

Quality Assessment of Ambulatory Electrocardiogram Signals by Noise Detection using Optimal Binary Classification

^{*1}V. Supraja, ²P. Nageswara Rao, ³M. N. Giriprasad

¹Research Scholar, Department of ECE, JNTU College of Engineering, Anantapur, India.

¹vsupraja.phd@gmail.com

²Professor, Vardhaman College of Engineering, Hyderabad, India

²nraop@yahoo.com

³Professor, Department of ECE, JNTU College of Engineering, Anantapur, India

³mahendragiri1960@gmail.com

Abstract: In order to improve the diagnostic capability in Ambulatory Electrocardiogram signal and to reduce the noise signal impacts, there is a need for more robust models in place. In terms of improvising to the existing solutions, this article explores a novel binary classifier that learns from the features optimized by fusion of diversity assessment measures, which performs Quality Assessment of Ambulatory Electrocardiogram Signals (QAAES) by Noise Detection. The performance of the proposed model QAAES has been scaled by comparing it with contemporary models. Concerning performance analysis, the 10-fold cross-validation has been carried on a benchmark dataset. The results obtained from experiments carried on proposed and other contemporary models for cross-validation metrics have been compared to signify the sensitivity, specificity, and noise detection accuracy.

Keywords: Noise Detection, Optimal Binary Classification, ECG, IoT etc.

1 Introduction

The ECG (electrocardiogram) has been a basic device in cardiac health screening programs. In this approach, cost-effectiveness has been exhibited in overall age groups such as neonates, adults as well as young athletes [1]. Nevertheless, because of this test shortness probably, definite abnormalities of heart rhythm cannot be detected or identified. It is because of deficiency of continuous measurement started a shift towards monitoring devices or equipment. Moreover, these devices enable monitoring over a long period, hence improves the likelihood of abnormalities detection.

Because of these devices' ambulatory nature, loose electrodes, motion artifacts as well as interference from other types of electric devices might cause signal violations or distortions. These distortions are indicated as artifacts, which might lower the monitoring device diagnostic capabilities and make a wrong decision in treatment.

Divergent technologies have been projected for identifying artifacts and improve the quality of the signal. One of such models has been ICA, which is a separation of blind sources [2]. Further, this model performs on the entire signal and leads to an improvement in ECG signal quality.

Nevertheless, none of the information on the location of the artifact might be extracted through this approach. A commonly used approach for overcoming this issue is to divide signals. The classification step and feature extraction have conventionally followed it. The statistical and spectral information have been used often as features in [3], [4].

Moreover, different ML (machine learning) models have been used for segment classification. This technique downside is required for the gold benchmark. The novel approach has been projected for detecting the location of the artifact automatically in ECG recording for the long term. Also, this model begins by dividing the signal of ECG and characterizing every segment through their ACF (autocorrelation function). Features have been derived and are fed as RUS-Boost algorithm towards classification. The selection for ACF has been motivated because it considers the advantage of ECG signal repetitiveness [5].

Often, ECG signals have been corrupted with distinct types of artifacts, and noises like drift and baseline wander muscle artifacts and powerline interference, making it impossible almost for performing RR interval and morphological analysis of such contaminated signals of ECG as in [6-8]. Many of the current ECG analysis models have

been designed for handling comparatively ECG signals that are free from noise [6]. Current systems in such scenarios render unreliable and inaccurate measurements that result in generating maximal false alarm rates for ECG noisy signals as in [8]. Subsequently, common false alarms have been not only disturbing and annoying both patients and clinicians however also result in cardiac arrhythmias misdiagnosis [8-12]. Moreover, problem-related to alarm rates of heart rate and maximal false arrhythmia impacts the usage of ECG due to 2 reasons: (a) artifacts and noise in ECG signal isoelectric line are identified incorrectly as abnormal or normal beats, (b) severe contamination of ECG beats have been misclassified because of imprecise measurements of required ECG feature aspects [9-12]. Hence, ECG noisy signals shall be either filtered or discarded before feature aspects extraction.

ECG signals have been extensively acquired for several applications like arrhythmias detection, sleep apnea identification, abrupt cardiac arrest estimations, physical and emotional activity recognition, or identification systems as in [13]-[18]. Overall, these application systems of ECG demand highly the correct fiducial point's determination of ECG signal for reliable and accurate morphological features measurements like U-wave, QRS-complex, P-wave and T-wave and finally interval features as in [16-18]. Many ECG analysis models need comparatively ECG signals free from noise in order to attain the measurements of ECG more reliably as well as precisely [18]. Often, ECG signals have been corrupted with distinct noises types like baseline wander, the noise of instrumentation and making it impossible almost for performing morphological investigation of such contaminated beats of ECG as in [18-21], [6], [7]. Hence, automatic measurement of ECG signal has been demanded highly by lowering false alarms because of unacceptable noise levels.

In this paper, electrocardiogram introduction and several other models has been discussed in section 1. In section 2, related work in terms of electrocardiogram for detecting the noise has been explored and various literatures have been reviewed and presented. Section 3 contains methods and materials. Section 4 contains experimental study, where the performance of the proposed model and contemporary model are compared using several metrics. Section 5 is the conclusion followed by references.

2 Related work

Many of the IoT-enabled devices in real-time applications perform on restricted power batteries over a long period. The consumption of energy becomes critical design consideration for improving the overall lifespan of the network [22]. Hence,

data exchange controlling might lower the expenditure and enhances the usage of the network. The work [22] researched the possible data-driven triggering in real-time scenarios like tracing public transport. The research exhibited that even data-driven triggering can enhance the usage of the network and effectiveness of the network. The long-term cardiac continuous health monitoring system demands maximum battery-power for ECG signals transmission and enhances the cost of treatment and bandwidth and traffic load diagnostic server. Also, event triggering has been a favorable term, which can address the above-stated significant challenges adequately.

Moreover, in our former contributions, the automated less complexity robust event cardiac change identification model has been projected for health monitoring applications long-term [23]. The review exhibited that a lifetime of battery for IoT-enabled devices has been impacted highly by power consumption. Hence, intelligent solutions have been demanded to enhance the IoT device's battery life and to lower the usage costs of bandwidth and diagnostic traffic. Subsequently, muscle artifacts decrease the quality of the signal, resolution of the spectrum, and outcomes in higher amplitudes. For instance, small local amplitude waves (U, T, P) have been obscured by higher amplitudes of muscle artifacts. Hence, it has been complexing for physicians to observe and to locate the non-existence or existence of these minimal amplitude waves, which might offer prominent clinical information needed for diagnosing definite abnormalities of the heart. Furthermore, it has been intricate for performing higher reliable and accurate morphological aspects measurements like polarity, timings, amplitude, shape, and interval pause for the local-waves, which are significant for accurate monitoring of heart rate, recognition of heartbeat pattern, and arrhythmia detection as in [24].

The work [25] examined the MSEnt (multiscale entropy) performance to measure the quality of the ECG signal. Moreover, the outcomes exhibited that MSEnt for the clean ECG and LF noise monotonously enhance with the increment in the scale factor. Also, MSEnt for noise PL (periodic signals) shows variation. The model has been assessed on MITBIHNS and MITBIHA databases. Outcomes exhibited that MSEnt has been sensitive towards noise-level within ECG signals and hence depicted valuable device for ECG signal measurement quality. Also, this contribution has primarily eradicated the baseline wander from original signals of ECG due to baseline wander comprised in ECGs might result in imprecise outcomes and later designed novel real-time signals of ECG.

The work [26] presented a model for quality estimation of ECG signal under divergent activities and

postures. The model has been developed depending on features, and PCA like complete range among 0.2 & 15mV, power ratio in the range of frequency is 5-20Hz towards overall 0-62.5Hz, correlation towards averaged beat, correlation towards other leads. Also, minimal distributed ECG segment in leads has been detected automatically and utilized for reconstruction of overall results in such segment. Also, this model has been measured by utilizing 12 lead ECGs for 4 hours comprising divergent activities incorporating controlling breathing for 36 min and exercise for 96 min from 6 healthy related subjects. Moreover, estimation of confidence has been changed from 0.58-0.90 for major movement towards remaining recordings of ECG in respective order.

The work [27] presented an enhanced quality of ECG signal using diversity and correlation-based models. Intolerable values have been derived from criteria based on scores incorporating the following: saturation, flat line; stable voltage superior towards 2mV; channel off and baseline drift; drifts over 2.5mV, minimal amplitude as maximal amplitude greater than 125 μ V.

The work [28] developed an evaluation model for ECG quality based on the Kors matrix and correlation. The model has been measured using a database called PICC and correct results of 92% classification with 75.1 % Sp and 97% Se.

The work [29] presented a quality ECG classifier based on residuals among noticed and filtered signals and among effective subset linear estimations without filtered data and constant term. The coefficients of prediction or estimation have been identified from the acceptable quality of ECG using the robust model. This model attained an accuracy of 96.9 % Se, 88.3% NPV, 94.5% NPV, 80.4% Sp on database PICC. The model yielded 90.0% of accuracy over test data.

The work [30] discussed the heuristic policies, which have been used for identifying frequent issues in recordings of ECG in real-scenario over a phone. The approved or tolerable signals have extremely high or small ranges of amplitude. When the range has been higher than 15mV and lower than 0.2mV, then the signal has been likely unacceptable or intolerable. Five rules have been developed to identify poor contact, maximal amplitude artifact, missing lead as in [4]. The outcomes exhibited that integration of various rules has been capable of identifying the most of poor ECG signal quality correctly, as shown by using database PICC. Moreover, this model attained 0.896 and 0.913 for testing and training sets in respective order.

The work [31] depicted an algorithm depending on thresholds set for measuring the signal quality of ECG over mobile. This model identifies poor signal quality, minimal amplitude, steep slope, saturation, and many more. Moreover, the model has been measured by using 500 ECGs considered from the database PICC. Also, it has 85.7% of accuracy. The work [32] explained the false alarm reduction algorithm using one ABP waveform (arterial blood pressure). By utilizing simultaneous ABP and ECG recordings, the entire rate of FA has been decreased from 42.7 – 17.2%

The work [9] presented ECG index quality based on modulation spectrum for the applications of telehealth. Moreover, MSSR (modulation spectral signal representation) has been utilized for quantifying the variance among noise sources and ECG waveforms. The model has been measured on 12-lead ECGs and 2- lead ambulatory-ECGs from MITBIHA and PICC databases in respective order. The artificial ECGs with changing levels of noise and ECGs in real-time have been attained by utilizing Hexoskin garment at the time of 3 activities such as running, sitting, and walking. The outcomes exhibited that the model performs better in quality metrics like sample kurtosis ECG and spectral power-ratio.

The work [33] discussed several models incorporating analysis of time, frequency, cross-correlation, entropy, time-frequency for measuring the quality of ECG. Also, the quality of ECG has been evaluated by regularity matrix spectral radius. The measurement of the PICC database exhibited that algorithms are having 93.5% and 90% in training and testing set in respective order.

The work [34] presented noise artifact and automatic motion identification in Holter ECGs by utilizing EMD (experimental mode decomposition) as well as statistical features like variance, Shannon entropy, and mean for primary IMF (intrinsic mode-function). Also, the model attained a Sp and Se of 94.73 % and 96.63% in respective order for 30 sets of testing data and 15 sets of training data.

The work [35] presented a model depending on RMS (root mean square) and residual errors after isolating the noises from the ECG signal by utilizing Karhunen-Loeve-Transform. Also, this model had 99.98% of Sp and 99.57% of Se on the European database ST-T. Many of the models extracted the distinct features from noisy as well as noise-free signals of ECG. The work [36] presented a novel quality index of ECG based on spectral modulation signal representation. The SVM and LDA classifiers have been tested for discriminating among non-usable and usable segments of ECG.

The work [37] presented a model depending on the cross-covariance matrix amongst several leads, covariance matrix eigenvalues, and SVM and supervised-decision tree classifier. The outcomes on database PICC exhibited that the model is having 0.89 of accuracy over test data.

The work [38] presented a model for physiological signals quality measurement at ambulatory measurements depending on cumulative plots frequency and KS-test. They also presented the quality indices of signal and fusion of data to define ECGs acceptability gathered in noisy environments. Moreover, in this contribution, 6 SQLs like Psql, kSQL, Fsql, bSQL, and iSQL and five classifiers (ANN, SVM, NB, and MLP) have been utilized to measure the quality of the signal. The model has been assessed by utilizing the database PICC. Moreover, the model had 99% of accuracy in set a, and 92.6% of accuracy in test data, 95% of accuracy in set b.

The work [39] presented a model for automatic recognition of reliable heart rates computed from PPG and ECG waveforms. Also, this model has been measured by utilizing 158 records gathered at the time of helicopter transport from suffering patients. This model attained an exact classification of 92%. The review exhibited that; single signal processing model has not been enough for eradicating divergent types of artifacts and noises in signals of ECG as in [40]. Furthermore, denoising outcomes exhibited that every filtering strategy might introduce distinct types of the waveform-distortion. Hence, it has been prominent for recognizing the noise's nature in ECG recordings and selecting suitable signal processing models suitable for noise types. Moreover, in some recording scenarios of ECG, the artifacts and noises might partially occur in 10s of ECG signal as in [19]. Hence, it has been prominent for determining localized boundaries noisy ECG portions either for marking as unreliable evaluations or for eradicating the noisy segments of ECG from the extraction of feature

The earlier studies exhibited that; contaminated signals of ECG might cause a reduction of classification accuracy because of detection errors of R-peak in the existence of noise or artifacts in the signal of ECG [16]. Moreover, in such cases, the incorrectly detected beats might be categorized as abnormal or normal beats inexactly. Hence, preprocessing phase of many models uses the denoising model for eradicating noise and artifacts from signals of ECG. Moreover, every denoising strategy distorts distinct local waves like T-wave, P-wave, and QRS complex. It introduces divergent waveform distortions types like reduction of amplitude, shape variation, and width widening.

Moreover, normal heartbeats have been misclassified as normal heartbeats and vice versa because of detection errors

of R-peak and delineation errors of ECG waveform. Hence, some heartbeat classification models used the quality measurement of ECG signal for discriminating among bad and good quality signals of ECG. Nevertheless, many of these models have been presented depending on interval and morphological features and ML models [10], [41]. Current models demand huge collections of abnormal and normal bad and good quality ECG signals for capturing distinct features of divergent kinds of noises, artifacts, and heartbeat waveforms.

In a previous study, "Supervised Learning-based Noise Detection to Improve the Performance of Filter-based ECG Signal Denoising (SLND)" [42], the topic of supervised learning-based noise detection in electrocardiograms was covered. The approach discussed in [42] uses supervised learning to identify unlabeled electrocardiograms based on the y-coordinates of x-factors in electrocardiograms tagged as noisy and benign. The other recent approach is indeed an ensemble technique constructed with RUSBOOST [43], which uses auto correlation to find artefacts in the provided ECG by comparing its five second epochs to other labelled electrocardiograms. Despite these contributions showing a performance with little false alarms, the model's detection accuracy for electrocardiograms with little noise at targeted peaks is in doubt.

3 Methods and Materials

The suggested method for detecting noise in electrocardiogram signals uses supervised learning and learns from the best digital features that are combined from metrics for diversity evaluation. Using a hierarchical label prediction process, a classifier has been created using the heuristic search method known as cuckoo search. Based on the Likert Scale, we have developed a system to assign relative importance to the positive (noise-prone) and negative (noise-free) labels for features [44]. The main idea behind the strategy is as follows.

The next section's features must be extracted from the electrocardiogram signals that are supplied and fall under one of the labels for noise or benign. A two-dimensional matrix will be used to display the features of recordings with positive (noise-prone) labels as a result. The electrocardiogram signal's related features are shown in each row, and their values are represented in each column. A two-dimensional matrix corresponding to the feature values of the negative recordings (electrocardiogram signals devoid of noise) must be generated.

Both the features of the positive and negative labels will be used in the fusion assessment metrics. This relates to making the best feature-features for both labels. For the

purposes of determining whether or not an electrocardiogram signal is noisy (positive), the ideal column to use is the one that corresponds to the observed diversity (negative). The "cuckoo search" heuristic search approach is used to build the classifier in the latter phase, which creates n-grams from the best features of each label. The proposal's last stage involves determining whether or not a specific electrocardiogram signal is noise-prone.

Regarding the fusion of diversity evaluations, the suggestion used the KS-Test (Kolmogorov-Smirnov test), the MWU-Test (Mann-Whitney U Test), and the dual tailed t-test (T-test) as three distribution diversity evaluations [46, 47, 48].

3.1 The data structure and features

Irregularities in the heartbeat rhythm are referred to as arrhythmia, and the abnormalities scale to the conditions of a higher or lower rate of heartbeats which are generally non-conductive for good health. The data format in terms of ECG signal series is the pattern used for understanding the heartbeat rate and any variations imperative in heart functioning. In the contemporary diagnosis models, such signals are clustered based on distinct tridimensional features explored in the section below.

3.1.1 Intervals

RR Interval: Over the ECG readings, the time-lapse observed for the sub-sequential R-waves in the QRS signals are technically termed as the RR intervals. While the objective is to understand the autonomic influence conditions and sinus node aspects, alongside the heart rate functions, the role of RR intervals can be more significant.

PR Interval: Over the instances of developing a QRS complex, the time consumed towards the successful onset of the P wave is termed as PR interval. It refers to the conduction depending on the node AV. In general, the PR frequency is within 120-200ms of duration. In the instances of PR interval resulting above >200ms, it is imperatively indicating the possibility of first-degree heart block.

QT interval is the parameter enabling track of heart performance over multiple durations. In general, the QT interval is around 0.4 to 0.44 seconds. It is a medically examined fact that the QT intervals in the female patients are seen to be higher than that of the males. At times, the tendency of lower heart rate results in longer QT interval conditions.

QRS duration is rated in general around 0.08 to 0.10 seconds, and any range reported in terms of 0.10 to 0.12 seconds, wherein such range is considered as moderate, and

the values that could be seen to have slightly prolonged. Any kind of QRS duration results with more than 0.1 seconds is considered aberrant.

QTc frequency is about defining the generic range of 0.40s to 0.44s. If there is an immediate arrest of heart functions, there is a spectrum range of QTc envisaged for male patients having 0.43 to 0.45 seconds, and the female patient ranges are at 0.45 to 0.47 seconds.

3.1.2 Axis

Axis works as a critical metric referring to the key directions in terms of electrical conditions of the body in aggregate taking place over the varying heart performance. Such an instance can result in any direction per say right or left or normal. In certain conditions, it can lead to significant variations resulting in inclination to the northwest axis.

P wave Axis variance refers to the conditions wherein the atrial depolarization emerges, with many of the sinus nodes active, and is determined as a sinoatrial node. It triggers the conditions in terms of depolarizing the atria. P wave in a conducive scenario should result in lead II, with action potential resulting in SA nodes more effectively.

The intensity of T wave Axis: T wave in an ECG denotes the repolarization aspect of the ventricles. An absolute refractory period is seen when the frequency range amidst the T waves apex and the QRS complex is complex and are usually considered as abnormally inverted T wave conditions that could be transforming to a set of cardiac and non-cardiac constraints. In exceptions to the right precordial leads, generally when the T wave is intensely transiting to the QRS conditions.

QRS Wave Axis Intensity: The axis range is advocated as normalcy if the band is of -30 to +90 degrees. Variations envisaged towards the left side can be vectors ranging variation of -30 to -90 degrees. Similarly, any variation of QRS towards the right side is imperative for a range of +90 to +180 degrees.

However, in either case of P or T wave, the procedures, in general, refer to the QRS axis conditions, and the limb refers to the conditions that require investigation. In general, the QRS axis needs to be in the interval of -30 to +90 degrees. As mentioned above, the left axis and right axis ranges too are to be considered. Thus, the indeterminate range of the axis can be termed as $\{(+/-)190, -90\}$.

3.1.3 Signal

Periodical impact leading to changes in the statistical features of biomedical signals is a common phenomenon. Wavelet

transform plays a vital role in signal representation in handling both time and significant frequency domains. It shall also help in understanding the quasiperiodic signals like the ones determined in ECG. In processing the ECG signals for a more effective range of feature extractions [49], heartbeat recognition [50], and de-noising [51]. The role of wavelet transformation can be impeccable. In the prototyped solution, if the DWT is used in tasking the feature extraction, the possibility of dissection into lower or higher frequency levels of approximation factors is evident.

DWT is used as it is vividly used for reasons like orthogonality properties like the Coiflets [52], Symlets, Daubechies, and Discrete Meyer [52], [53]. Heartbeat-related input is fragmented according to definitive impulse response instance emerging for the Discrete Mayer's wavelet transform frequency intensity as fourth-level approximation sub-band in the range of 011.25Hz and level four detail sub-band in the range of 11.2522 Hz.

Around two hundred coefficients are garnered using the wavelet features processed by ICA, envisaging potential address of dimensionality reduction issues. In aggregate, six of the components of ICA are chosen from the basket of sub-bands DWT, and the requisite twelve structural features essential are generated.

3.2 Preprocessing

In the preprocessing stages, the input ECGs are classified towards positive or negative as arrhythmia and benign, respectively, from the datasets adapted for training.

In furtherance, the features of the formats are enlisted as interval CS , Axis AS ,

- Accordingly, the respective signals FS of each input ECG value for the training datasets were chosen.
- The record of frequencies $\{r \exists r \in CS\}$, otherwise denoted as intervals, signifies the varying frequencies stacked in corresponding ECG.
- The count of axis values $\{ar \exists ar \in AS\}$ of the ECG depicts the divergences imperative in the axis formats.
- Also, it attains the count of signal n-grams $\{fo \exists fo \in FS\}$ of each ECG of the chosen datasets.

3.3 Mann-Whitney U Test

The MWU-Test (Mann-Whitney U Test) [47] is one of the diversity assessment labels that excludes consideration of the distribution structure that is appropriate for the majority of datasets with labelled diversity. The MWU-Test implementation procedure is described as follows:

The vector distributions v_1, v_2 have been used as input to the MWU-Test, which is the technique used to determine the degree of diversity between related vectors as follows:

All of the entries from the vectors v_1, v_2 are first transferred to the new vector v . The vector should also be sorted by ascending order of values, and the indices of the vector's ordered values should be treated as the corresponding ranks R . The rank of any individual identical value will be determined by the average of its indices. The ranks given to the values of the vector v_1 are denoted as the set R_1 as well as the ranks allocated to the values of the vector v_2 are denoted as the set R_2 . The rank-sum threshold (RST_1) of the vector v_1 is afterwards determined by the procedure using the aggregate of the items in the set R_1 as RS_1 , which is subsequently utilised in (Eq 1) as follows:

$$RST_1 = RS_1 - ((|v_1| + 1) * |v_1|) * 2^{-1} \dots \text{(Eq 1)}$$

$$RST_2 = RS_2 - ((|v_2| + 1) * |v_2|) * 2^{-1} \dots \text{(Eq 2) // rank-sum threshold } RST_2 \text{ of the vector } v_2$$

// the notations $|v_1|, |v_2|$ denotes the size of the vectors v_1 and v_2 .

Then the rank-sum threshold RST of the vectors v_1, v_2 can be estimated as follows:

$$RST_+ = RST_1 \dots \text{(Eq 3)}$$

$$RST_+ = RST_2$$

The z-score [53] can be found as follows:

$$m_{RST} = RST * 2^{-1} \dots \text{(Eq 4) //average of the RST}$$

$$d_{RST} = \left\{ \frac{\sqrt{((|v| + 1) * |v| * |v_2|) * |v|^{-1} - ((|v| + 1) * |v| * |v_2|) * |v|^{-1})}}{\left(\sum_{i=1}^k (t_i^3 - t_i) * (|v| * (|v| - 1))^{-1} \right)^{-1}} \right\} \dots \text{(Eq 5)}$$

//deviation of the RST

The 'k' denotes the total distinct ranks, the 't_i' denotes the total entries ranked as 'i'

Further, the z-core assesses as follows:

$$z = (d_{RST})^{-1} * (RST - m_{RST}) \dots \text{(Eq 6)}$$

Then, in the z-table [54], locate the p-value for the z-score that is shown. The distribution of v_1, v_2 vectors is determined to be more diverse if the p-value is larger than the specified probability threshold. Otherwise, the distribution is equivalent.

3.4 KS-test

KS -test (Kolmogorov-Smirnov test) is the diversity in terms of distribution imperative among two datasets mentioned as distance metric [46]. Distance metrics do not depend on the size of the information about the distribution kind of data differing KS-test in lines with the other methods for detecting distribution-related diversities. Application of the KS-test is carried out based on the following information.

Let Values of two vectors be v_a, v_b and KS -test shall be implemented based on evaluating the distributions for two vectors similar or divergent to the conditions.

The procedure leads to aggregating $Ag(v_a), Ag(v_b)$ values of v_a, v_b vectors, in sequential and accordingly, the cumulative ratio is predicted for each entry of the vectors chosen.

$$pr = 0$$

$$\forall_{i=1}^{|v_j|} \{e_i, \exists e_i \in v_j\} \text{ begin}$$

$$pr = \frac{e_i}{Ag(v_j)} + pr \quad \dots(\text{Eq 7})$$

$$CR_{v_j} \leftarrow pr$$

$$\text{end}$$

Thus, as depicted in the equation.7, e_i implies each element of a vector v_j . The interpretation pr refers to the aggregate ratio of the prior element in iteration; the depiction $Ag(v_j)$ refers to the summation values denoted in the v_j , and notion CR_{v_j} refers to the set constituting accumulated ratios of fundamental elements available as v_j . The values depicting the conditions mentioned above refers to identifying aggregated ratios of values represented in specified v_a, v_b vectors as corresponding sets CR_{v_a}, CR_{v_b} .

Further discovers the nonnegative difference of “aggregated ratios” related to values implicit at a related index of both v_a, v_b vectors.

$$\max_{i=1}^{(CR_{v_a}, CR_{v_b})} \{c_i(v_a), c_i(v_b) \exists c_i(v_a) \in CR_{v_a} \wedge c_i(v_b) \in CR_{v_b}\}$$

Begin // for each index i , values present in CR_{v_a}, CR_{v_b}

$$AD_{CR_{v_a} \leftrightarrow CR_{v_b}} \leftarrow abs(c_i(v_a) - c_i(v_b))$$

// the non-negative difference AD of aggregated ratios presented in sets” CR_{v_a}, CR_{v_b} at the index i , which is further stored in the set $AD_{CR_{v_a} \leftrightarrow CR_{v_b}}$.

End

In furtherance, observe for d-stat, which contains max value implied in the $AD_{CR_{v_a} \leftrightarrow CR_{v_b}}$

Observe for d-critic in the KS-table targeting at $Ag(v_a), Ag(v_b)$ mentioned as “degree of probability threshold” $p\tau$ (generally 0.05).

In the instance of d-stat being more than the d-critic value, the distribution of the vectors could be resulting as similar.

3.5 Dual Tailed Variance-Test

Assessment of distribution diversity in the suggested fusion procedure, a dual-tailed diversity test, often known as a t-test, has been employed. Applying this technique, we may ascertain the variation in projected feature values (y-coordinates) across records with distinct labels (x-coordinates). The dual-tailed t-test results improved in recent contributions [48], [55], [56], [57] have prompted to use this approach in the suggested fusion strategy.

The t-score is the diversity amongst the chosen vectors $\{r_i, r_j \exists i \neq j\}$ distributions, is estimated based using equation-8 mentioned below.

$$f(r_i, r_j) = (\mu(r_i) - \mu(r_j)) * \left(\sqrt{\sigma_{(r_i)} + \sigma_{(r_j)}} \right)^{-1} \dots(\text{Eq 8})$$

The vector’s r_i, r_j t-score is represented by the function $f(r_i, r_j)$. The symbols $\mu(r_i), \mu(r_j)$ represent the mean values regarding the respective vectors. The associated vectors’ deviation is denoted by the notations $\sigma_{(r_i)}, \sigma_{(r_j)}$.

The significant variability among the selected vector distributions is shown by t-score observations that are smaller than the stated p-value (probability value) [54], [53].

3.6 The classifier

Modern meta-heuristic algorithms were developing and gaining popularity for tackling a variety of engineering challenges after being inspired by nature. The CS (cuckoo search) method is used in this work. Even though they may destroy other people's eggs to increase the likelihood that their own eggs would hatch, these cuckoos place their eggs in nests. A significant number of species have accepted the responsibility of brood parasitism by depositing eggs in the nests of other host-birds. There were three main categories of brood parasitism: Cooperative breeding, nest invasion, and intra-specific brood parasitism are the first three. The host bird may either discard the alien eggs or just quit the nest and build a new one after realizing the eggs are not their own.

3.7 Optimal Feature Selection

Find the best features in comparison to the opposite set for each pair of positive as well as negative labels representing the respective sets M_+ , M_- . If the values anticipated for the corresponding feature x_i of the counterpart set M_- match those projected for the corresponding feature in the set M_+ , then that feature is in its optimum state. The algorithm will estimate the diversity weight that is the absolute difference among maximal similarity one and likely similarity observed ($0 < p\text{-value} \leq 1$), with each feature x_i of the set M_+ when compared to the values of the feature x_i in the set M_- . The following description is a mathematical model for extracting the best features out of each pair of sets.

```


$$\forall_{i=1}^{cl} \{c_i^+, c_i^-, \exists c_i^+ \in M_+, c_i^- \in M_-\} \text{ Begin // for all the features}$$


$$dw_i = 1 // \text{ the diversity of the } i^{th} \text{ feature towards } M_+, M_-$$


$$\left. \begin{aligned} p_{ks} &= KS - test(c_i^+, c_i^-) \\ p_{mwu} &= MWU - test(c_i^+, c_i^-) \\ p_{dt} &= DT - test(c_i^+, c_i^-) \end{aligned} \right\} // \text{ diversity}$$

    assessment of the  $i^{th}$  feature towards  $M_+, M_-$ 

$$d(x_i)_{D_j \leftrightarrow D_k} = d\tau // \text{ Diversity of the feature } x_i \text{ towards } D_j \leftrightarrow D_k \text{ is set to } d\tau$$

    if  $((p_{ks} < p\tau) \vee (p_{mwu} < p\tau) \vee (p_{dt} < p\tau))$ 
    Begin // if one of the diversity assessment measures
  
```

exhibit the similarity that is greater than the given threshold $p\tau$

```


$$dw_i = 1 - (p_{ks} \otimes p_{mwu} \otimes p_{dt}) // \text{ the diversity } dw_i \text{ of the } i^{th} \text{ feature between the sets } M_+, M_- \text{ has been discovered from the fusion of the diversity estimation measures.}$$

  
```

End // of the condition

```

    if  $(dw_i \geq d\tau)$  // if the diversity weight  $dw_i$  of the  $i^{th}$  feature is greater than or at least equal to  $d\tau$ 
    Begin // denotes that the corresponding feature is optimal
    oF  $\leftarrow i$ 
    End
  
```

End // of iterations

// Preprocess the datasets of diversified labels//

```


$$\forall_{i=1}^{cl} \{i \in oF\} \text{ Begin // for each optimal feature } x_i \text{ of the set } D_j \text{ which is labelled as } j$$

  
```

```


$$M_+ \setminus c_i^+ // \text{ discarding the column of values projected to the corresponding } i^{th} \text{ feature from the set } M_+$$

  
```

```


$$M_- \setminus c_i^- // \text{ discarding the column of values projected to the corresponding } i^{th} \text{ feature from the set } M_-$$

  
```

End

// n-grams selection//

```


$$\forall_{i=1}^{|M_+|} \{r_i \in M_+\} \text{ Begin // for each record } r_i \text{ of the set } M_+$$

  
```

```


$$nG_+ \leftarrow (nG_+ \cup nGrams(r_i)) // \text{ finding the possible combination of n-grams of sizes 1 to n, and adding the unique, non-existent n-grams that result to the collection } nG_+.$$

  
```

End

```


$$\forall_{i=1}^{|M_-|} \{r_i \in M_-\} \text{ Begin // for each record } r_i \text{ of the set } M_-$$

  
```

```


$$nG_- \leftarrow (nG_- \cup nGrams(r_i)) // \text{ finding the possible combination of n-grams of sizes 1 to n, and adding the unique, non-existent n-grams that result to the collection } nG_-$$

  
```

End

The next step is to calculate the label level positive, negative, purity, and decision-supporting n-gram occurrence probabilities for the relevant set.

$\forall_{i=1}^{|nG|} \{ng_i \exists ng_i \in nG\}$ Begin // for each n-gram $\{ng_i \exists ng_i \in nG\}$ of the set nG

$ngp_i^+ = \left(\sum_{k=1}^{|M_+|} \{1 \exists ng_i \subseteq r_k \wedge r_k \in M_+\} \right) * (|M_+|)^{-1}$ // the n-gram

ng_i 's probability of being positive ngp_i^+ with respect to set M_+

$ngp_i^- = \left(\sum_{k=1}^{|M_-|} \{1 \exists ng_i \subseteq r_k \wedge r_k \in M_-\} \right) * (|M_-|)^{-1}$ // the n-gram

ng_i 's probability of being negative ngp_i^- with respect to set M_-

$ngw_i = 1 - (ngp_i^+ * ngp_i^-)$ //The purity of the n-gram ng_i towards the set D_j that derived through Gini-impurity estimation [56]

End //of the iterations

Find-positive-decision weights: Begin //Decision weights of the n-grams towards the positive label

$ip_i = 1$ // initializing the impurity ip_i of the n-gram to 1

$\forall_{i=1}^{|nG_+|} \{ng_i \exists ng_i \in nG_+\}$ Begin // For each n-gram $\{ng_i \exists ng_i \in nG_+\}$ of the set D_j

$ip_i = 1 - (ip_i * \{ngw_i \exists ng_i \in nG_-\})$ // Updating the impurity ip_i of the n-gram ng_i

End

$\forall_{i=1}^{|nG_+|} \{ng_i \exists ng_i \in nG_+\}$ Begin // For each n-gram $\{ng_i \exists ng_i \in nG_+\}$ of the set M_+

$ngdw_i^+ = ngw_i - ip_i$ // The decision weight $ngdw_i^+$ of the n-gram $ng_{[j,i]}$ towards the set M_+

if $(ngdw_i^+ < d\tau)$ begin // If the decision weight $ngdw_i^+$ of the n-gram ng_i towards the set M_+ is less than the decision weight threshold $d\tau$

$nG_+ \setminus ng_i$ // Discarding the n-gram ng_i of the set M_+

End

End

Find-negative-decision weights: //Decision weights of the n-grams towards the negative label

$ip_i = 1$ // The overall impurity ip_i of the n-gram ng_i is initialized to maximum, which is 1

$\forall_{i=1}^{|nG_-|} \{ng_i \exists ng_i \in nG_-\}$ Begin // For each n-gram $\{ng_i \exists ng_i \in nG_-\}$ of the set M_-

$ip_i = 1 - (ip_i * \{ngw_i \exists ng_i \in nG_+\})$ // Updating the impurity ip_i of the n-gram ng_i towards the negative label

End

$\forall_{i=1}^{|nG_-|} \{ng_i \exists ng_i \in nG_-\}$ Begin // for each n-gram $\{ng_i \exists ng_i \in nG_-\}$ of the set M_-

$ngdw_i^- = ngw_i - ip_i$ // the decision weight $ngdw_i^-$ of the n-gram ng_i towards the set M_-

if $(ngdw_i^- < d\tau)$ Begin // If the decision weight $ngdw_i^-$ of the n-gram ng_i towards the set M_- is less than the decision weight threshold $d\tau$

$nG_- \setminus ng_i$ // Discarding the n-gram ng_i of the set M_-

End

End // of assessing negative decision weights of the n-grams

3.8 Build the Classifier

Regarding the performance of the proposed cuckoo search-based classifier, the classifier's learning phase depicts the perch hierarchy towards each label in the training corpus, with each level of the hierarchy having one or more perches. Each level's perch displays the same-sized n-gram features. Each perch of the hierarchy represents a feature (n-gram representation of x-coordinates). The n-gram features of the perches listed in a level of the hierarchy are smaller than the n-grams of the perches in the antecedent level and larger than the n-gram features of the perches in the descendant level. Placing distinct n-gram feature values (n-gram y-coordinates) of related n-gram features on the perches of hierarchy (as eggs) of each label is the next job of the learning phase.

3.9 Label Prediction Phase

In the testing phase, the predictive analytic task that forecasts the label of the given record executes as follows.

Let the test record be represented by the notation tr .

Let nG_{tr} denote the set of all conceivable unique n-gram patterns, with size $|tr|$ equal to the size of the test record tr .

$$f_{tr}^+ = \sum_{i=1}^{|nG_{tr}|} \sum_{idx=1}^{|H_+|} \sum_{j=1}^{|l_{idx}|} \left\{ \begin{array}{l} ngdw_i^+ \exists \\ ng_i \in p_j^+ \wedge \\ p_j^+ \in l_{idx} \wedge \\ l_{idx} \in H_+ \end{array} \right\}$$

// denotes the test record's positive fitness, which is determined by adding the decision weights assigned to each of its n-grams in relation to the positive label hierarchy.

$$f_{tr}^- = \sum_{i=1}^{|nG_{tr}|} \sum_{idx=1}^{|H_-|} \sum_{j=1}^{|l_{idx}|} \left\{ \begin{array}{l} ngdw_i^- \exists \\ ng_i \in p_j^- \wedge \\ p_j^- \in l_{idx} \wedge \\ l_{idx} \in H_- \end{array} \right\}$$

//The decision-weights of all the n-grams in the test record are also combined; these decision weights indicate how well the test record is suited for the negative label.

$$epdw_{tr}^+ = f_{tr}^+ \times \left(\sum_{idx=1}^{|H_+|} |l_{idx}| \right)^{-1}$$

//Determining the empirical-probability of the decision weights of the test record's n-grams in favor of the negative label hierarchy.

$$epdw_{tr}^- = f_{tr}^- \times \left(\sum_{idx=1}^{|H_-|} |l_{idx}| \right)^{-1}$$

//Determining the divergence of the n-gram judgement weights from the positive label hierarchy in the test record.

$$\sigma_{tr}^+ = \sum_{i=1}^{|nG_{tr}|} \sum_{idx=1}^{|H_+|} \sum_{j=1}^{|l_{idx}|} \left\{ \begin{array}{l} \sqrt{(f_{tr}^+ - ngdw_i^+)^2} \exists \\ ng_i \in p_j^+ \wedge \\ p_j^+ \in l_{idx} \wedge \\ l_{idx} \in H_+ \end{array} \right\}$$

//Determining the decision weights of the test record's n-grams that deviate from the hierarchy of the positive label.

$$\sigma_{tr}^- = \sum_{i=1}^{|nG_{tr}|} \sum_{idx=1}^{|H_-|} \sum_{j=1}^{|l_{idx}|} \left\{ \begin{array}{l} \sqrt{(f_{tr}^- - ngdw_i^-)^2} \exists \\ ng_i \in p_j^- \wedge \\ p_j^- \in l_{idx} \wedge \\ l_{idx} \in H_- \end{array} \right\}$$

//Determining the decision weights of the test record's n-grams that deviate from the hierarchy of the negative label.

The empirical-probabilities of the decision-weights obtained for all feasible n-grams of the provided test record and the corresponding deviations for both labels, as mentioned in the explanation above, will be evaluated.

3.9.1 Label Prediction

The decision about the recommended label positive (noise prone) or negative (noise-free) of the given test record tr shall perform as follows:

$$ip_{tr}^+ = 1 - (epdw_{tr}^- - \sigma_{tr}^-)$$

//impurity ip_{tr}^+ of the test record tr towards the positive label

$$mf_{tr}^+ = (epdw_{tr}^+ - \sigma_{tr}^+) - ip_{tr}^+$$

// minimal fitness mf_{tr}^+ of the given record tr towards the positive label

$$ip_{tr}^- = 1 - (epdw_{tr}^+ - \sigma_{tr}^+)$$

//impurity ip_{tr}^- of the test record tr towards the negative label

$$mf_{tr}^- = (epdw_{tr}^- - \sigma_{tr}^-) - ip_{tr}^-$$

// minimal fitness mf_{tr}^- of the given record tr towards the positive label

End

Finally, it prioritizes the provided record through the empirical-probability of decision-weights, relative deviation, and the minimal fitness towards the associated labels. This is true for both labels that signal positive and negative emotions. This strategy was inspired by the Likert scale notion [58], [59], [60], [61].

The empirical-probability as well as deviation are in the appropriate order after the minimal fitness of the unlabeled record towards the positive as well as negative labels, which is strongly valued. Using the Likert-scale, the parameter having the least importance will have an index of 1, while the progressively more important factors will each

have index increases of 1. The values representing the parameters must be multiplied by their normalized value indexed in the manner shown below in order to determine the correlation among the given record and the related label.

The range of the index will be 1, 2, as well as 3, representing deviation, empirical probability, and minimal fitness in that order because three parameters are taken into consideration for the correlation assessment. From this point forward, the test record's correlation with both labels will be calculated as follows.

$$cr_{tr}^+ = 1 - \left((epdw_{tr}^+ * (1-1^{-1})) * (\sigma_{tr}^+ * (1-2^{-1})) * (mf_{tr}^+ * (1-3^{-1})) \right)$$

// finding correlation cr_{tr}^+ of the test record tr towards positive (noise prone) label

$$cr_{tr}^- = 1 - \left((epdw_{tr}^- * (1-1^{-1})) * (\sigma_{tr}^- * (1-2^{-1})) * (mf_{tr}^- * (1-3^{-1})) \right)$$

// finding correlation cr_{tr}^- of the test record tr towards negative (noise-free) label

Further concludes that the test record is noise prone if the positive correlation cr_{tr}^+ is greater than or equal to the negative correlation cr_{tr}^- . Else the record shall be stated as noise-free

4 Experimental Study

This section of the study provides insights into how the experimental studies for the proposed model indicate the efficacy of the proposed model. However, to ensure the authenticity of the proposed model, two of the other comparable models, SLND [42], RUSBOOST [43], are used alongside the proposed model under similar testing conditions to understand how the proposed and the reviewed models are faring in terms of the metrics referring to the effectiveness of the models.

For the proposed study, using the datasets, the comparative analysis is carried out among three distinct models, one proposed in this manuscript and the other two SLND [42] and RUSBOOST [43]. Based on the feature selections for the respective models and the classification used, the following analysis based on certain critical parameters is discussed.

Table 1 is exhibiting the mean values and respective deviations of the cross-validation metrics. The values obtained from QAAES are outperforming the other two contemporary models. The true positive rate and true negative rate of the proposed method QAAES are exhibiting that both labels' detection imbalanced prediction accuracy is robust and far better than the contemporary models. The QAAES model has approximately 7% more imbalanced prediction accuracy than SLND and 12% more than RUSBOOST

The table-1 detailed below, represents the conditions wherein the average and the deviation values for various parameters are considered for all the three models used for comparative analysis. Based on the descriptive statistics of the datasets used, the positive predictive value, true negative rate, true positive rate, imbalanced prediction accuracy, and harmonic mean values related mean value are garnered, and even the variance for the same is estimated in significant ways.

In terms of simple interpretation of the values towards its impact and coherence to the mean values, the standard deviation assessment can be more significant. Low standard deviation refers to the conditions wherein the data are clustered around the mean, and higher levels of standard deviation refer to the conditions wherein it is spread out. Standard deviation is closer to zero refers to the data points being closer to the mean value, and any high or low standard deviation refers to the points being far above or below the mean value.

Table 1: The table exhibiting average and deviation values.

Average and deviation values			
	QAAES	RUSBOOST	SLND
Positive Predictive Value	0.94312± 0.005376579	0.92951± 0.011729744	0.87186± 0.013604499
True Negative Rate	0.94243± 0.005503099	0.92992± 0.011477526	0.86923± 0.016763237
True Positive Rate	0.960522222± 0.010787683	0.933755556± 0.021049238	0.893788889± 0.013265297
Imbalanced Prediction Accuracy	0.95235± 0.006133881	0.93274± 0.015653639	0.88164± 0.008954798
Harmonic Mean	0.94274± 0.005424426	0.9297± 0.011586371	0.87052± 0.01514865
Phi coefficient	0.90497± 0.012438895	0.86555± 0.031333824	0.76383± 0.017743396

In the table-1 above, it is evident that across all the metrics assessed for the three models respectively, the mean value is

comparatively higher in the QAAES when compared to the other models like RUSBOOST and SLND. Also, in the

standard deviation values, though the majority of the values in the matrix are lesser, and it signifies that the standard deviation values are clustered around the mean value. However, in the case of the QAAES, the values are more clustered compared to the other models and their variance values.

4.1 Positive Predictive Value (PPV)

The Positive Predictive Value is the probability referring to the positive results over a hypothesis test indicating there is a significant effect. In an illustrative scenario, it can be stated that the probability of patients testing positive being positive. Despite that, the degree of variance is feasible and could grade the level of impact. Still, the baseline impact is seen as positive. For instance, when a patient is tested for heart

problem identification using specific metrics, if the values remain positive, it refers to the possibility of a person suffering from a heart problem. However, the intensity is not evident in the PPV values.

The PPV values for the datasets are estimated using the following formulae.

$$PPV = \frac{\text{Number of true positives}}{\text{Number of true positives} + \text{Number of false positives}}$$

Based on the datasets used for the analysis and the ten-fold data set modeling used for testing the values, the positive predictive values across the ten-folds are higher in the case of the QAAES, in comparison to the other two models RUSBOOST and SLND.

Precision (Positive Predictive Value)										
	1	2	3	4	5	6	7	8	9	10
QAAES	0.9513	0.9444	0.9395	0.9318	0.9407	0.9454	0.9495	0.9449	0.9451	0.9386
RUSBOOST	0.9273	0.9175	0.9149	0.9107	0.9298	0.9268	0.9395	0.9444	0.9447	0.9395
SLND	0.8719	0.8709	0.8662	0.8684	0.8517	0.8483	0.8922	0.8883	0.8832	0.8775

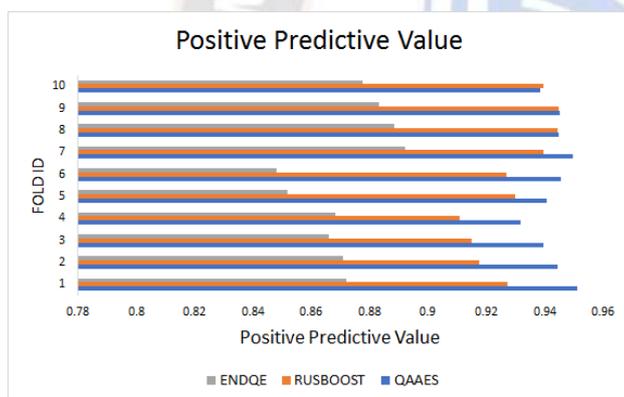


Figure 1: Precision (positive predict value) has noticed for QAAES, SLND [42] & RUSBOOST [43].

The figurative representation of the values in the figure-1 for PPV-related performance metrics indicates that the QAAES ranks higher in the performance, followed by RUSBOOST. The PPV value in the case of the SLND is lower in comparison to the other two models.

4.2 True Negative Rate (TNR)

TNR, also referred to as specificity of a test, is the proportion of samples that test negative among the data in question that are genuinely negative. For instance, in the dataset constituting 100 records, if genuinely 80 are negative for an implication test, if the prediction from the model refers to 78 as negative, then the TNR rates are evident in the model as more significant. For the other two records that were missed by the prediction values, it can be seen as False Positive Rate.

The formulae mentioned below are adapted to detail the specificity or the TNR values for the datasets.

$$specificity(TNR) = \frac{|TN|}{|TN| + |FP|}$$

// |TN| denotes total true negatives
 // |FP| denotes total false positives

True Negative Rate										
	1	2	3	4	5	6	7	8	9	10
QAAES	0.9513	0.9433	0.9398	0.9323	0.9402	0.9432	0.9502	0.9442	0.9439	0.9359
RUSBOOST	0.9293	0.9198	0.9161	0.9101	0.9296	0.926	0.94	0.945	0.944	0.9393
SLND	0.8693	0.8708	0.8654	0.8674	0.8415	0.8388	0.8921	0.8874	0.8828	0.8768

Figure 2 based on the analysis of the datasets for ten-fold analysis, the true negative rate of QAAES has outperformed the other two models. More in specific, QAAES has performed better than SLND across all the ten-fold analyses. Whereas in comparison to RUSBOOST, in six-folds, QAAES has shown more difference in the performance, and in the other four folds (7-10 in the above table), the margin of performance difference is narrow. However, focusing on the averaging of the performance factors, it is evident that the QAAES has depicted significant performance.

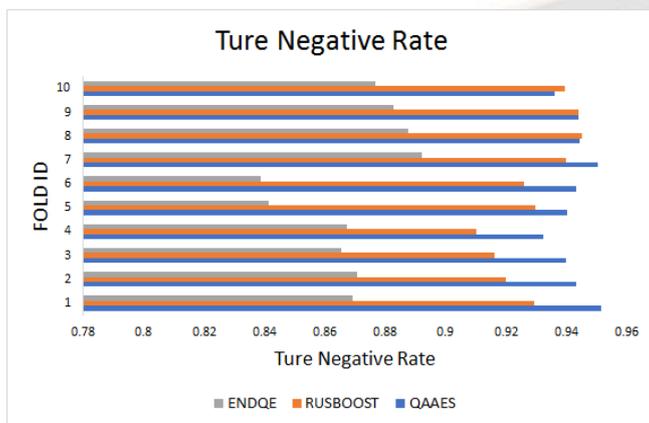


Figure 2: Specificity (True Negative Rate) has noticed for QAAES, SLND & RUSBOOST

4.3 True Positive Rate (TPR)

TPR is alternatively called sensitivity rate, refers to the conditions wherein the proportion of people with the disease being shown as positive. In simple terms, if the dataset has 100 records constituting 80 positive records and if the system can predict all the 80 positive records, the sensitivity of the model is much significant. However, considering the scope of processing errors in the medical interpretation, the higher the sensitivity rate, the higher the efficacy of the model can be.

The sensitivity factors in the model are assessed based on the following formulae.

$$TPR = \frac{\text{Number of True Positives}}{(\text{Number of True Positives} + \text{Number of False Negatives})}$$

The Figure 3 TPR test carried out on the dataset for the ten-folds, using the formulae mentioned above stands the computation of the below-mentioned table. In the probability of all the ten-folds, the performance of the QAAES has been outstanding compared to the other two models used for analysis. For instance, the average TPR rate of the SLND stands at 0.89, 0.93, and 0.97 in the respective cases of SLND, RUSBOOST, and QAAES. The performance indicates that the proposed model has the upper edge in defining the true positives more effectively.

True Positive Rate

	1	2	3	4	5	6	7	8	9	10
QAAES	0.95	0.9562	0.9531	0.9436	0.9733	0.9736	0.9616	0.9635	0.9698	0.9776
RUSBOOST	0.9132	0.9116	0.9083	0.9118	0.9476	0.9376	0.9552	0.9559	0.9626	0.9548
SLND	0.8877	0.8804	0.8785	0.8765	0.9107	0.9144	0.9101	0.8962	0.8896	0.8972

4.4 Imbalanced Prediction Accuracy

Imbalanced prediction accuracy is profoundly about focusing rate classification errors wherein the minority classes are given more weightage than the majority class. Unlike the standard evaluation metrics, wherein the classes are considered equally important. In the imbalanced prediction model, the emphasis is on the minority class, wherein there could be preliminary observations to train the model. Applying the following formulae for estimating the imbalanced prediction accuracy, the following tabulation is computed for the ten-fold analysis across all three models.

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}}$$

From the data interpretation, it is evident that the QAAES has an upper hand performance than the other two models. However, in a few of the folds like (8-10) in the below table, the performance of RUSBOOST and QAAES is

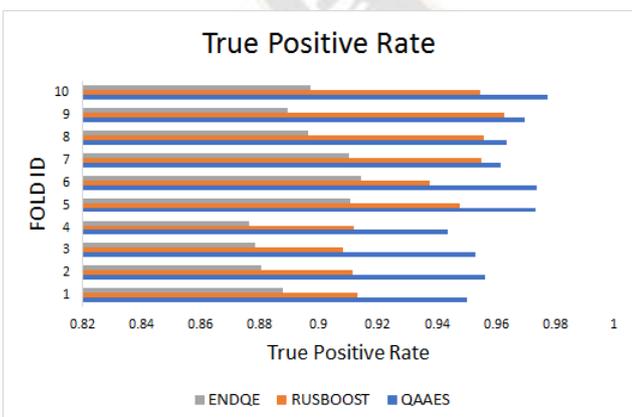
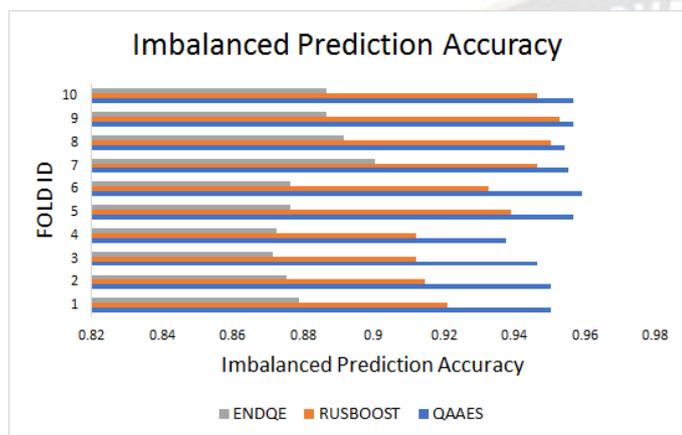


Figure 3: Sensitivity (True Positive Rate) has noticed for QAAES, SLND & RUSBOOST

seen as highly competitive. But the performance of SLND has remained below par with the performance of QAAES. The simple visual interpretation of the table in the form of a graph

indicates that QAAES ranks high in the model, followed by RUSBOOST, while SLND stands last in Figure 4.

Imbalanced Prediction Accuracy										
	1	2	3	4	5	6	7	8	9	10
QAAES	0.9503	0.9503	0.9465	0.9376	0.9567	0.9592	0.9554	0.9541	0.9567	0.9567
RUSBOOST	0.921	0.9146	0.9121	0.9121	0.9389	0.9325	0.9465	0.9503	0.9529	0.9465
SLND	0.879	0.8752	0.8713	0.8726	0.8764	0.8764	0.9006	0.8917	0.8866	0.8866



reciprocal of each of the numbers over the series. In simple terms, the Harmonic mean can be termed as the reciprocal arithmetic mean of the reciprocals. Focusing on the harmonic mean is to understand the mean performance value across various folds used in the analysis.

$$\frac{\sum_{i=1}^n w_i}{\sum_{i=1}^n \frac{w_i}{x_i}}$$

The harmonic mean formulae as depicted above is on the consideration of weights being equal to 1, and the weighted harmonic mean of if seen as X1, X2, X3, then the corresponding weights as W1, W2, and W3.

Figure 4: Imbalanced Prediction Accuracy has noticed for QAAES, SLND & RUSBOOST

4.5 Harmonic Mean

Harmonic mean refers to a kind of numerical average, wherein the number of observations is divided by the

Harmonic Mean										
	1	2	3	4	5	6	7	8	9	10
QAAES	0.9513	0.9438	0.9396	0.932	0.9404	0.9443	0.9498	0.9445	0.9445	0.9372
RUSBOOST	0.9283	0.9186	0.9155	0.9104	0.9297	0.9264	0.9397	0.9447	0.9443	0.9394
SLND	0.8706	0.8708	0.8658	0.8679	0.8466	0.8435	0.8921	0.8878	0.883	0.8771

In line with the computation interpretation, like the earlier metrics, even in the case of the harmonic mean values, the impact performance of QAAES is imperative compared to the other models assessed in the study. While the

RUSBOOST still stands as a strong contender, the performance of QAAES has the marginal edge in some folds, and in many folds, the performance of the QAAES is superior Figure 5.

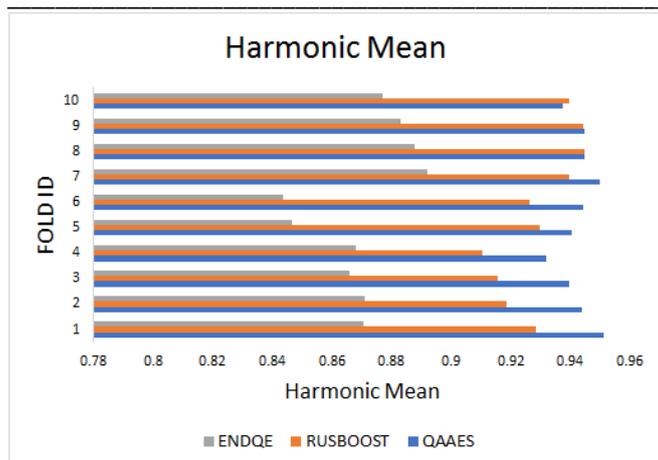


Figure 5: The F-Measure (Harmonic Mean) has noticed for QAAES, SLND & RUSBOOST

4.6 PHI Coefficient

Phi Coefficient refers to the measure of association among two binary values and is alternatively termed as Mean Square

PHI Coefficient										
	1	2	3	4	5	6	7	8	9	10
QAAES	0.9006	0.9007	0.8931	0.8752	0.914	0.9189	0.9109	0.9085	0.9137	0.9141
RUSBOOST	0.8421	0.8293	0.8242	0.8242	0.8779	0.865	0.8931	0.9007	0.9059	0.8931
SLND	0.7581	0.7504	0.7427	0.7453	0.7547	0.7553	0.8015	0.7835	0.7733	0.7735

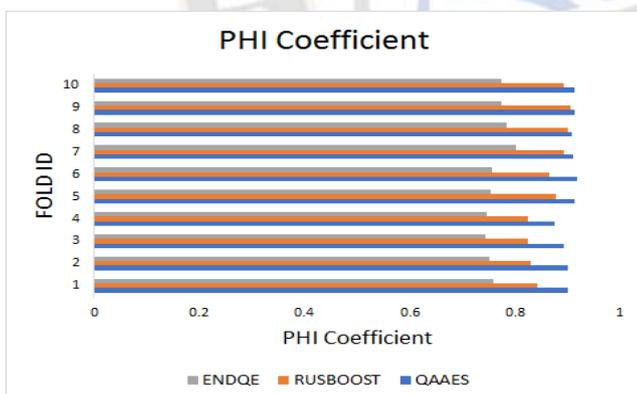


Figure 6: The PHI Coefficient has noticed for QAAES, SLND & RUSBOOST

Based on the overall metrics assessed in the experimental study, it is evident that the model proposed as QAAES in this manuscript stands superior in attaining the desired objectives in comparison to the other models reviewed in this study. Thus, considering the emerging requirements of the medical domain to have more accurate analysis, the proposed machine learning solution can be pragmatic for the futuristic diagnosis.

Contingency Coefficient. The association of the positive and negative results diagnosis-related coherence is assessed in this model using the Phi Coefficient.

Profoundly, the Phi Coefficient values are symmetrical statistics, wherein the independent variable and dependent variables are significantly interchangeable. Interpretation of the Phi Coefficient is the assessment of correlation coefficients ranging from -1 to 1, wherein 0 refers to hardly any relationship, 1 depicting perfect positive relationship, and -1 depicting perfect negative relationship Figure 6.

The Phi Coefficient values of all the three models across its ten-fold analysis refer to a condition wherein there is a perfect positive correlation. However, the degree of positive correlation is superior in QAAES when compared to RUSBOOST and SLND.

5 Conclusion

Predictive models applied for the ECG analysis are in high demand, and with the emerging technological advancements, more contemporary solutions are evident in the domain. This research manuscript discusses a contemporary solution in QAAES, in line with the model narrative discussed in the above sections. Based on the experimental analysis of the model over the ten-fold training sets from the datasets used, it is evident that the success of QAAES is comparatively higher than that of the other two relative models RUSBOOST and SLND. Despite that in some metrics performance, RUSBOOST has significant performance too in relative terms, and absolute results indicate QAAES having cutting edge performance than the other two. Thus, it is advocated that the recommended model QAAES applied as a machine learning solution for the pragmatic medical assessment can be a sustainable and futuristic solution.

References

- [1]. Quaglini, S., Rognoni, C., Spazzolini, C., Priori, S. G., Mannarino, S., & Schwartz, P. J. (2006). Cost-effectiveness of neonatal ECG screening for the long QT syndrome. *European heart journal*, 27(15), 1824-1832.

- [2]. Chawla, M. P. S., Verma, H. K., & Kumar, V. (2008). RETRACTED: Artifacts and noise removal in electrocardiograms using independent component analysis.
- [3]. Zaunseder, S., Huhle, R., & Malberg, H. (2011, September). CinC challenge—Assessing the usability of ECG by ensemble decision trees. In 2011 Computing in Cardiology (pp. 277-280). IEEE.
- [4]. Chudáček, V., Zach, L., Kužilek, J., Spilka, J., & Lhotská, L. (2011, September). Simple scoring system for ECG quality assessment on android platform. In 2011 Computing in Cardiology (pp. 449-451). IEEE.
- [5]. Varon, C., Testelmans, D., Buyse, B., Suykens, J. A., & Van Huffel, S. (2012, August). Robust artefact detection in long-term ECG recordings based on autocorrelation function similarity and percentile analysis. In 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 3151-3154). IEEE.
- [6]. Friesen, G. M., Jannett, T. C., Jadallah, M. A., Yates, S. L., Quint, S. R., & Nagle, H. T. (1990). A comparison of the noise sensitivity of nine QRS detection algorithms. *IEEE Transactions on biomedical engineering*, 37(1), 85-98.
- [7]. Oster, J., Behar, J., Sayadi, O., Nemati, S., Johnson, A. E., & Clifford, G. D. (2015). Semisupervised ECG ventricular beat classification with novelty detection based on switching Kalman filters. *IEEE Transactions on Biomedical Engineering*, 62(9), 2125-2134.
- [8]. Li, Q., & Clifford, G. D. (2012). Signal quality and data fusion for false alarm reduction in the intensive care unit. *Journal of electrocardiology*, 45(6), 596-603.
- [9]. Falk, T. H., & Maier, M. (2014). MS-QI: A modulation spectrum-based ECG quality index for telehealth applications. *IEEE Transactions on Biomedical Engineering*, 63(8), 1613-1622.
- [10]. Orphanidou, C., Bonnici, T., Charlton, P., Clifton, D., Vallance, D., & Tarassenko, L. (2014). Signal-quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring. *IEEE journal of biomedical and health informatics*, 19(3), 832-838.
- [11]. Castro, D., Félix, P., & Presedo, J. (2014). A method for context-based adaptive QRS clustering in real time. *IEEE journal of biomedical and health informatics*, 19(5), 1660-1671.
- [12]. Takalokastari, T., Alasaarela, E., Kinnunen, M., & Jämsä, T. (2013). Quality of the wireless electrocardiogram signal during physical exercise in different age groups. *IEEE journal of biomedical and health informatics*, 18(3), 1058-1064.
- [13]. Brüser, C., Antink, C. H., Wartzek, T., Walter, M., & Leonhardt, S. (2015). Ambient and unobtrusive cardiorespiratory monitoring techniques. *IEEE reviews in biomedical engineering*, 8, 30-43.
- [14]. Forouzanfar, M., Dajani, H. R., Groza, V. Z., Bolic, M., Rajan, S., & Batkin, I. (2015). Oscillometric blood pressure estimation: past, present, and future. *IEEE reviews in biomedical engineering*, 8, 44-63.
- [15]. Arteaga-Falconi, J. S., Al Osman, H., & El Saddik, A. (2015). ECG authentication for mobile devices. *IEEE Transactions on Instrumentation and Measurement*, 65(3), 591-600.
- [16]. Kiranyaz, S., Ince, T., & Gabbouj, M. (2015). Real-time patient-specific ECG classification by 1-D convolutional neural networks. *IEEE Transactions on Biomedical Engineering*, 63(3), 664-675.
- [17]. Yathiraju, D. . (2022). Blockchain Based 5g Heterogeneous Networks Using Privacy Federated Learning with Internet of Things. *Research Journal of Computer Systems and Engineering*, 3(1), 21–28. Retrieved from <https://technicaljournals.org/RJCSE/index.php/journal/article/view/37>
- [18]. Ye, C., Kumar, B. V., & Coimbra, M. T. (2015). An automatic subject-adaptable heartbeat classifier based on multiview learning. *IEEE journal of biomedical and health informatics*, 20(6), 1485-1492.
- [19]. Satija, U., Ramkumar, B., & Manikandan, M. S. (2017). Real-time signal quality-aware ECG telemetry system for IoT-based health care monitoring. *IEEE Internet of Things Journal*, 4(3), 815-823.
- [20]. Satija, U., Ramkumar, B., & Manikandan, M. S. (2017). Automated ECG noise detection and classification system for unsupervised healthcare monitoring. *IEEE Journal of biomedical and health informatics*, 22(3), 722-732.
- [21]. Clifford, G. D., Azuaje, F., & Mcsharry, P. (2006). ECG statistics, noise, artifacts, and missing data. *Advanced methods and tools for ECG data analysis*, 6(1), 18.
- [22]. Shanthi, D. N. ., & J, S. . (2022). Machine Learning Architecture in Soft Sensor for Manufacturing Control and Monitoring System Based on Data Classification. *Research Journal of Computer Systems and Engineering*, 2(2), 01:05. Retrieved from <https://technicaljournals.org/RJCSE/index.php/journal/article/view/24>
- [23]. Van Alste, J. A., & Schilder, T. S. (1985). Removal of baseline wander and power-line interference from the ECG by an efficient FIR filter with a reduced number of taps. *IEEE transactions on biomedical engineering*, (12), 1052-1060.
- [24]. Kolios, P., Panayiotou, C., Ellinas, G., & Polycarpou, M. (2016). Data-driven event triggering for IoT applications. *IEEE Internet of Things Journal*, 3(6), 1146-1158.
- [25]. Satija, U., Ramkumar, B., & Manikandan, M. S. (2016). Robust cardiac event change detection method for long-term healthcare monitoring applications. *Healthcare technology letters*, 3(2), 116-123.
- [26]. Satija, U., Ramkumar, B., & Manikandan, M. S. (2016). Robust cardiac event change detection method for long-term healthcare monitoring applications. *Healthcare technology letters*, 3(2), 116-123.
- [27]. Zhang, Y., Wei, S., Long, Y., & Liu, C. (2015). Performance analysis of multiscale entropy for the assessment of ECG signal quality. *Journal of Electrical and Computer Engineering*, 2015.
- [28]. Mann, S. A., & Orglmeister, R. (2014, September). A flexible PCA-based ECG-reconstruction algorithm with

- confidence estimation for ECG during exercise. In *Computing in Cardiology 2014* (pp. 33-36). IEEE.
- [29]. Martinez-Tabares, F. J., Espinosa-Oviedo, J., & Castellanos-Dominguez, G. (2012, January). Improvement of ECG signal quality measurement using correlation and diversity-based approaches. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 4295-4298). IEEE.
- [30]. Maan, A. C., van Zwet, E. W., Man, S., Oliveira-Martens, S. M., Schali, M. J., & Swenne, C. A. (2011, September). Assessment of signal quality and electrode placement in ECGs using a reconstruction matrix. In *2011 Computing in Cardiology* (pp. 289-292). IEEE.
- [31]. Noponen, K., Karsikas, M., Tiinainen, S., Kortelainen, J., Huikuri, H., & Seppänen, T. (2011, September). Electrocardiogram quality classification based on robust best subsets linear prediction error. In *2011 Computing in Cardiology* (pp. 365-368). IEEE.
- [32]. Moody, B. E. (2011, September). Rule-based methods for ECG quality control. In *2011 Computing in Cardiology* (pp. 361-363). IEEE.
- [33]. Langley, P., Di Marco, L. Y., King, S., Duncan, D., Di Maria, C., Duan, W., ... & Murray, A. (2011, September). An algorithm for assessment of quality of ECGs acquired via mobile telephones. In *2011 Computing in Cardiology* (pp. 281-284). IEEE.
- [34]. Aboukhalil, A., Nielsen, L., Saeed, M., Mark, R. G., & Clifford, G. D. (2008). Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform. *Journal of biomedical informatics*, 41(3), 442-451.
- [35]. Xia, H., Garcia, G. A., McBride, J. C., Sullivan, A., De Bock, T., Bains, J., ... & Zhao, X. (2011, September). Computer algorithms for evaluating the quality of ECGs in real time. In *2011 Computing in Cardiology* (pp. 369-372). IEEE.
- [36]. Lee, J., McManus, D. D., Merchant, S., & Chon, K. H. (2011). Automatic motion and noise artifact detection in Holter ECG data using empirical mode decomposition and statistical approaches. *IEEE Transactions on Biomedical Engineering*, 59(6), 1499-1506.
- [37]. Oh, S. (2004). A new quality measure in electrocardiogram signal (Doctoral dissertation, University of Florida).
- [38]. Tobón, D. P., & Falk, T. H. (2015, May). Online ECG quality assessment for context-aware wireless body area networks. In *2015 IEEE 28th Canadian Conference on Electrical and Computer Engineering (CCECE)* (pp. 587-592). IEEE.
- [39]. Morgado, E., Alonso-Atienza, F., Santiago-Mozos, R., Barquero-Pérez, Ó., Silva, I., Ramos, J., & Mark, R. (2015). Quality estimation of the electrocardiogram using cross-correlation among leads. *Biomedical engineering online*, 14(1), 1-19.
- [40]. Schumm, J. (2010). Quality assessment of physiological signals during ambulatory measurements. ETH Zurich.
- [41]. Yu, C., Liu, Z., McKenna, T., Reisner, A. T., & Reifman, J. (2006). A method for automatic identification of reliable heart rates calculated from ECG and PPG waveforms. *Journal of the American Medical Informatics Association*, 13(3), 309-320.
- [42]. Satija, U., Ramkumar, B., & Manikandan, M. S. (2017). Noise-aware dictionary-learning-based sparse representation framework for detection and removal of single and combined noises from ECG signal. *Healthcare technology letters*, 4(1), 2-12.
- [43]. Abdulaal, A. H., Shah, A. F. M. S., & Pathan, A.-S. K. (2022). NM-LEACH: A Novel Modified LEACH Protocol to Improve Performance in WSN. *International Journal of Communication Networks and Information Security (IJCNIS)*, 14(1). <https://doi.org/10.17762/ijcnis.v14i1.5127>
- [44]. Naseri, H., & Homaeinezhad, M. R. (2015). Electrocardiogram signal quality assessment using an artificially reconstructed target lead. *Computer Methods in Biomechanics and Biomedical Engineering*, 18(10), 1126-1141.
- [45]. V. Supraja, P. Nageswara Rao, M. N. Giriprasad, (2021). "Supervised Learning-based Noise Detection to Improve the Performance of Filter-based ECG Signal Denoising" *Journal of Artificial Intelligence and Soft Computing Research*, 2022.
- [46]. Moeyersons, J., Smets, E., Morales, J., Villa, A., De Raedt, W., Testelmans, D., ... & Varon, C. (2019). Artefact detection and quality assessment of ambulatory ECG signals. *Computer methods and programs in biomedicine*, 182, 105050.
- [47]. Mu, E., & Pereyra-Rojas, M. (2016). Practical decision making: an introduction to the Analytic Hierarchy Process (AHP) using super decisions V2. Springer.
- [48]. Yang XS, D. S. (2014, Jan 1). Cuckoo search: recent advances and applications. *Neural Computing and Applications*., 24(1), 169-74.
- [49]. Ghasemi A, Z. S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International journal of endocrinology and metabolism*., 10(2), 486.
- [50]. McKnight PE, N. J. (2010, Jan 30). Mann-Whitney U Test. *The Corsini encyclopedia of psychology*., 1-1.
- [51]. Budak H, T. S. (2016, Dec 20). A modified t-score for feature selection. *Anadolu Üniversitesi Bilim Ve Teknoloji Dergisi A-Uygulamalı Bilimler ve Mühendislik*., 17(5), 845-52.
- [52]. Huikuri, H. V., Castellanos, A., & Myerburg, R. J. (2001). Sudden death due to cardiac arrhythmias. *New England Journal of Medicine*, 345(20), 1473-1482.
- [53]. Kadambe, S., Murray, R., & Boudreaux-Bartels, G. F. (1999). Wavelet transform-based QRS complex detector. *IEEE Transactions on biomedical Engineering*, 46(7), 838-848.
- [54]. Ye, C., Kumar, B. V., & Coimbra, M. T. (2012). Heartbeat classification using morphological and dynamic features of ECG signals. *IEEE Transactions on Biomedical Engineering*, 59(10), 2930-2941.
- [55]. Addison, P. S. (2005). Wavelet transforms and the ECG: a review. *Physiological measurement*, 26(5), R155.

-
- [56]. Reddy, K. Uday Kumar, S. Shabbiha, and M. Rudra Kumar. "Design of high-security smart health care monitoring system using IoT." *Int. J* 8 (2020).
- [57]. Nam, T. S., Thuc, H. V., & Long, N. V. (2022). A High-Throughput Hardware Implementation of NAT Traversal For IPSEC VPN. *International Journal of Communication Networks and Information Security (IJCNIS)*, 14(1). <https://doi.org/10.17762/ijcnis.v14i1.5260>
- [58]. Suneel, Chenna Venkata, K. Prasanna, and M. Rudra Kumar. "Frequent data partitioning using parallel mining item sets and MapReduce." *International Journal of Scientific Research in Computer Science, Engineering and Information Technology* 2.4 (2017)
- [59]. Sahoo, P., & Riedel, T. (1998). Mean value theorems and functional equations. World Scientific.
- [60]. Z-table. (n.d.). Retrieved from <http://www.sjsu.edu/faculty/gerstman/StatPrimer/z-table.PDF>
- [61]. Savoy. (2012, Jan). Feature selection in sentiment analysis. *Proc. CORIA*, no., 273-84.
- [62]. Rudra Kumar, M., Rashmi Pathak, and Vinit Kumar Gunjan. "Diagnosis and Medicine Prediction for COVID-19 Using Machine Learning Approach." *Computational Intelligence in Machine Learning*. Springer, Singapore, 2022. 123-133.
- [63]. 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.
- [64]. Rutkowski L, J. M. (2020). Misclassification Error Impurity Measure. *InStream Data Mining: Algorithms and Their Probabilistic Properties*, 63-82.
- [65]. Xue Y, H. M. (2017). Active learning of classification models with likert-scale feedback. In *Proceedings of the 2017 SIAM International Conference on Data Mining*, (pp. 28-35).