients score by incorporating developing clinical data into estimation. Initially, they endeavored to predict the reactivity of cerebrovascular pressure that often signifies the scope of secondary brain loss like altered cerebral flow of blood or cerebral edema in brain traumatic injury patients. Also, they endeavored to utilize clinical progressing notes for estimating the mortality of patient.

The work [25] presents that how accumulation of patient's active sign trends for patients preceding measured momentary critical signs enlarged the predictive power approach. Moreover, they utilized the information on critical sign trends from 5 of the hospitals over the span of 5 years in order to predict the transfer of hospital, death

and cardiac arrest. Ultimately, they defined that including the trends improved the accuracy while compared with approach comprising only momentary critical signs AUC of 0.78 vs 0.74.

The study [26] projected 3-layer DGEC (deep genetic ensemble classifiers) for identification of cardiac arrhythmia by utilizing ECG signal. Furthermore, 744 segments of imbalanced data from the 29 people ECG signal have been used, which are achieved from MIT-BIH arrhythmia database. 99.37% of accuracy has been achieved by this model with 0.8736 single samples classification time in identification of entire 17 kinds of arrhythmias.

The work [27] proposed minimal complex and simple novel hexadecimal-ternary-pattern model for cardiac arrhythmia identification automatically. Moreover, they used multi-level feature extraction wavelet for 17 kinds of ECG arrhythmia & attained 95% of classification accuracy on signal of ECG attained from database MIT-BIH. The study [28] presents a nonlinear morphological feature based automatic heartbeat classification and higher voting-based strategy known as ICEEMED, by utilizing arrhythmia database called MIT-BIH. In this approach, they extracted significant ECG signals features, where they utilized for classifying distinct arrhythmia types. Their approach attained 90.4% of classification accuracy & 100% on unknown and fusion classes. Regardless of these prominent outcomes of this model, some of the heartbeat's performance such as aberrated atrial, junctional premature-beats, supra-ventricular and atrial-premature contraction has been low still and requires to be enhanced when compared with other classes. Also, they carried out that, their approach might be utilized for monitoring the system in real-time in healthcare.

The study [29] presented new evolutionary neural model by utilizing SVM classifier. Moreover, in this contribution, they examined longer ECG signal fragments for categorizing 17 ECG arrhythmia classes on arrhythmia database called MIT-BIH.

The contribution [30] utilized arrhythmia database called MIT-BIH for signal classification of ECG of 17 kinds of arrhythmia by using novel genetic-ensemble classifiers model depending on SVM. Furthermore, their approach attained maximal classification accuracy as 98.99%, sensitivity as 91.40% and specificity as 99.46%. Moreover, it is possible for applying this approach on mobile-devices because of their minimal computational complexity. Nevertheless, because of inadequate appropriate signals in database called MIT-BIH, this approach has been tested with minimal amount of ECG signal fragments such as [29].

The study [31] presented a new technique for effective classification of 17 cardiac arrhythmia types by utilizing 1000 fragments from ECG signals from a database called MIT-BIH for 1 lead, MLII, from 45 patients. Moreover, their approach attained maximal classification, where specificity is 99.93%, accuracy is 98.85%, sensitivity is 90.20% and classification time for 1 sample is 0.0023. Nevertheless, because of inadequate suitable signals in database MIT-BIH, this approach has been examined with minimal amount of ECG-signals fragments identical to projected approaches of [30] [29].

The models in the contributions [32] [33] presented the heartbeats classification as negative or positive. The former study ESCPF stated in [32] depicted regression-heuristics. Nevertheless, the model confines to train only internal-subjects, & accuracy has been offensive when the specified training data is having higher dimensionality. Also, similar confines have occurred in other existing methods, which categorizes the heartbeat called ARTM (Automated real-time model) [33], [34].

3. Materials and Methods

The projected model of detecting the scope of arrhythmia from ECG has been described in this section. Moreover, this section, we have divided into several subsections that covers information regarding the data along with their framework utilized in the projected approach, the adopted approach is to detect the optimum features for training the classifier, which has been elaborated in other subsection

Dataset and Records format

Let dataset ECG represents reports set of ECG of subjects in digital format that have been labeled as either positive or negative. The ECG has been an input corpus should divide into 2 sets E_+, E_- that comprises the records as positive as well as negative labeled respectively. The order or ECG elements should be divided into tuples of n size. Every tuple t of each ECG depicts the order of elements of n size of corresponding electrocardiogram. Each electrocardiogram reflects the set of tuples of size n.

Further, for each label $\{positive, negative\}$ for each electrocardiogram $\{ecg_i^+, ecg_i^-\}$, for each tuple $\left\{\left\{t_j^+ \exists t_j^+ \in ecg_i^+, ecg_i^+ \in E_+\right\}, \left\{t_j^- \exists t_j^- \in ecg_i^-, ecg_i^- \in E_-\right\}\right\}$, find the status-indicators $\{o, h, l, e\}$ representing the "open", "high", "low", and "end" value of the corresponding tuple in respective order.

Moving averages of the status-indicators of the given electrocardiogram signals of both positive and negative classes will be calculated using the obtained values of each status-indicator of each tuple of each electrocardiogram, where coefficient represents the sequence of the values considered to assess moving averages. For each status-indicator of the metric, the moving averages will be evaluated as follows:

These set of status-indicators of each electrocardiogram shall be considered as features to train the model that

predicts the arrhythmia scope from the given unlabeled records.

Further, the resultant moving averages of status-indicators of electrocardiograms of both positive and negative labels shall present as two-dimensional matrices M_+, M_-. Each of these matrices reflects the moving averages of status-indicators of electrocardiograms of positive and negative label in respective order.

Each i^{th} row of these matrices presents the moving averages of status-indicators of the i^{th} record (electrocardiogram) of the set $\{E_+ \square E_-\}$ label represented by corresponding matrix $\{M_+ \square M_-\}$. Each j^{th} column of the matrix represents the j^{th} status indicators' moving averages of the all records (electrocardiograms) of the set $\{E_+ \square E_-\}$ label represented by corresponding matrix $\{M_+ \square M_-\}$. Matrix template is figure 1

Concerning to the proposed model, the features are the status indicators' moving averages (open, high, low, and end) of the n-tuples of the input electrocardiograms of both positive and negative labels. The status indicators of i^{th} tuple of all electrocardiograms of both positive label and negative label shall be considered as optimal if and only if the corresponding status indicators of positive label and negative label reflects distribution diversity.

$find-moving-averages(sM)$ Begin //discovers moving averages for all status-indicators sM of electrocardiogram signal ecg

$$ma_{sM}^{o} = \overset{(|sM|-mac)}{\underset{p=1}{\forall}} \left\{ \sum_{j=i}^{mac} \left\{ sM\left(o_j\right) \right\} \right\}$$ // status-indicator open-value's Moving average of the electrocardiogram

$$ma_{sM}^{h} = \overset{(|sM|-mac)}{\underset{p=1}{\forall}} \left\{ \sum_{j=i}^{mac} \left\{ sM\left(h_j\right) \right\} \right\}$$ // status-indicator high-value's Moving average of the electrocardiogram

$$ma_{sM}^{l} = \overset{(|sM|-mac)}{\underset{p=1}{\forall}} \left\{ \sum_{j=i}^{mac} \left\{ sM\left(l_j\right) \right\} \right\}$$ // status-indicator low-value's moving average of the electrocardiogram

$$ma_{sM}^{e} = \overset{(|sM|-mac)}{\underset{p=1}{\forall}} \left\{ \sum_{j=i}^{mac} \left\{ sM\left(e_j\right) \right\} \right\}$$ // status-indicator end-value's moving average of the electrocardiogram

End

$ma_{SM} \leftarrow \left\{ ma_{sM}^{o}, ma_{sM}^{h}, ma_{sM}^{l}, ma_{sM}^{e} \right\}$ // moving average coefficients of status-indicators

Return ma_{sM}

End

The proposed model discovers the optimal features, which are the status-indicators of the column having diversity between matrices M_+, M_-. The j^{th} column of the matrices is said to be optimal, if and only if the j^{th} column of the matrix M_+ and the j^{th} column of the matrix M_- shall have diversity in distribution. The proposed approach discovers optimal features (status indicators) of the both labels using distribution diversity measure called Wilcoxon Rank Sum Test (WRS-Test)

	col_1	col_2	.	.	.	col_j	col_{j+1}	.	.	.	col_n
ecg_1	$\{o,h,l,e\}_{1X1}$	$\{o,h,l,e\}_{1X2}$	.	.	.	$\{o,h,l,e\}_{1Xj}$	$\{o,h,l,e\}_{1X(j+1)}$	.	.	.	$\{o,h,l,e\}_{1Xn}$
ecg_2	$\{o,h,l,e\}_{2X1}$	$\{o,h,l,e\}_{2X2}$	.	.	.	$\{o,h,l,e\}_{2Xj}$	$\{o,h,l,e\}_{2X(j+1)}$	.	.	.	$\{o,h,l,e\}_{2Xn}$
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
ecg_i	$\{o,h,l,e\}_{iX1}$	$\{o,h,l,e\}_{iX2}$	.	.		$\{o,h,l,e\}_{iXj}$	$\{o,h,l,e\}_{iX(j+1)}$	.	.		$\{o,h,l,e\}_{iXn}$
ecg_{i+1}	$\{o,h,l,e\}_{(i+1)X1}$	$\{o,h,l,e\}_{(i+1)X2}$	.	.	.	$\{o,h,l,e\}_{(i+1)Xj}$	$\{o,h,l,e\}_{(i+1)X(j+1)}$	.	.		$\{o,h,l,e\}_{(i+1)Xn}$
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
ecg_m	$\{o,h,l,e\}_{mX1}$	$\{o,h,l,e\}_{mX2}$	.	.	.	$\{o,h,l,e\}_{mXj}$	$\{o,h,l,e\}_{mX(j+1)}$				$\{o,h,l,e\}_{mXn}$

Figure 1- Two-dimensional matrix representation of the status-indicators

Feature Optimization by distribution diversity measure

The distribution state diversity of the specified 2 datasets might represent through a distance metric known as WRS-test (Wilcoxon rank sum) as stated in [34]. This test does not need information regarding the distribution type of the data, which is a conventional need of the data stream concepts. The description of the WRS-Test implementation process is as follows:

Let the 2 vectors values be $\boldsymbol{v_1, v_2}$. Also, the above stated test might implement in order to evaluate these 2 vectors distributions are same or different in the following way:Initially, all the entries of vectors $\boldsymbol{v_1, v_2}$ are moved to new vector $\boldsymbol{v}$. Further, sort the vector $\boldsymbol{v}$ in ascending order of the values and let the indices of the ordered values of the vector $\boldsymbol{v}$ as corresponding ranks $\boldsymbol{R}$ .The average of the indices assigned to the identical values will be the rank of all the respective identical values. Further description denotes the ranks assigned to the values of the vector $\boldsymbol{v_1}$ as set $\boldsymbol{R_1}$ and the ranks assigned to the values in vector $\boldsymbol{v_2}$ as set $\boldsymbol{R_2}$. Later the process finds the aggregate of the entries in set $\boldsymbol{R_1}$ as $\boldsymbol{RS_1}$, which is further used to determine the rank sum threshold $\boldsymbol{RST_1}$ of the vector $\boldsymbol{v_1}$ as follows:

$\boldsymbol{wrs-test(v_1, v_2)}$ Begin	The function of Wilcoxon rank sum test
$\boldsymbol{RST_1 = RS_1 - \frac{\lvert v_1 \rvert \times (\lvert v_1 \rvert + 1)}{2}}$	// the notation $\lvert v_1 \rvert$ denotes the size of the vector $\boldsymbol{v_1}$.

Similarly, the rank sum threshold $\boldsymbol{RST_2}$ of the vector $\boldsymbol{v_2}$ will be determined as follows

$\boldsymbol{RST_2 = RS_2 - \frac{\lvert v_2 \rvert \times (\lvert v_2 \rvert + 1)}{2}}$	// the notation $\lvert v_2 \rvert$ denotes the size of the vector $\boldsymbol{v_2}$, and the notation $\boldsymbol{RS_2}$ denotes the sum of the ranks of the entries in vector $\boldsymbol{v_2}$ those listed in set $\boldsymbol{R_2}$.
$\boldsymbol{RST = RST_1 + RST_2}$	Then the rank sum threshold $\boldsymbol{RST}$ of the entries in both vectors $\boldsymbol{v_1, v_2}$ is the sum of rank sum thresholds $\boldsymbol{RST_1, RST_2}$ of the vectors $\boldsymbol{v_1, v_2}$.

Then find the z-score [35]:

$\boldsymbol{m_{RST} = \frac{RST}{2}}$	Initially, find the mean $\boldsymbol{m_{RST}}$

$$d_{RST} = \left(\sqrt{\frac{|v_1|*|v_2|*(|v|+1)}{|v|}} - \sqrt{\frac{|v_1|*|v_2|}{|v|}\left((|v|+1) - \sum_{i=1}^{k} \frac{t_i^3 - t_i}{|v|*(|v|-1)}\right)} \right)$$

//find and standard deviation d_{RST}, the notation k denotes the number of distinct ranks, t_i denotes the number of entries sharing the same rank i

$$z = \frac{RST - m_{RST}}{d_{RST}}$$

z-core assessed

$if\ ((pValue > p\tau)$ return true $else$ return false

Then find the p-value $pValue$ of the depicted z score in z-table [36]. If the p-value found to be greater than the given probability threshold (usually 0.01, 0.05, or 0.1) then the distribution of the vectors v_1, v_2 found to be diversified, else the distribution is similar.

Analytical Progression Scale

The analytical progression scale that proposed to predict the arrhythmia scope in electrocardiogram is detailed in this section. The APS, initially partitions the given training dataset ECG of electrocardiograms having positive label as one set E_+ and the electrocardiograms labeled as negative as other set. Further, for each dataset $\{E_+ \| E_-\}$, for each electrocardiogram $\{ecg_+ \| ecg_-\}$, partitions in to set of tuples, and finds status indicators open, high, low, and end $\{o,h,l,e\}$ for each tuple $\{t \exists t \in eT\}$ of the set eT contains all the partitioned tuples of the corresponding electrocardiogram $\{ecg_+ \| ecg_-\}$. Further, for each electrocardiogram $\{ecg_+ \| ecg_-\}$ discovers the status indicators' moving averages of the corresponding partitioned tuples eT. Afterwards, the resultant status indicators' moving averages discovered from the electrocardiograms of the set E_+ shall be projected as matrix M_+ with two dimensions, and projects a two dimensional matrix M_- represents the status indicators' moving averages discovered from the electrocardiogram signals of the set E_-. These projected matrices M_+, M_- shall be used further as input to determine the optimal features using distributed diversity measuring method WRS-Test. The resultant columns of the matrix, those representing optimal features of both labels shall be used further to derive analytical progression scale. The algorithmic model of the APS follows.

//for electrocardiograms with positive label//

$\overset{|E_+|}{\underset{i=1}{\forall}} \{e \exists e \in E_+\}$ Begin // for each electrocardiogram e of the set E_+

$$eT \leftarrow \overset{|e|}{\underset{j=1}{\forall}} \left\{ \overset{[j+n]}{\underset{x=j}{\forall}} \{et \leftarrow y_x\} \wedge \{j = j+n\} \right\}$$

//partitioning the y-coordinates of the given electrocardiogram e in to multiple tuples

$$sM \leftarrow \overset{|eT|}{\underset{j=1}{\forall}} \{\{o,h,l,e\} \exists \{o,h,l,e\} \in t_j\}$$

//collects status-indicators $\{o,h,l,e\}$ of each tuple t_j of the given electrocardiogram e

$$M_+[i] \leftarrow find - moving - averages(sM)$$

End

//for electrocardiograms of the negative label//

$\overset{|E_-|}{\underset{i=1}{\forall}} \{e \exists e \in E_-\}$ Begin // for each electrocardiogram e of the set E_-

$$eT \leftarrow \overset{|e|}{\underset{j=1}{\forall}} \left\{ \overset{[j+n]}{\underset{x=j}{\forall}} \{et \leftarrow y_x\} \wedge \{j = j+n\} \right\}$$

//partitioning the y-coordinates of the given electrocardiogram e in to multiple tuples

$$sM \leftarrow \overset{|eT|}{\underset{j=1}{\forall}} \{\{o,h,l,e\} \exists \{o,h,l,e\} \in t_j\}$$

//collects status-indicators $\{o,h,l,e\}$ of each tuple t_j of the given electrocardiogram e

$$M_-[i] \leftarrow find - moving - averages(sM)$$

End

// optimal feature selection//

$foreach\{j = 1,2,3,...,n\}$ Begin //

$c_+ \leftarrow \overset{m}{\underset{i=1}{\forall}} \{M_+^{[i,j]}\}$ // collecting all the values of the j^{th} column of the matrix M_+

$c_- \leftarrow \overset{m}{\underset{i=1}{\forall}} \{M_-^{[i,j]}\}$ // collecting all the values of the j^{th} column of the matrix M_-

$v_+ \leftarrow \overset{|c_+|}{\underset{k=1}{\forall}} \left\{ \left(o_k + h_k + l_k + e_k \right) * 4^{-1} \right\}$ //finding the average of each set of status indicators listed in c_+

$v_- \leftarrow \overset{|c_-|}{\underset{k=1}{\forall}} \left\{ \left(o_k + h_k + l_k + e_k \right) * 4^{-1} \right\}$ //finding the average of each set of status indicators listed in c_-

$if\ (wrs - test(v_+, v_-))$ Begin // find the diversity between two vectors v_+, v_- is true or false

$oF_+ \leftarrow v_+$ // preparing a matrix oF_+ representing the optimal features of the positive label

$oF_- \leftarrow v_-$ // preparing a matrix oF_- representing the optimal features of the negative label

End

// discovering analytical progression scale measures for negative label//

$foreach\{ j = 1, 2, 3, ..., n \}$ Begin //

$\mu_-^j = \frac{1}{m} \sum_{i=1}^{m} oF_-^{[i,j]}$ // finding the mean μ_-^j of the j^{th} column of the matrix oF_-

$\delta_-^j = \frac{\sum_{j=1}^{m} \left[\sqrt{\mu_-^j - oF_-^{[i,j]}} \right]^2}{m}$ // finding the deviation δ_-^j of the j^{th} column of the matrix oF_-

$\left. \begin{aligned} l_-^j &= \mu_-^j - \delta_-^j \\ u_-^j &= \mu_-^j + \delta_-^j \end{aligned} \right\}$ // finding lower and upper bound of the j^{th} column of the matrix oF_-

End

// discovering analytical progression scale measures for positive label//

$foreach\{ j = 1, 2, 3, ..., n \}$ Begin //

$\mu_+^j = \frac{1}{m} \sum_{i=1}^{m} oF_+^{[i,j]}$ // finding the mean μ_+^j of the j^{th} column of the matrix oF_+

$\delta_+^j = \frac{\sum_{j=1}^{m} \left[\sqrt{\mu_+^j - oF_+^{[i,j]}} \right]^2}{m}$ // finding the deviation δ_+^j of the j^{th} column of the matrix oF_+

$\left. \begin{aligned} l_+^j &= \mu_+^j - \delta_+^j \\ u_+^j &= \mu_+^j + \delta_+^j \end{aligned} \right\}$ // finding lower and upper bound of the j^{th} column of the matrix oF_+

End

Label prediction by APS

For a given electrocardiogram ecg, the label shall be predicted using APS as follows

$eT \leftarrow \overset{|ecg|}{\underset{j=1}{\forall}} \left\{ \overset{[j+n]}{\underset{x=j}{\forall}} \{ et \leftarrow y_x \} \wedge \{ j = j + n \} \right\}$

//partitioning the y-coordinates of the given electrocardiogram ecg in to multiple tuples

$sM \leftarrow \overset{|eT|}{\underset{j=1}{\forall}} \left\{ \{o,h,l,e\} \exists \{o,h,l,e\} \in t_j \right\}$

//collects status-indicators $\{o,h,l,e\}$ of each tuple t_j of the given electrocardiogram ecg

$ma(o,h,l,e) \leftarrow find - moving - averages(sM)$

$\mu_{ma}^{ecg} = \frac{(o + h + l + e)}{4}$ //finding average of the status-indicators of the electrocardiogram ecg

Find the weight of the mean μ_{ma}^{ecg} towards positive and negative label as follows

$aps_+^{ecg} = \frac{1}{n} \sum_{j=1}^{n} \left\{ 1 \exists l_+^j \le \mu_{ma}^j \le u_+^j \right\}$ // finding weight aps_+^{ecg} of the mean μ_{ma}^{ecg} towards positive label

$aps_-^{ecg} = \frac{1}{n} \sum_{j=1}^{n} \left\{ 1 \exists l_-^j \le \mu_{ma}^j \le u_-^j \right\}$ // finding weight aps_-^{ecg} of the mean μ_{ma}^{ecg} towards negative label

$if\ \left(aps_+^{ecg} \ge aps_-^{ecg} \right)$ Label the given electrocardiogram as positive (prone to arrhythmia)

$else$ Label the given electrocardiogram as negative (benign)

4. Experimental study

The focus is on examining the effectiveness of projected approach as well as other contemporary relative trend approaches that have been utilized in standard datasets. Also, the prominence of the approaches has been measured significantly by concentrating on outcomes noticed for the important metrics such as Sensitivity, F-measure, specificity, accuracy, matthews correlation coefficient (MCC) and precision. The suggested model APS has been compared with other contemporary models ARTM [33], [34], [35] and ESCPF [32], [36] in order to forecast the arrhythmia scope in ECG.

The dataset

In this, the dataset used in the simulation is ECG heart beat classification dataset (EHCD) [37], [38], [39], [40], [41] which have been significantly utilized in existing research contributions for assuring the responsibility of experimental analysis of projected and other existing approaches. Moreover, the digital depiction of ECG

waveforms, which have been gathered from database EHCD is bee utilized for examining the proposed and existing models' performance.

Table 1: standard deviation mean values of the performance metrics has been listed

METRICS	APS	ESCPF	ARTM
ACCURACY	0.97±0.0033	0.92±0.0055	0.88±0.0104
FMEASURE	0.97±0.0038	0.95±0.0029	0.91±0.0070
MCC	0.92±0.0077	0.82±0.0112	0.73±0.0221
PRECISION	0.98±0.0020	0.97±0.0016	0.95±0.0041
SENSITIVITY	0.97±0.0043	0.92±0.0080	0.88±0.0121
SPECIFICITY	0.97±0.0056	0.93±0.0041	0.88±0.0095

Precision

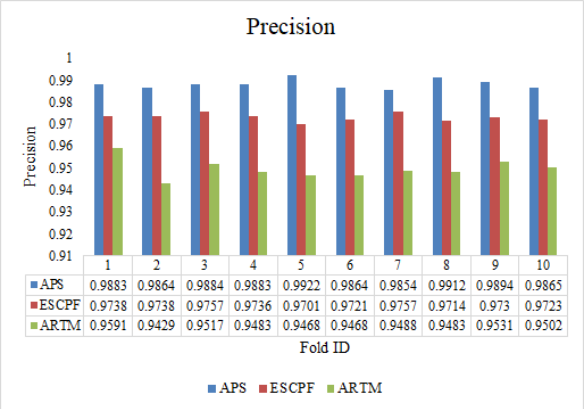

Figure 2- Value of precision perceived for the suggested model APS and other contemporary models ARTM and ESCPF over the 10 folds

The precision metric reflects on diversified rations noticed as positive for the cumulative false records set that have been signified as positive result. Graph is depicted for tenfold cross validation values of the metric precision estimated for the suggested model APS and comparison contemporary models ARTM and ESCPF as shown in figure 2. The average weight of precision of the APS, ARTM, and ESCPF are 0.98±0.0020, 0.95±0.0041 and 0.97±0.0016 respectively. It has been finally concluded that, the performance of APS in terms of precision is more effective and added an advantage while compared to other contemporary models used in the study ARTM and ESCPF.

Specificity

The specificity metric has been used in order measure the performance of suggested model and contemporary models of this contribution. Specificity is defined as ratio of correctly selected labeled records that are negative when compared to actual number of negative records specified as input in phase of testing. Graph is depicted for tenfold cross validation values of the metric specificity estimated for the suggested model APS and comparison contemporary models ARTM and ESCPF as shown in figure 3. The average standard means deviation of specificity for the APS, ARTM, and ESCPF are 0.97±0.0056, 0.88±0.0095 and 0.93±0.0041 respectively. It has been finally concluded that, the performance of APS in terms of specificity added an advantage while compared to other contemporary models used in the study ARTM and ESCPF.

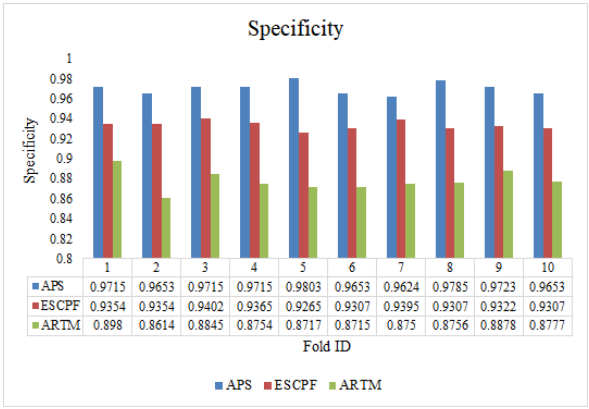

Figure 3- Value of specificity perceived for the suggested model APS and other contemporary models ARTM and ESCPF over the 10 folds

Sensitivity

Graph is depicted for tenfold cross validation values of the metric sensitivity estimated for the suggested model APS and comparison contemporary models ARTM and ESCPF as shown in figure 4. The average standard means deviation of specificity for the APS, ARTM, and ESCPF are 0.97±0.0043, 0.88±0.0121 and 0.92±0.0080 respectively. From the statistics, it has been concluded that, the performance of APS in terms of sensitivity is more superior while compared to other contemporary models used in the study ARTM and ESCPF.

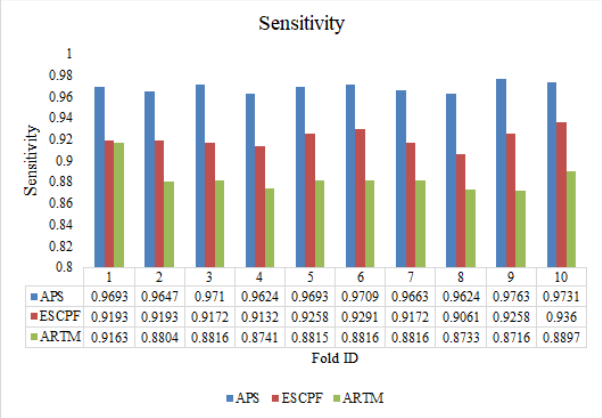

Figure 4- Value of sensitivity perceived for the suggested model APS and other contemporary models ARTM and ESCPF over the 10 folds

Accuracy

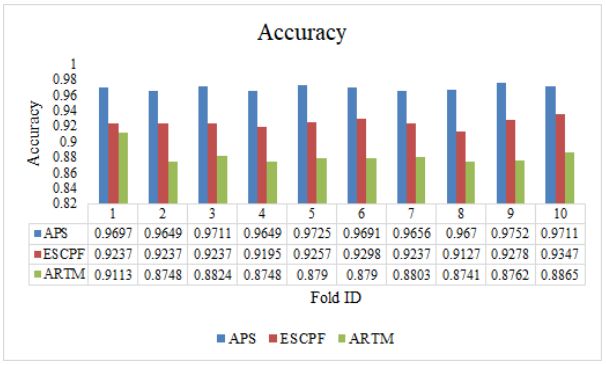

Figure 5: Value of accuracy perceived for the suggested model APS and other contemporary models ARTM and ESCPF over the 10 folds

Graph is depicted for tenfold cross validation values of the metric accuracy estimated for the suggested model APS and comparison contemporary models ARTM and ESCPF as shown in figure 5. The average standard means deviation of accuracy for the APS, ARTM, and ESCPF are 0.97±0.0033, 0.88±0.0104 and 0.92±0.0055 respectively. From the statistics, it has been concluded that, the performance of APS in terms of accuracy is more superior while compared to other contemporary models used in the study ARTM and ESCPF.

F-Measure

The metric F-measure reflects the relation among recall and precision considered for the system accuracy. Graph is depicted for tenfold cross validation values of the metric F-measure estimated for the suggested model APS and comparison contemporary models ARTM and ESCPF as shown in figure 6. The average standard means deviation of F-measure for the APS, ARTM, and ESCPF are 0.97±0.0038, 0.91±0.0070 and 0.95±0.0029 respectively. From the statistics, it has been concluded that, the performance of APS in terms of F-measure is more effective while compared to other contemporary models used in the study ARTM and ESCPF.

Figure 6: Values of F-measure perceived for the suggested model APS and other contemporary models ARTM and ESCPF over the 10 folds

MCC

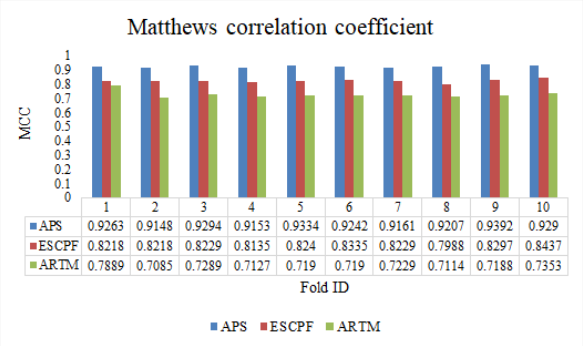

Figure 7- Values of MCC perceived for the suggested model APS and other contemporary models ARTM and ESCPF over the 10 folds

The MCC metric considers often for evaluating the quality all over the binary classifications that has been utilized for measuring the overall 3 approaches performance. Graph is depicted for tenfold cross validation values of the metric MCC estimated for the suggested model APS and comparison contemporary models ARTM and ESCPF as shown in figure 7. The average standard mean deviation of MCC for the APS, ARTM, and ESCPF are 0.92±0.0077, 0.73±0.0221 and 0.82±0.0112 respectively. From the statistics, it has been concluded that, the performance of APS in terms of MCC is higher marginally while compared to other contemporary models used in the study ARTM and ESCPF.

5. Conclusion

Reducing false alarming in arrhythmia prediction from electrocardiograms using machine learning is the objective of this contribution. Concerning this, a novel Analytical progression Scale (APS) has been proposed in this manuscript. The given labeled electrocardiograms have been used to define the scales those reflect the prediction of arrhythmia scope in given electrocardiogram. Moving averages of the open, low, close, and end of each tuple of the electrocardiogram shall be taken as features, which represents in matrix format. Further discovers optimal features to by using diversity measures. Further derives Analytical progression scales for both positive and negative label. The results obtained for statistical assessment metrics of the 10-fold cross validation have

compared with the other contemporary methods, which concluding that the proposed model outperforming the contemporary models with minimal false alarming. The future research shall consider the pattern of features introduced in this manuscript to define a appropriate fitness method for soft computing models

References

[1] Goldberger, A. L., G. B. Moody, and M. D. Costa. "Physionet." Nonlinear dynamics, fractals, and chaos theory: implications for neuroanatomic heart rate control in health and disease. http://www. physionet. org/tutorials/ndc (1999).

[2] Qurraie, Safa Sultan, and Rashid Ghorbani Afkhami. "ECG arrhythmia classification using time frequency distribution techniques." Biomedical engineering letters 7.4 (2017): 325-332.

[3] Afkhami, Rashid Ghorbani, Ghanbar Azarnia, and Mohammad Ali Tinati. "Cardiac arrhythmia classification using statistical and mixture modeling features of ECG signals." Pattern Recognition Letters 70 (2016): 45-51.

[4] Krasteva, Vessela, et al. "Superiority of classification tree versus cluster, fuzzy and discriminant models in a heartbeat classification system." PloS one 10.10 (2015): e0140123.

[5] Boyapati, B. ., and J. . Kumar. “Parasitic Element Based Frequency Reconfigurable Antenna With Dual Wideband Characteristics for Wireless Applications”. International Journal on Recent and Innovation Trends in Computing and Communication, vol. 10, no. 6, June 2022, pp. 10-23, doi:10.17762/ijritcc.v10i6.5619.

[6] Elhaj, Fatin A., et al. "Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals." Computer methods and programs in biomedicine 127 (2016): 52-63.

[7] Martis, Roshan Joy, U. Rajendra Acharya, and Lim Choo Min. "ECG beat classification using PCA, LDA, ICA and discrete wavelet transform." Biomedical Signal Processing and Control 8.5 (2013): 437-448.

[8] Xiao, Xuan, Zhi-Cheng Wu, and Kuo-Chen Chou. "A multi-label classifier for predicting the subcellular localization of gram-negative bacterial proteins with both single and multiple sites." PloS one 6.6 (2011): e20592.

[9] Kiranyaz, Serkan, Turker Ince, and Moncef Gabbouj. "Real-time patient-specific ECG classification by 1-D convolutional neural networks." IEEE Transactions on Biomedical Engineering 63.3 (2015): 664-675.

[10] Zhang, Zhancheng, et al. "Heartbeat classification using disease-specific feature selection." Computers in biology and medicine 46 (2014): 79-89.

[11] Huang, Huifang, et al. "A new hierarchical method for inter-patient heartbeat classification using random projections and RR intervals." Biomedical engineering online 13.1 (2014): 1-26.

[12] Garcia, G., et al. "Inter-patient ECG heartbeat classification with temporal VCG optimized by PSO. Sci. Rep. 7, 10543–10544 (2017)."

[13] Raj, Sandeep, Kailash Chandra Ray, and Om Shankar. "Cardiac arrhythmia beat classification using DOST and PSO tuned SVM." Computer methods and programs in biomedicine 136 (2016): 163-177.

[14] De Lannoy, Gaël, et al. "Weighted conditional random fields for supervised interpatient heartbeat classification." IEEE Transactions on Biomedical Engineering 59.1 (2011): 241-247.

[15] Wu, Zhiyong, Xiangqian Ding, and Guangrui Zhang. "A novel method for classification of ECG arrhythmias using deep belief networks." International Journal of Computational Intelligence and Applications 15.04 (2016): 1650021.

[16] Acharya, U. Rajendra, et al. "A deep convolutional neural network model to classify heartbeats." Computers in biology and medicine 89 (2017): 389-396.

[17] Al Rahhal, Mohamad Mahmoud, et al. "Deep learning approach for active classification of electrocardiogram signals." Information Sciences 345 (2016): 340-354.

[18] Li, Hongqiang, et al. "Genetic algorithm for the optimization of features and neural networks in ECG signals classification." Scientific reports 7.1 (2017): 1-12.

[19] Gunjan, Vinit Kumar, and Madapuri Rudra Kumar. "Predictive Analytics for OSA Detection Using Non-Conventional Metrics." International Journal of Knowledge-Based Organizations (IJKBO) 10.4 (2020): 13-23.

[20] Ye, Can, BVK Vijaya Kumar, and Miguel Tavares Coimbra. "An automatic subject-adaptable heartbeat classifier based on multiview learning." IEEE journal of biomedical and health informatics 20.6 (2015): 1485-1492.

[21] Zidelmal, Zahia, et al. "ECG beat classification using a cost sensitive classifier." Computer methods and programs in biomedicine 111.3 (2013): 570-577.

[22] Gupta, D. J. . (2022). A Study on Various Cloud Computing Technologies, Implementation Process, Categories and Application Use in Organisation. International Journal on Future Revolution in

Computer Science &Amp; Communication Engineering, 8(1), 09–12. https://doi.org/10.17762/ijfrcsce.v8i1.2064

[23] Ye, Can, BVK Vijaya Kumar, and Miguel Tavares Coimbra. "Heartbeat classification using morphological and dynamic features of ECG signals." IEEE Transactions on Biomedical Engineering 59.10 (2012): 2930-2941.

[24] Choi, Edward, et al. "Using recurrent neural network models for early detection of heart failure onset." Journal of the American Medical Informatics Association 24.2 (2017): 361-370.

[25] Choi, Edward, et al. "Doctor ai: Predicting clinical events via recurrent neural networks." Machine learning for healthcare conference. PMLR, 2016, 301-318.

[26] Ghassemi, Marzyeh, et al. "A multivariate timeseries modeling approach to severity of illness assessment and forecasting in ICU with sparse, heterogeneous clinical data." Proceedings of the AAAI conference on artificial intelligence. Vol. 29. No. 1. 2015.

[27] Churpek, Matthew M., Richa Adhikari, and Dana P. Edelson. "The value of vital sign trends for detecting clinical deterioration on the wards." Resuscitation 102 (2016): 1-5.

[28] Pławiak, Paweł, and U. Rajendra Acharya. "Novel deep genetic ensemble of classifiers for arrhythmia detection using ECG signals." Neural Computing and Applications 32.15 (2020): 11137-11161.

[29] Tuncer, Turker, et al. "Automated arrhythmia detection using novel hexadecimal local pattern and multilevel wavelet transform with ECG signals." Knowledge-Based Systems 186 (2019): 104923.

[30] Kandala, Rajesh NVPS, et al. "Towards real-time heartbeat classification: evaluation of nonlinear morphological features and voting method." Sensors 19.23 (2019): 5079.

[31] Pławiak, Paweł, and Moloud Abdar. "Novel methodology for cardiac arrhythmias classification based on long-duration ECG signal fragments analysis." Biomedical signal processing. Springer, Singapore, 2020. 225-272.

[32] Pławiak, Paweł. "Novel genetic ensembles of classifiers applied to myocardium dysfunction recognition based on ECG signals." Swarm and evolutionary computation 39 (2018): 192-208.

[33] Pławiak, Paweł. "Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system." Expert Systems with Applications 92 (2018): 334-349.

[34] Fuad Ali Mohammed Al-Yarimi, Nabil Mohammed Ali Munassar, Fahd Awesabi. Electrocardiogram stream level correlated patterns as features to classify heartbeats for arrhythmia prediction. Data Technologies and Applications Emerald Publishing Limited. 2020.

[35] Ortín González, Silvia, et al. "Automated real-time method for ventricular heartbeat classification." (2019).

[36] Rosemaro, E. . (2022). Understanding the Concept of Entrepreneurship Management and Its Contribution in Organization. International Journal of New Practices in Management and Engineering, 11(01), 24–30. https://doi.org/10.17762/ijnpme.v11i01.159

[37] Madapuri, Rudra Kumar, and P. C. Mahesh. "HBS-CRA: scaling impact of change request towards fault proneness: defining a heuristic and biases scale (HBS) of change request artifacts (CRA)." Cluster Computing 22.5 (2019): 11591-11599.

[38] Rudra Kumar, M., and Vinit Kumar Gunjan. "Machine Learning Based Solutions for Human Resource Systems Management." ICCCE 2021. Springer, Singapore, 2022. 1239-1249

[39] Haynes, Winston. "Wilcoxon rank sum test." Encyclopedia of Systems Biology. Springer New York, 2013. 2354-2355.

[40] Sahoo, Prasanna, and Thomas Riedel. Mean value theorems and functional equations. World Scientific, 1998.

[41] Chalapathi, M. M., et al. "Ensemble Learning by High-Dimensional Acoustic Features for Emotion Recognition from Speech Audio Signal." Security and Communication Networks 2022 (2022).

[42] Kumar, S., Gornale, S. S., Siddalingappa, R., & Mane, A. (2022). Gender Classification Based on Online Signature Features using Machine Learning Techniques. International Journal of Intelligent Systems and Applications in Engineering, 10(2), 260–268. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2020

[43] Rudra Kumar, M., Rashmi Pathak, and Vinit Kumar Gunjan. "Machine Learning-Based Project Resource Allocation Fitment Analysis System (ML-PRAFS)." Computational Intelligence in Machine Learning. Springer, Singapore, 2022. 1-14

[44] http://www.sjsu.edu/faculty/gerstman/StatPrimer/z-table.PDF

[45] https://www.kaggle.com/shayanfazeli/heartbeat.