Study On AI-Assisted Health Detection Mechanism Based On ECG Data For Public Health

Kamlesh Kumar Yadav[1], Dhablia Dharmesh Kirit[2]

[1]*Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India*

[2]*Research Scholar, Department of CS & IT, Kalinga University, Raipur, India*

KEYWORDS

Electrocardiogram, Dynamic Ant Colony Optimized Intelligent Gaussian Naïve Bayes (DACO-IGNB), Atrial Fibrillation, Public Health, Artificial Intelligence (AI)

ABSTRACT

Health care is being revolutionized by Artificial Intelligence (AI) technologies that provide more precise diagnosis, customized therapies, streamlined administrative procedures and better patient care. A health detection mechanism assesses electrocardiogram (ECG) data to detect cardiac anomalies, offering precise and immediate atrial fibrillation diagnosis. We propose artificial intelligence (AI) assisted public health monitoring using ECG data. This study introduces the dynamic Ant Colony Optimized Intelligent Gaussian Naïve Bayes (DACO-IGNB) method to identify the atrial fibrillation detection mechanism. The Atrial Fibrillation (AF) ECG data were collected and Baseline Wander was to facilitate pre-processing. The Principal Component Analysis (PCA) method was used to extract the feature. Compared to other traditional methods, it demonstrated superior performance across accuracy (90.9%), F1-Score (0.87), recall (0.89) and precision (0.88) metrics. These findings underscore its efficiency in DACO-IGNB assisting health detection and diagnostics, significantly improving the accuracy and reliability of public health monitoring systems.

1. Introduction

Public health is focused on promoting and safeguarding population health, coordinated action is frequently necessary [1]. Automated detection and enhanced access for patients, physicians, and rural areas are two ways that digital technologies in conjunction with cloud products are revolutionizing healthcare [4]. Patient care and health tracking are improved with EGC devices with internet access and smart beds [2]. It can diagnose patients with the help of vital sign data transmitted via remote monitoring equipment. Patients with chronic illnesses, such as those with diabetes (sugar levels) or heart issues (EGC) are routinely observed. Significant mortality is caused by chronic diseases; 48% of deaths are related to cancer and heart disease [3]. Heart conditions such as myocardial infarction and arrhythmia are frequently diagnosed using ECG data. These signals are analyzed by computer-assisted medical diagnosis systems, which then use the patterns found in the ECG signals to make expert direct diagnoses [14]. When atrial fibrillation is diagnosed, the ECG usually displays an erratic rhythm without any discernible P waves and the QRS complexes are frequently spaced unevenly. This is indicative of both the quick and irregular ventricular response and the chaotic electrical activity in the atrial [13]. To effectively use AI-assisted health detection, it must be possible to manage computational complexity, ensure model interpretability, integrate with current systems, navigate ethical and legal issues and maintain generality and durability [5]. To develop early disease recognition, public health is monitored by creating and validating a DACO-IGNB-assisted public health finding method utilizing ECG data.

2. Related works

A multiclass advance for ECG study was accessible in the research. ECGs were segmented utilizing Convolutional Bidirectional Long Short Term Memory of Haleem et.al [6] algorithms as a central tool. On ECG rates, to predict heart failure, arrhythmia, and unexpected cardiac death from ECG rates, there was a contribution to public health [10]. To utilize the virtual healthcare (VH) doctor tool, Venkataramanaiah and Kamala [15] provided accessible healthcare to remote communities [7]. It outperformed existing machine learning (ML) algorithms with 99% accuracy by processing ECG signals, extracting features, and working a K-Nearest Neighbour (KNN) classifier, thereby enhancing public health through online remedial consultations. Atrial fibrillation (AF), a wide spread cardiac arrhythmia concurrent to a five-fold supplement in the risk of caress and humanity, was intentionally using the hierarchical attention network (HAN) ECG model Mousavi et al. [8]. The method analysed multi-resolution ECG patterns and found clinically important waves and heartbeats with the use of three attention mechanism levels. The conventional neural network (CNN) model of Panganiban et al. [9] was used to eliminate noise from ECG data that has been transformed into 2D images to extract distinctive maps through pooling and convolution. Aziz et al. [16] developed a diagnostic support system for gathering, analyzing, and evaluating medical records to diagnose heart disease in underprivileged areas, aiming to improve public health by retraining Google's Inception V3 model [12]. Beyond state-of-the-art procedures, a novel approach that located ECG peak sites used two-event-related moving-average (TERMA). Whereas TERMA was used to identify regions of interest and signals were spun in the time-frequency plane. Rashed et al. [11] proposed a ML technique for autonomous heart disease categorization, which was trained using ECG peaks, intervals, and characteristics. With time-frequency encoding of ECG values, they proposed a unique ECG beat classifier that used a modified VGG16-based CNN to find interpretable frequency contributions. This approach was used by clinicians intended for clinicians to improve the automation of cardiovascular diagnosis, contributing to public health.

3. Methodology

The ECG rhythm data were collected from kaggle, the data is pre-processed to reduce noise through baseline wander method. After noise was removed, ECG signals were segmented into smaller pieces and features were extracted using PCA. Subsequently, the signals were classified with probability class using GNB and the high probability classes were further searched with ACO, high probability class's debits the health detection mechanism based on ECG rhythm data. The proposed flow for this method is given in Figure 1.

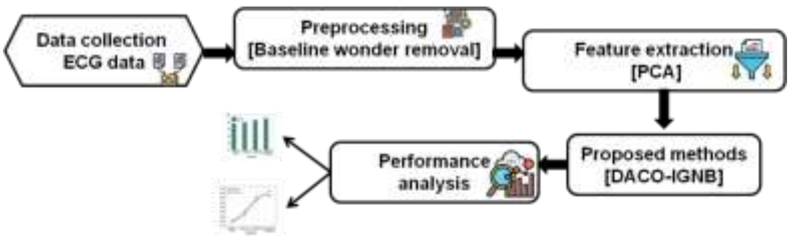

Figure 1. Proposed flow

Data collection

The data were collected from Kaggle dataset [https://www.kaggle.com/datasets/arjunascagnetto/ptbxl-atrial-fibrillation-detection,](https://www.kaggle.com/datasets/arjunascagnetto/ptbxl-atrial-fibrillation-detection) which is an open-labelled dataset of 5000 ECG heart rhythm signal data supplied by PTB-XL, to assess the efficacy of the proposed method. The data, which had a short waveform of 10 s, was captured. 3268 ECG recordings make up the dataset. The dataset includes three distinct classes: sinus rhythm (SR), atrial fibrillation (AF) and another type of arrhythmias. The record's 500 Hz sampling frequency.

Data preprocessing

Pre-processing the ECG heart rhythm signal helps to reduce noise and ensure that the results were consistent for all training cases. Baseline Wander: Breathing or electrode movement can produce low 5Hz frequency disturbances in ECG readings that skew data before processing. Numerous methods are intended to alleviate this problem. We used a wavelet-based technique by efficiently splitting the data into 0.5 to 50 Hz frequencies using Daubechies wavelets. Figure 2a depicts the normal ECG signal data and 2b shows the baseline wander removal.

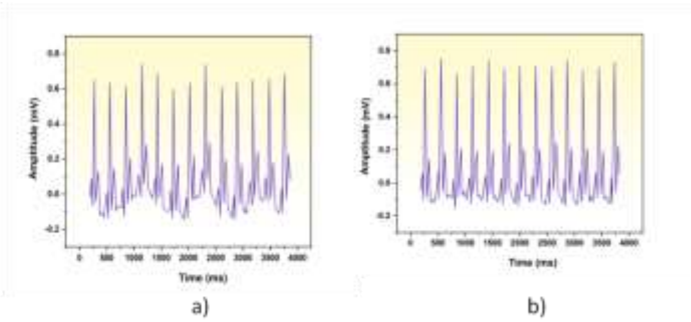

Figure 2. a) AF ECG signal data, b) baseline wander removal ECG data

Principal Component Analysis (PCA) is used for feature extraction.

After pre-processing utilized PCA to extract features for exploratory and predictive modeling, big ECG rhythm signal datasets can be divided into smaller, simpler ones to analyze using Principal Component Analysis (PCA), an unsupervised linear technique. Finding the points of maximum variance and cantering the signal data are how PCA functions. The covariance matrix is computed using either singular value decomposition or Eigenvalue decomposition to do this. In doing so, PCA efficiently eliminates noise from the ECG rhythm data and makes it easier to see underlying patterns and structures, a critical step in the accurate detection and monitoring of public health.

Take into consideration n-samples using l-variables to symbolize the ECG signal data in Equation 1.

$$Z = \begin{bmatrix} Z_{11}Z_{12} & \cdots & Z_{13} \\ Z_{21}Z_{22} & \cdots & Z_{2l} \\ \vdots \quad \vdots & \ddots & \vdots \\ Z_{m1}Z_{m2} & \cdots & Z_{m3} \end{bmatrix} = [Z_1, Z_2, \dots, Z_l] \qquad (1)$$

The ECG rhythm data standardization is accomplished using Equations 2 and 3.

$$Z'_{ji} = \frac{z_{ji} - \bar{z}_i}{\sqrt{\frac{1}{m-1}\sum_{j=1}^{m}(z_{ji} - \bar{z}_i)^2}} \qquad (2)$$

$$\bar{z}_i = \frac{1}{m}\sum_{j=1}^{m} z_{ji} \qquad (3)$$

Creating the correlation matrix as provided by, Equation 4, and the associated eigenvalue $\lambda_i$is determined by using Equation 5.

$$D = (d_{ji})_{l \times l} \qquad (4)$$

$$|\lambda F - D| = 0 \qquad (5)$$

PCA improves computing efficiency, model performance, and visualization by reducing dimensionality and noise in ECG rhythm signal data. It also prevents overfitting, which leads to more accurate and dependable public health identification.

Access health detection mechanism using Dynamic Ant colony optimized intelligent Gaussian Naive Bayes (DACO-IGNB)

To properly classify noise removed ECG signal data, ASO-GNB combines Gaussian Naive Bayes

(GNB) with Ant Colony Optimization (ACO) for improved feature selection. It is essential for developing AI-assisted public health detection and diagnostics since this synergy improves detection accuracy by dynamically adjusting feature relevance.

Intelligent Gaussian naïve bayes (IGNB)

Within the group of supervised learning methods, GNB operates based on the Naïve variation of Bayes' theorem, which postulates that each set of attributes is detached from one another, given the actual value of the ECG signal data.

A Gaussian naïve Bayes classifier makes predictions about the conditional probability spectrum of W provided $\log W$ for each factor W selected from the ECG signals and the dependent variable J is categorical. This is done by applying Bayes' rule:

$$Oq(J|\log W) = \frac{Oq(\log W|J)Oq(J)}{Oq(\log W)} \quad (6)$$

Using Bayes' rule, the conditional chance of seeing $\log \mathrm{W}$ upon highest posterior class predicting the ECG data for public health diagnosis class I is represented by Oq ($\log \mathrm{W}$) and the priori chances for J with $\log \mathrm{W}$, accordingly, are Oq (j) and Oq ($\log \mathrm{W}$). Based on this specific scenario, the following is the probability that factor W will acquire the value $\log w_j$and have activity class l:

$$Oq\big(j = l\big| \log W = \log w_j\big) = \frac{Oq(\log W = \log w_j|J = l)Oq(J = l)}{\sum_i(\log W = \log w_j|J = i)Oq(J = i)} \quad (7)$$

We empirically estimated each ECG signal parameter using maximum likelihood. By using data analysis, it was possible to determine the distribution of frequencies of the classes found in the data set or $Oq(J = l)$. M_{tot} stands for the total amount of data, whereas M_l indicates the amount of values in class J.

$$Oq(J = l) = \frac{M_l}{M_{tot}} \quad (8)$$

We calculated the conditional chance Oq $\left({}^{\log \mathrm{W}}/_{J}\right)$ for every sensitivity class k in the set of ECG signal data using the pre-estimated distribution of normality with a mean value of μ_l^* and an average standard deviation of σ_{DTC}^2.

$$Oq(\log W = \log w_j|J = l) = \frac{1}{\sqrt{2\pi\sigma_{DTC}^2}^{\,f^{-\frac{(\log w_j - \mu_l^*)^2}{2\sigma_{DTC}^2}}}} \quad (9)$$

The continuous conditional distribution of chances is computed for every input ECG signal quantity overall amplitude class by fitting the probabilities over the whole ECG rhythm dataset using GNB. Focusing on the collection with the highest posterior class associated probability value allowed us to calculate the best estimation of J.Gaussian Naive Bayes is an excellent method for precise, scalable public health diagnostics since it is resistant against extraneous features and has high interpretability and efficiency when used with ECG data for public health diagnosis.

DynamicAnt colony optimization (DACO)

When looking for a public health diagnosis, it searches the high probability class in the ECG rhythm signal data was improved by the use of DACO. To address optimization issues, the DACO algorithm mimicked the feeding habits of ant colonies. In this case, M represents the number of nodes and D is a $M \times M$ matrix that represents the path distance matrix, withstanding for the distance between j to i. The pheromone matrix amounts on paths τ_{ji}^{α} accumulate over time, forming the $M \times M$ pheromone matrix. Each ant, starting from a node, used heuristic factions η_{ji} in path selection (Equation 9),

$$\eta_{ji} = \frac{1}{c_{ji}} \quad (10)$$

At time s, the next city was selected using the probability Equation 10,

$$o_{ji}^{l}(s) = \frac{\tau_{ji}^{\alpha}(s)\eta_{ji}^{\beta}}{\sum_{iM_j^l}\tau_{ji}^{\alpha}(s)\eta_{ji}^{\beta}} if i \in M_j^l \qquad (11)$$

Guided by α and β parameters influencing pheromone and heuristic information. After completing a cycle, path pheromones are updated using Equation 11.

$$\tau_{ji}(s+1) = \rho\tau_{ji}(s) + \sum_{l=1}^{n}\Delta\tau_{ji}^{l} \qquad (12)$$

Based on ρ and the pheromone increment, it shows in Equation 12, where R is the pheromone constant and S^l is the path distance. Ant groups iterate until termination conditions yield near-optimal solutions.

$$\Delta\tau_{ji}^{l} = \begin{cases} \frac{R}{S^l}, if\ ant\ l\ moves\ through\ the\ path\ (j,i) \\ 0, other\ situations \end{cases} \qquad (13)$$

The public health detection systems' accuracy and reliability are increased by the combined action of IGNB and DACO, which optimize feature selection and effectively classify ECG rhythm signal data.

4. Results and discussion

The study evaluates classifiers for ECG-based public health detection. DACO-IGNB outperforms signal quality index- densely connected convolutional neural networks SQI-DCNN [13], Ensembling Convolutional and Long Short-Term Memory EDLM [13] and long short term memory-convolutional neural network (LSM-CNN) [13] are compared with the proposed method in terms of accuracy, precision, recall and F1-score metrics. It significantly enhances public health identification, emphasizing its effectiveness in AI-assisted diagnostic systems. Table 1 depicts the outcomes of the research.

Table 1. Outcomes of proposed and existing methods

Methods	Accuracy	Recall	Precision	F1-Score
SQI-DCNN [13]	77.16	0.76	0.77	0.76
EDLM [13]	80	0.79	0.80	0.80
LSM-CNN[13]	86.5	0.85	0.86	0.86
DACO-IGNB [Proposed]	90.9	0.89	0.88	0.87

Accuracy: Accuracy is the percentage of all the model's accurate forecasts. The accuracy results for both the suggested and current approaches are shown in Figure 3. In comparison to the SQI-DCNN (77.16%), EDLM (80%), and LSM-CNN (86.5%) approaches, the recommended DACO-IGNB method yielded a higher accuracy of 90.9%.

Precision: Precision can be defined as the relationship between all of the algorithm's projected positive situations and accurately predicted optimistic events. Figure 3 shows the results of comparative precision. The suggested DACO-IGNB approach produced a higher number of 0.88, while the SQI-DCNN achieved 0.77, the EDLM achieved 0.80, and the LSM-CNN displayed 0.86 precision.

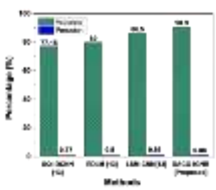

Figure 3. Outcomes of Accuracy and Precision

Recall: Recall is the system's ability to accurately identify relevant examples of the practices from among all available real instances. The recommended DACO-IGNB technique outperformed the SQI-

DCNN (0.76), EDLM (0.79), and LSM-CNN (0.85) methods with an advanced recall of 0.89.

F1-Score: An indicator that is commonly employed to assess test performance in binary identification is the F1 Score. Precision and recall are computed using harmonic means. Figure 4 presents the F1 Score and Recall values of the proposed and current methods. The DACO-IGNBapproach, which was proposed, performed better than other methods that were in use, such as SQI-DCNN (0.76), EDLM (0.80), and LSM-CNN (0.86), with an F1 Score of 0.87.

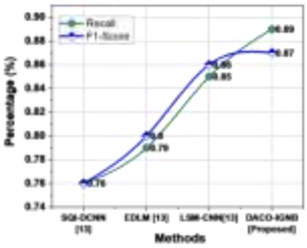

Figure 4. Evaluations of Recall and F1-Score

5. Conclusion

The study successfully identifies atrial fibrillation by applying the novel techniques of DACO-IGNB for classification and PCA for feature extraction. We used a wavelet-based pre-processing method to improve feature clarity and noise reduction in ECG data. Although DACO-IGNB enhanced feature selection and classification, PCA was used to minimize dimensionality and noise. As a result, DACO-IGNB was found to be more effective in AI-assisted public health detection and diagnostics than previous techniques SQI-DCNN [13], EDLM [13] and LSM-CNN [13] regarding F1-score (0.87), recall (0.89), accuracy (90.9%) and precision (0.88). Systems for detecting public health issues become much more accurate and reliable with this method. Future developments involve improving the device's ability to recognize a variety of heart problems as well as real-time deployment and interaction with wearable technology for continuous monitoring.

Reference

[1] A. Pan, L. Liu, C. Wang, H. Guo, X. Hao, Q. Wang, J. Huang, N. He, H.Yu, X. Lin, Wei, "Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China", *Jama* 323(19), pp. 1915-1923, 2020. https://doi.org/10.1001/jama.2020.6130

[2] P. Manickam, S. A. Mariappan, S. M. Murugesan, S. Hansda, A. Kaushik, R. Shinde, and S. P. Thipperudraswamy, "Artificial intelligence (AI) and internet of medical things (IoMT) assisted biomedical systems for intelligent healthcare", Biosensors 12,(8), p. 562, 2022. https://doi.org/10.3390/bios12080562

[3] N. Iqbal, Imran, S. Ahmad, R. Ahmad, and D.H. Kim, "A scheduling mechanism based on optimization using IoT-tasks orchestration for efficient patient health monitoring," Sensors 21(16), p. 5430, 2021. https://doi.org/10.3390/s21165430

[4] Stephen, K. V. K., Mathivanan, V., Manalang, A. R., Udinookkaran, P., De Vera, R. P. N., Shaikh, M. T., & Al-Harthy, F. R. A. (2023). IOT-Based Generic Health Monitoring with Cardiac Classification Using Edge Computing. Journal of Internet Services and Information Security, 13(2), 128-145.

[5] R. G. L. da Silva, "The advancement of artificial intelligence in biomedical research and health innovation: challenges and opportunities in emerging economies," Globalization and Health 20(1), p. 44, 2024. https://doi.org/10.1186/s12992-024-01049-5

[6] M. S. Haleem, R. Castaldo, S. M. Pagliara, M. Petretta, M. Salvatore, M. Franzese, and L. Pecchia, "Time adaptive ECG driven cardiovascular disease detector," Biomedical Signal Processing and Control 70, p. 102968, 2021. https://doi.org/10.1016/j.bspc.2021.102968

[7] Sathyanarayanan, S., & Srikanta, M.K. (2024). Heart Sound Analysis Using SAINet Incorporating CNN and Transfer Learning for Detecting Heart Diseases. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), 15(2), 152-169. https://doi.org/10.58346/JOWUA.2024.I2.011

[8] S. Mousavi, F. Afghah, and U. R. Acharya, "HAN-ECG: An interpretable atrial fibrillation detection model using hierarchical attention networks," Computers in biology and medicine 127, p. 104057, 2020.

https://doi.org/10.1016/j.compbiomed.2020.104057

[9] E. B. Panganiban, A. C. Paglinawan, W. Y. Chung, and G. L. S. Paa, “ECG diagnostic support system (EDSS): A deep learning neural network based classification system for detecting ECG abnormal rhythms from low-powered wearable biosensors,” Sensing and Bio-Sensing Research 31, p. 100398, 2021. https://doi.org/10.1016/j.sbsr.2021.100398

[10] M. Rashed-Al-Mahfuz, M. A. Moni, P. Lio’, S. M. S. Islam, S. Berkovsky, M. Khushi, and J. M. Quinn, “Deep convolutional neural networks based ECG beats classification to diagnose cardiovascular conditions,” Biomedical engineering letters 11, pp. 147-162, 2021. https://doi.org/10.1007/s13534-021-00185-w

[11] EEG Dataset collected from kaggle website https://www.kaggle.com/datasets/arjunascagnetto/ptbxl-atrial-fibrillation-detection dataset.

[12] Neamah, Kawther Abbood. "Motion stability study of double and ball pendulums." *Results in Nonlinear Analysis* 6.4 (2023): 83-96.

[13] S. Neelima, Manoj Govindaraj, Dr.K. Subramani, Ahmed ALkhayyat, & Dr. Chippy Mohan. (2024). Factors Influencing Data Utilization and Performance of Health Management Information Systems: A Case Study. Indian Journal of Information Sources and Services, 14(2), 146–152. https://doi.org/10.51983/ijiss-2024.14.2.21

[14] L. Xie, Z. Li, Y. Zhou, Y. He, and J. Zhu,”Computational diagnostic techniques for electrocardiogram signal analysis,” Sensors 20(21), p. 6318, 2020. https://doi.org/10.3390/s20216318

[15] B. Venkataramanaiah and J. Kamala, “ECG signal processing and KNN classifier-based abnormality detection by VH-doctor for remote cardiac healthcare monitoring,” Soft Computing 24(22), pp. 17457-17466 ,2020. https://doi.org/10.1007/s00500-020-05191-1

[16] S. Aziz, S. Ahmed, and M. S. Alouini. "ECG-based machine-learning algorithms for heartbeat classification,” Scientific reports 11(1), p. 18738, 2021. https://doi.org/10.1038/s41598-021-97118-5

[17] S. Liaqat, K. Dashtipour, A. Zahid, K. Assaleh, K. Arshad, and N. Ramzan, “Detection of atrial fibrillation using machine learning approach,” Information 11(12), p. 549, 2020. https://doi.org/10.3390/info11120549