The Scientific Temper (2024) Vol. 15 (3): 2968-2977
E-ISSN: 2231-6396, ISSN: 0976-8653
Doi: 10.58414/SCIENTIFICTEMPER.2024.15.3.70
https://scientifictemper.com/

RESEARCH ARTICLE

An optimized cardiac risk levels classifier based on GMM with min- max model from photoplethysmography signals

Divya R.[1*], Vanathi P. T.[2], Harikumar R.[3]

Abstract

As per the latest study, coronary artery disease and hemorrhagic stroke are the predominant factors contributing to over 80% of cardiovascular diseases (CVDs). To reduce the mortality rate due to CVDs, researchers are proposing techniques for early detection of these CVDs. For the preliminary investigation of cardiovascular disease, photoplethysmography (PPG) can be used. Using PPG signals, it is possible to infer the risk levels like CVD with low risk, CVD with medium risk and respiratory disorder. To classify the risk levels of CVD, a model incorporating a Gaussian mixture model (GMM) classifier with a min-max decision model has been implemented. The proposed model resulted in better performance than existing classifiers like logistic regression-GMM (LR-GMM), detrend fluctuation analysis (DFA) and Cuckoo search algorithm (CSA) using the min-max model. Based on the results, GMM reflects a peak 95.9% classification accuracy with minimal false alarms of 7.1 and 0.99% miss classification when compared to other post-classifiers.

Keywords: Gaussian mixture model, Min-max decision model, Cardiovascular disease, Photoplethysmography, Singular value decomposition.

Introduction

Cardiovascular disease (CVD) ranks among the foremost global causes of mortality. It is reported that approximately 80% of deaths associated with CVD occur in developing countries like India (Gupta *et al*., 2013; Sowmiya *et al*.,2023). CVDs are chronic diseases that occur due to the lack of exercise, smoking, diabetes, unhealthy eating, age and family history. There is a demand for diagnosis and treatment of this chronic disease increases. Therefore, it is necessary to make efforts for early diagnosis of cardiovascular disease using effective risk prediction methods (Lin *et al*., 2013; Divya & Vanathi, 2019). In order to sense the CVD at the initial stage, a low-cost, simple diagnostic tool called photoplethysmography (PPG) may be used. This is a non-obtrusive, optical measurement approach for detecting the fluctuations in blood volume that occur in the vasculature during each heartbeat (Elgendi, 2012). Blood vessels have flexibility and elasticity to handle the flow of blood. Measuring these properties enables us to identify the risk of CVDs.

In general, it is inferred that PPG signals afford important data regarding the above CVD (Allen, 2007). PPG technology featuring a light emitter diffused into the tissue, utilizes a photosensor for quantifying how much light is reflected back. The variations of the blood volume are equivalent to the reflected light. PPG signal is taken from the index finger because finger PPG is easy to record, and it is a commonly used technique in medical applications. By analyzing the recordings of the PPG signals, doctors will get a better understanding of a patient regarding their medical conditions before going for surgery. On the other hand, it is not possible to examine the recordings of all the generated signals (Sukor *et al*., 2011). Thus, it is critical to design and build an intelligent system so that these automated classifiers increase the speed of doctor's pre-surgical assessments and save medical expenses.

[1]Department of Electrical and Electronics Engineering, PSG College of Technology, Coimbatore, Tamil Nadu, India.

[2]Department of Electronics and Communication Engineering, PSG College of Technology, Coimbatore, Tamil Nadu, India.

[3]Department of Electronics and Communication Engineering, Bannari Amman Institute of Technology, Sathyamangalam, Tamil Nadu, India.

***Corresponding Author:** Divya R., Department of Electrical and Electronics Engineering, PSG College of Technology, Coimbatore, Tamil Nadu, India., E-Mail: divyayuvaraj2012@gmail.com

How to cite this article: Divya, R., Vanathi, P. T., Harikumar, R. (2024). An optimized cardiac risk levels classifier based on GMM with min- max model from photoplethysmography signals. The Scientific Temper, **15**(3):2968-2977.

Doi: 10.58414/SCIENTIFICTEMPER.2024.15.3.70

Source of support: Nil

Conflict of interest: None.

Received: 07/07/2024 **Accepted:** 08/08/2024 **Published** : 30/09/2024

Literature Review

The following are some of the significant contributions and advancements in the analysis and application of PPG signals for cardiovascular health assessment and disease detection: The study assessed the characterization of cardiovascular dynamics non-invasively from the pulsation of blood volume using PPG device. A model was employed to identify and extract the periodic non-sinusoidal component present within the data series (Bhattacharya *et al*., 2001). PPG measurements for medical applications and computerized pulsatile analysis approaches for the diagnosis of vascular disease from various features of the pulse are described (Allen *et al*., 2006). There was an approach to detect ventricular premature beats in Photoplethysmography signals involving using a sliding window technique to extract features, with a particular emphasis on peak-to-peak internals (PP) and PPG pulse power. These feature vectors were tested using an artificial neural network (ANN) with both non-linear and linear outputs. Better classification results were obtained for PVC detection using a non-invasive approach compared to ECG based detection system (Sološenko *et al*., 2015).

In literature, the heart rate variability (HRV) was estimated by analyzing the duration between peaks in photoplethysmography waveforms (Selvaraj *et al*., 2008). The results showed that PP variability was more accurate than the ECG-derived RR intervals for the identification and classification of arrhythmia. Heart abnormalities were identified by analyzing PPI results from PPG signals of healthy and abnormal patients. The results indicated that abnormal heartbeats have higher PPI values compared to normal heartbeats (Umadi *et al*., 2016). PPG signals were analyzed to diagnose the cardiac risk of the patient without any support of an added instrument. Arterial stiffness is a key factor for cardiac disorder and the augmentation index is used to measure Arterial stiffness, which is derived from the pulse wave (Manimegalai *et al*., 2012).

A methodology was put forth to categorize the patients between different CAD conditions by deriving a range of temporal features inherent in PPG signals, employing the K-Nearest Neighbor classifier for classification (Hosseini *et al*., 2015). A system for predicting cardiac risk was also developed to categorize different levels of cardiovascular diseases using PPG signals. This model involved extracting wavelets, SVDs and statistical features and then employing GMM and softmax classifier (SDC). SVD, along with statistical features and SDC, shows better performance, achieving a high accuracy rate (Divya *et al*. 2020). Recently, a novel CVD risk detection model, CRDHHO-EL, was introduced, incorporating the Harris Hawks algorithm alongside Ensemble Learning for analyzing PPG signals (Divya *et al*., 2023).

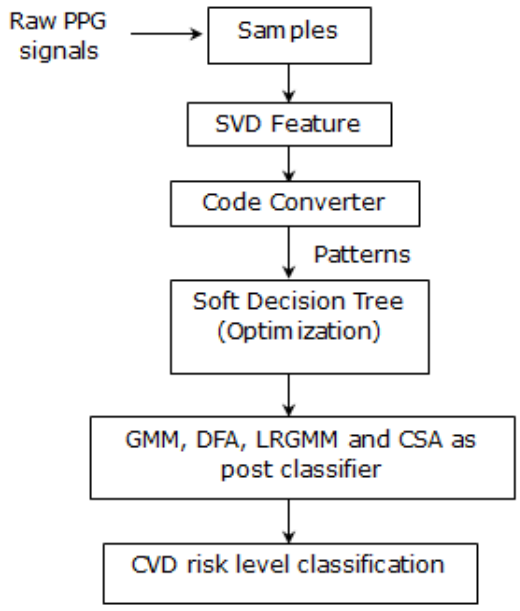

Figure 1: Workflow of heart health risk prediction system

Proposed Methodology

The main goal of the investigation involves assessing various risk phases of cardiac vascular disease using photoplethysmography signals. The PPG signal for this study has been obtained from the capnobase pulse oximetry benchmark dataset which contains 41 cases (29 children and 13 adults) of 8 minutes duration comprising 720 segments. The recordings of regular ventilation, inadequate ventilation, spontaneous breathing, rebreathing, hyper and hypo ventilation, cardiac oscillations and apnea were observed in these cases. Also, the dataset contains instantaneous heart rates derived from PPG pulse peaks and ECG. The segments of the PPG signals have equal intervals and have 200 samples per segment (Karlen, 2010).

In this study, the feature vector, namely singular value decomposition (SVD) was derived from the PPG signals and this represents the sole feature extracted from the signal. Then with the help of the expert's knowledge, three linguistic risk phases, namely CVD low risk (P), CVD medium risk (Q) and respiratory problem (R), has been determined.

Then, the above risks are encoded as strings of alphabets that relate each and every signal parameter. The encoded strings are optimized using a soft decision tree (SDT) to minimize the computational complexity and to approximate the global decision region. Subsequently the output of the SDT is used as input for the Gaussian mixture model (GMM) for categorizing health status stratification using PPG. In comparative analysis and performance evaluation of the GMM, several other methods were employed to classify PPG signals, including Logistic regression-GMM (LR-GMM), Detrend fluctuation analysis (DFA) and Cuckoo search algorithm (CSA). The flow chart for CVD risk prediction system is presented in Figure 1. Each element within the flow chart is thoroughly elucidated in the subsequent subsections.

Singular Value Decomposition

Singular Value Decomposition (SVD), initially formulated in 1970, has been used in extensive applications like feature extraction, image compression and texture processing (Karlen *et al.*, 2013). In this paper, SVD feature derived through PPG signals helps in determining cardiac and respiratory risk groups. Let $Y=[y_1, y_2, y_3, \ldots\ldots\ldots\ldots y_n]$ represents a matrix with dimensions $m \times n$. The SVD decomposes Y into $Y=U\Sigma V^T$, where U denotes an r x r orthogonal matrix, V^T also represents an m x m matrix and ∑ represents r x m diagonal matrix, $\Sigma = diag(\sigma_1, \sigma_2, \sigma_3, \ldots\ldots\ldots\ldots \sigma_p)$. ∑ is a singular value matrix, U and V are unitary matrices of Y. If the condition is $\sigma_1 \geq \sigma_2 \geq \sigma_3, \ldots\ldots\ldots\ldots \geq \sigma_p$ also if the matrix Y possesses a rank h < p, the p-k singular values at the end of the sequence becomes zero. The SVD can be written as $Y = \Sigma_{k=1}^{h} \sigma_k u_k v_k^T$ (Rajaguru & Prabhakar, 2016) and (Sadasivan & Dutt, 1996). By means of SVD, vectors taken in one space can be transformed into another space.

The merit of using SVD is that it can combine two dissimilar uncertainty representations as a whole uncertainty. These combined uncertainty measures encompassing aspects like probability possibility manifest as a group of vectors with varying units within a principal space. It is essential to take this feature because the uncertainty measures contain different units (cardiac risk levels) (Lee & Hayes, 2004). Only SVD feature has been extracted from each segment (fragment) of the photoplethysmography signal from the capnobase dataset.

With 720 partitions present within the captured PPG signal, the data underwent compression from 1,44,001 samples to 720, reducing its size significantly. Labeling of the segment can be done easily using the annotation of the SVD. Normal and abnormal segments can be easily identified using the SVD values. If the threshold value of the SVD is £ 40, then the particular segment is considered as normal; otherwise, it is labeled as abnormal. Segments with SVD thresholds falling between 41 to 50 are categorized as CVD with risk level 1, while those in the range of 51 to 60 are identified as CVD of risk level 2. Segments with a threshold exceeding 60 are attributed to respiratory disorders among abnormal segments.

Table 1 displays the extracted SVD feature from the PPG signals along with the corresponding segment labels. The total number of subjects considered for experimentation is 20 from the entire IEEE benchmark capnobase dataset. Figure 2 represents the subject characterized by normal segments within the PPG signal, along with a histogram representation showcasing the derived SVD feature from this normal signal.

Similarly, Figure 3 shows the inconsistency in the PPG signal indicating heart irregularities and the histogram representation of the SVD feature extracted from this abnormal signal. Figure 2 (b) and 3(b) depict the first 2000 samples of the normal and abnormal segments. Then, the obtained SVD feature vector is normalized using the min-max method. Min-max normalization is determined using to Equation 1.

$$Z_{norm} = \frac{a - \min(a)}{\max(a) - \min(a)} \quad (1)$$

Thus, all the data will be mapped in the range of 0 and 1 (Klema & Laub, 1980).

Code converter

Identifying the optimal risk threshold for individuals with CVD is crucial. This knowledge is vital for the advancement of automated systems capable of precisely categorizing the observed patient's CVD risk levels. In order to categorize the risk levels of patient, a code-converting process is carried

Table 1: Representation of sample SVD feature from photoplethysmogram data set and its labeling

Signal	*Fragment 1*	*Fragment 2*	*Fragment 3*	*Fragment 4*	*Fragment 5*	*Fragment 6*
Norm 1	32.686	35.651	34.474	26.253	31.946	32.105
Abnor 1	83.669	79.437	74.228	84.466	79.697	80.556
Norm 2	63.566	60.171	59.307	51.773	57.501	50.830
Norm 3	27.468	27.528	26.114	19.288	26.538	26.507
Abnor 2	100.896	85.514	73.426	89.087	95.664	75.989
Norm 4	26.371	27.124	27.776	28.456	26.283	24.293
Norm 5	12.054	11.717	13.695	14.217	13.850	14.097
Abnor 3	60.608	51.260	60.774	65.406	62.694	51.650
Abnor 4	77.892	61.010	58.705	64.422	79.831	68.853
Norm 6	25.944	28.130	29.813	26.358	23.963	28.712
Abnor 5	50.660	54.750	64.491	49.340	47.243	48.087
Abnor 6	74.088	83.531	79.460	73.487	83.804	82.497
Abnor 7	20.657	29.700	29.157	30.489	28.321	20.349
Abnor 8	69.463	67.307	67.634	76.927	54.699	54.795
Abnor 9	53.433	55.274	57.865	54.425	55.522	49.908

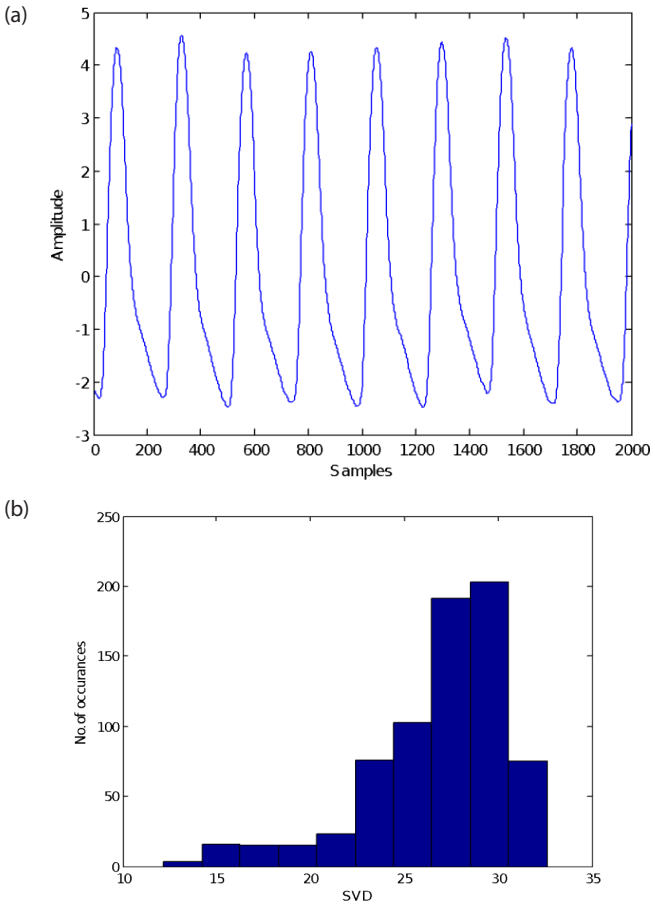

Figure 2: a) Normal segments of the PPG signal and its corresponding; b) histogram for SVD feature vector extracted from healthy subject

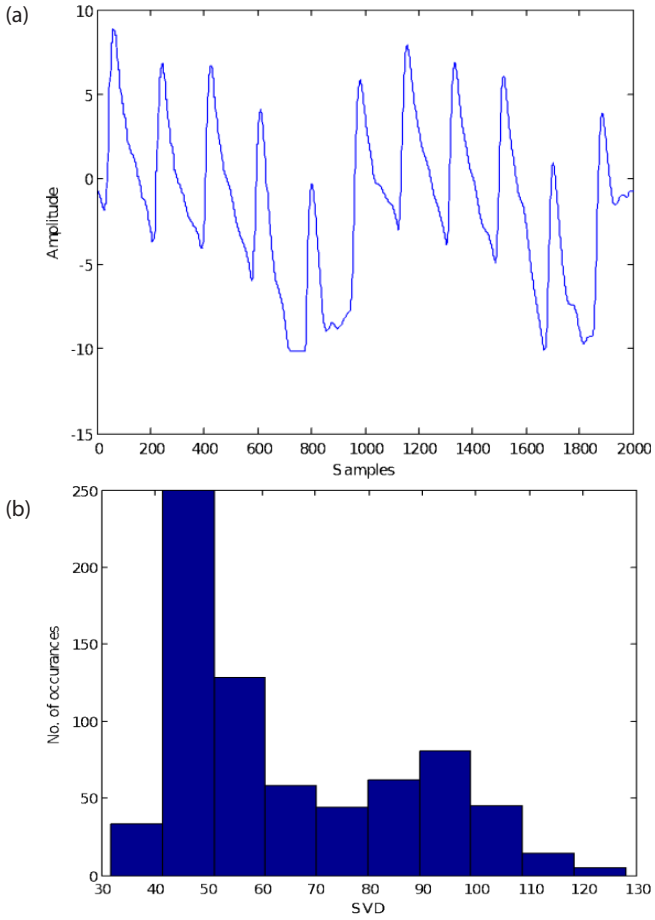

Figure 3: a) Abnormal segments of PPG signal along with its corresponding; b) histogram for SVD feature extracted from the abnormal subject

out to convert the PPG signals into encoded strings or patterns (Rahmad *et al*., 2018). Table 2 depicts the risk level classification using alphabetical representation.

The four risk levels are tabulated in descending order. The highest risk level is R and the lowest risk level is P. The patient's outputs are labeled as S, P, Q and R and eight labeled outputs form a pattern. Then, each pattern is given to the soft decision tree for optimization.

Soft decision tree model (min-max) for optimization

The PPG signals are intrinsically complex because of their instability and often non-linear nature. The main aim is to unify the PPG risk level depiction, with symbolic decision tree (SDT) models. This method is simple to apply. Decision trees are appealing because it has ability to decompose an intricate procedure into a set of more manageable steps, resulting in an outcome that is more easily interpretable. Unnecessary computations can be reduced in an SDT by testing a sample against only definite subsets of classes. In contrast, single-stage classifiers require testing every observation across all categories, thus potentially diminishing the efficacy of the classifier.

Table 2: Depiction of risk assessment levels

Risk assessment grade	Symbol
Healthy	S
CVD level 1	P
CVD level 2	Q
Respiratory Disorder	R

There are also some shortcomings related to SDT, which are as follows; i) an increase in the number of classes causes an increase in the number of terminals, which leads to memory constraint and more search time. ii) There may be accumulation of errors from stage to stage in a large tree. Thus, simultaneous optimization of both efficiency and accuracy is difficult. iii) Formulating the most efficient decision tree may pose challenges. The effectiveness of SDT is significantly influenced by the quality of the tree's design (Harikumar & Vijayakumar, 2015).

SDT Optimization Algorithm

The main goal of SDT is, to classify appropriately any number of normal and abnormal training samples with high accuracy.

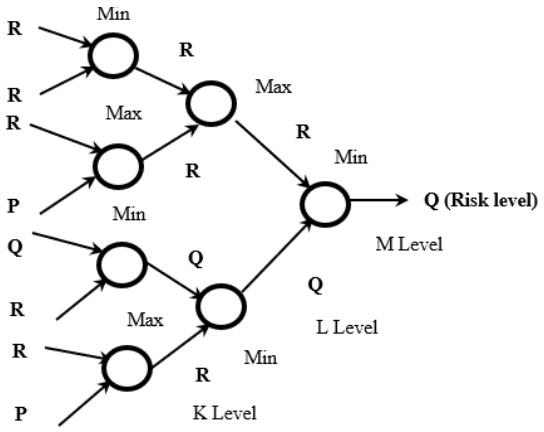

Figure 4: Optimization of cardiac risk category using min-max decision approach

There are several heuristic approaches for designing SDT, such as growing – pruning, top-down approaches, bottom-up approaches and hybrid approaches (Harikumar *et al*., 2012). In the bottom-up approach, employing a distance measure such as Mahalanobis distance involves calculating pairwise distances among classes that have been predefined. The two classes with lesser distances are combined in each step to create a cluster. Within each cluster, the mean vector and covariance matrix are computed based on the training samples from the classes within that cluster. The procedure is reiterated until a single cluster remains at the root.

In the top-down methodology, the procedure initiates from the root node, and by applying a splitting rule, classes are partitioned until a termination threshold is attained. The primary challenges in this methodology include finding conditions for node splitting ii) endpoint rules iii) assigning labels to the final nodes. The hybrid approach utilizes the bottom-up strategy to steer the top-down method. An iterative Growing-Pruning approach divides the dataset into two nearly equivalent-sized subsets. It systematically constructs the tree using one subset and prunes with the other subset in an iterative manner. Then, successfully exchange the roles of the two subsets. Starting at the end nodes, this methodology progresses upward through the tree, trimming without eliminating terminal nodes (Salembier & Garrido, 2000).

For a soft decision tree, it is crucial to make judicious choices regarding the tree topology, feature subsets and determining the decision rule for every inner node. The decision tree was constructed using a bottom-up approach. There are two decision techniques used at the node level of SDT, namely Min-Max and Max-Min. In this work, the realization of 8×1 matrix into one optimum cardiac risk level utilizing soft decision tree optimization with three tiers using the Min-max decision method has been carried out. By means of the Min-Max procedure, distances among priori-defined classes undergo computation in a pairwise manner. During each stage, the pair of classes linked to a nodal point are amalgamated, forming a consolidated group. This iterative merging procedure persists as long as more than a group remains at the root, signifying the ultimate formation of optimized cardiac health patterns. 8-4-2-1 model of SDT has been chosen for optimizing input patterns. Min-max decision approach was applied at each node in the 8-4-2-1 SDT model.

Table 3: Representation of different patterns of patient 1 and optimization using min-max

Patterns	*K level*	*L level*	*M level*
QPQRRPPP	PRPP	RP	P
RRRPQRRP	RRQR	RQ	Q
PQRRRQPR	QRQR	RQ	Q
RQPQRQRQ	QQQR	RQ	Q
PRRQPPRR	PRPR	RP	P

In this model, the terminal leaf nodes in the group (8x1) are referred as elements. The subsequent tier of a tree is designated K, featuring four decision nodes, succeeded by L tier has two decision nodes of two. The final tier denoted as E has a sole node and act as the tree's root. Subsequent decisions are performed for every tree node using row-wise optimization called min-max method.

Consider two consecutive leaves, denoted as ri and ri+1, to be determined at the next level, labeled Ki. The decision criteria are as follows:

For K_i, compute the minimum value between r_i and r_{i+1}. Simultaneously, determine K_{i+1} by finding the maximum value between r_{i+2} and r_{i+3}. Proceed to the next level labeled L_i.

- At L_i, calculate the maximum value between K_i and K_{i+1}, and determine L_{i+1} by finding the minimum value between K_{i+2} and K_{i+3}. Progress to the subsequent level labeled M_i.
- For M_i, compute the minimum value between L_i and L_{i+1}. Concurrently, determine M_{i+1} by finding the maximum value between L_{i+2} and L_{i+3}. Move on to the next level, labeled N.

The above process is illustrated with an example in Figure 4.

Table 3 depicts the few patterns of patient 1 and its optimization by means of the min-max procedure.

The optimized results obtained from the soft decision tree are supportive in identifying the cardiac risk levels from the PPG signals and these results are given to the different post-classifiers.

Classifiers

Various classifiers employed in categorizing the risk levels of cardiovascular diseases are explained as follows.

Logical Regression Gaussian Mixture Model Classifier

Cardiovascular disease risk classification was done using logistic regression Gaussian mixture model (LR-GMM)

from PPG signals. If random variable y is Gaussian, then the probability density function (PDF) modeled in the classical GMM is represented as

$$P_{G,b,\Sigma,w}(y) = \sum_{g=1}^{G} \phi_{w,g}\, M_{b_g,\Sigma_g}(y) \quad (2)$$

where ϕ_n represents the number of mixture components and $M(y|b_g, g = G|$ are the component of Gaussian densities of the covariance matrix Σ (variance 'σ', for univariate Gaussian distribution') and mean 'b'.

The general form for the density of each component is

$$M((y|b_g, \Sigma_g) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(y-b)^T \Sigma^{-1}(y-b)\right] \quad (3)$$

Mathematically definition for mixture weights from a G- tuples (w_1, w_G) with a particular logistic approach as (Montuelle & Le Pennec, 2014)

$$w, g = \frac{e^{w_g}}{\sum_{g}^{G} e^{w_g}} \quad (4)$$

A model is examined where both the mixture weights and means are based specifically upon a covariate. The dataset comprises '*k*' pair of random variables, denoted as $((V_j, U_j))_{1\le j\le n}$, where the covariates V_jS are independent and U_iS are independent given the V_jS. The conditional density $s_0(.|k)$ is assessed regarding Lebesgue measure of U relative to V. The conditional density can be represented through a Gaussian regression ensemble with changing logistic weights and is denoted in the form of

$$S_{G,b,\Sigma,w}(y|k) = \sum_{g=1}^{G} \phi_{w(k),g}\, M_{b_g(k),\Sigma_g}(y), \quad (5)$$

where $(b_1, \ldots\ldots\ldots\ldots, b_G)$ and $(w_1, \ldots\ldots\ldots\ldots, w_G)$ represents *F* tuples of the selected parameters across the entire dataset. Subsequently, the parameters b_g and w_g are assessed, accompanied the covariance matrix Σ_g with the overall class count denoted by g. Thus, minimum error is attained by comparing true and estimated conditional density.

Gaussian Mixture Model Classifier

The Gaussian mixture model (GMM) classifier serves the purpose of classifying large collection of N-dimensional signals. It is a parametric model comprising a vast array of Gaussian mixture densities. GMM as a whole is characterized by incorporating mean vectors, covariance matrices and weight coefficient for each Gaussian component. During the training process, the expectation maximization (EM) algorithm iteratively optimizes these parameters.

At times, GMM parameters are also determined through Maximum APosteriori (MAP) using a fine-grained model. Often model setup is chosen based on the total dataset present for GMM parameters estimation. In order to represent the comprehensive feature density, Gaussian components collaborate with each other and it's worth noting that even when the features exhibit statistical dependence, full covariance matrices are not needed. Maximum likelihood (ML) estimation is one of the techniques which are used to assess the parameters of GMM. The ML estimation is used for ascertaining the model parameters that maximize the likelihood of GMM with the given training dataset.

Given the K training vectors Y= {y_1, y_2……,y_K}, assuming each vector to be independent, GMM likelihood is expressed as

$$p(Y|c) = \prod_{y=1}^{K} p(Y_k|c) \quad (6)$$

The aforementioned expression signifies a function of parameter' c' that is non-linear, rendering direct maximization impractical.

Here c is utilized to denote the parameters as a whole, in the subsequent manner.

$$c = \{\phi_g, b_g, \Sigma_g\}, g = 1,2\ldots\ldots\ldots N \quad (7)$$

Where b_g signifies the mean vector and the covariance matrix denotes Σ_g. A condition $\sum_{i=1}^{N} \phi_g = 1$ must be satisfied for mixture weights. By means of the iteration process using EM algorithm, the best ML parameters can be easily obtained. This EM algorithm starts with some initial model 'c' and then continues to update 'c' iteratively until a convergence value is identified. Thus, to assess a new model $\bar{c}$, such that $p(Y/\bar{c}) \geq p(Y/c)$ (Permuter *et al*., 2016). GMM are capable of signifying distributions of huge classes of samples and they have the characteristics of forming smooth approximations to random shaped densities. However this algorithm is of a first-order nature, resulting in a gradual convergence to a fixed-point solution. Also, it is prone to misclassifying local maxima and demonstrates sensitivity to outset conditions.

Detrend Fluctuation Analysis

Detrend fluctuation analysis (DFA) is used in biomedical applications to deal with complex and non-stationary signals. DFA serves as an essential tool in identifying both long-range and short-range patterns within a distorted signal. The fractal dimension derived through the DFA technique shows valuable measures within different domains, including cardiovascular and hemodynamic variations, earthquake data, traffic fluctuation etc. This statistical technique is used to distinguish between normal and unhealthy subjects (Telesca *et al*., 2008). Random walk DFA involves a two-step algorithm. Step 1: Examine the time sequence Z (i) with i ranges from 1 to N. Here N represents the length of the time series. The time sequence Z (k) is shifted by the average value of the series Z_{avg} and cumulatively summed.

$$x(k) = \sum_{i=1}^{k} [Z(i) - Z_{avg}] \quad (8)$$

Step 2: Within every segment, a least-squares line applies to the data, indicating an underlying pattern of the specific segment. Fitting polynomial is represented as $x_{\Delta b}(k)$.

The computation of mean square variation for this detrended and integrated times series involves according to equation 8.

Table 4: Average performance analysis of DFA, LR-GMM, GMM and Cuckoo classifiers

Classifier	*PC (%)*	*MC (%)*	*FA (%)*	*PI (%)*	*Sensitivity (%)*	*Specificity (%)*	*Accuracy (%)*
GMM	91.81	0.99	7.19	90.66	92.81	99	95.91
DFA	81.29	0	18.71	76.01	81.30	100	90.65
LR-GMM	80.05	0.97	18.98	73.02	81.03	99.03	90.03
Cuckoo	85.47	1.88	12.65	82.54	87.35	98.13	92.74

$$F(\Delta b) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}[x(k) - x_{\Delta b}(k)]^2} \quad (9)$$

By averaging all the segments, $F(\Delta b)^2$ can be observed as the overall detrended function.

If F(Δb) exhibit power-law function in relation to Δb, then $F(\Delta b) \propto (\Delta b)^{\alpha}$. The scaling exponent signifies the slope derived from a fitting $\log F(\Delta b) to \log \Delta b$, can be used to describe the oscillations under these circumstances. In this study, the autocorrelation of the input is represented by the scaling exponent.

Cuckoo Search Algorithm

Cuckoo search algorithm (CSA), the swarm-intelligence-inspired algorithm, is crafted by mimicking cuckoo's breeding habits. CSA incorporates three rules that mimic the behavior exhibited by cuckoos, adapting them for computational algorithms (Fister *et al.*, 2013)

1) Every cuckoo deposit a single egg into a randomly selected nest at any given moment.

2) Nests containing high-quality eggs are carried forward to subsequent offspring.

3) The quantity of candidate nests is predetermined, with likelihood of encountering a cuckoo-laid egg governed by a value denoted as p_a, falling within the interval (0, 1). Under these circumstances, the host bird can opt to abandon the nest, construct another or discard the egg altogether.

Every egg within a nest signifies a potential candidate solution. While nests may contain multiple eggs, indicating a range of solutions, a cuckoo typically contributes only a single egg to a nest in general. The function of cuckoo search is geared towards producing new and prospectively enhanced solutions so that worse solutions will be replaced with better solutions within the existing population of nests. Assessment of the approach is made through the objective function pertinent to addressing the issue.

Performance Measures

In this work, GMM with the min-max method has been used to identify normal as well as anomalous segments of PPG signal. Within the abnormal segment, the model also identifies CVD with low risk, CVD with medium risk and respiratory illness from PPG signal. In this section, the relative performances of the classifiers are examined with a discussion of the corresponding metrics provided below.

True Positive (TP)

Cases labeled healthy and correctly identified as healthy subject.

False Positive (FP)

Cases categorized normally but misclassified individuals with CVD.

Table 5: Assessment of cardiac risk level classification in terms of PC, MC, FA and PI for all post-classifiers

Classifier	*Category*	*PC (%)*	*MC (%)*	*FA (%)*	*PI (%)*
DFA	S	78.961	0	21.040	72.959
	P	82.502	0	17.498	76.828
	Q	76.651	0	23.350	69.282
	R	87.057	0	12.943	84.984
LR-GMM	S	80.261	0	19.739	73.052
	P	79.688	3.541	16.772	72.959
	Q	90.280	0.347	9.372	88.620
	R	69.979	0	30.024	57.431
GMM	S	84.380	0	15.620	81.470
	P	94.584	2.499	2.914	94.214
	Q	91.843	0	8.157	90.776
	R	96.430	1.487	2.080	96.184
Cuckoo	S	86.459	7.5	6.041	84.253
	P	75.314	0	24.687	68.604
	Q	85.768	0	14.232	83.327
	R	94.346	0	5.654	93.979

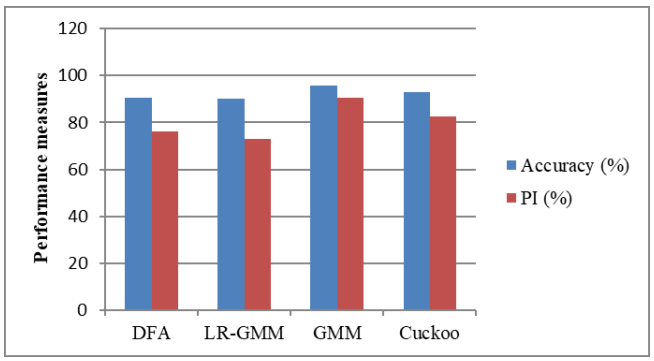

Figure 5: Average accuracy and performance index of DFA, LR-GMM, Cuckoo and GMM

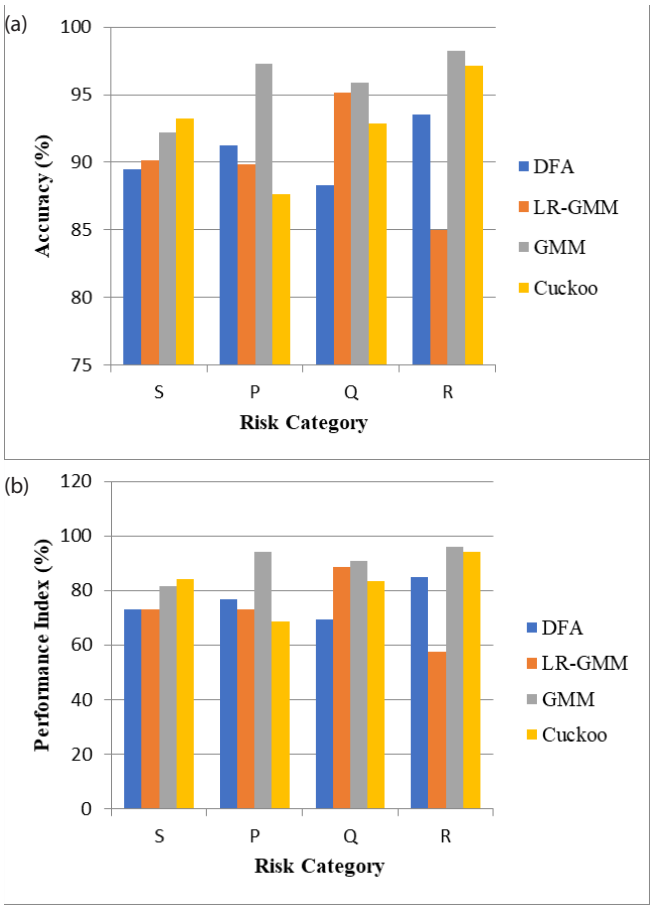

Figure 6: a) Accuracy and b) performance index for classifying various CVD risk types using all post-classifiers

Table 6: Assessment of cardiac risk level classification in terms of sensitivity, specificity and accuracy for all post-classifiers

Classifier	*Category*	*Sensitivity (%)*	*Specificity (%)*	*Accuracy (%)*
DFA	S	78.964	100	89.482
	P	82.502	100	91.251
	Q	76.656	100	88.328
	R	87.057	100	93.529
LR-GMM	S	80.264	100	90.132
	P	83.233	96.458	89.845
	Q	90.627	99.653	95.140
	R	69.984	100	84.992
GMM	S	84.380	100	92.190
	P	97.084	97.5	97.292
	Q	91.843	100	95.922
	R	97.917	98.513	98.215
Cuckoo	S	93.959	92.5	93.230
	P	75.323	100	87.661
	Q	85.768	100	92.884
	R	94.346	100	97.173

Table 7: Estimation of MSE obtained by all post classifiers

Classifier	S	P	Q	R	Average
DFA	1.66E-05	1.45E-05	2.14E-05	5.45E-06	1.4474E-05
LR-GMM	2.00E-05	1.79E-05	4.16E-06	3.42E-05	1.9068E-05
GMM	7.45E-06	6.20E-07	2.82E-06	4.07E-07	2.8241E-06
Cuckoo	5.36E-06	2.09E-05	5.83E-06	3.21E-07	8.1E-06

True Negative (TN)

Instances pertaining to CVD patients correctly classified as such.

False Negative (FN)

Instances from CVD subjects incorrectly classified as belonging to normal subjects.

Based on the values of these parameters, the following mathematical formulas are employed to calculate perfect classification (PC), missed classification (MC), accuracy, false alarm (FA), sensitivity, specificity and performance index (PI) (Ramachandran *et al*., 2020).

$$\mathrm{PC}(\%) = \frac{\mathrm{TN+TP}}{\mathrm{TP+TN+FP+FN}} \times 100 \quad (10)$$

$$\mathrm{MC}(\%) = \frac{\mathrm{FN+FP}}{\mathrm{TP+TN+FP+FN}} \times 100 \quad (11)$$

$$FA(\%) = \frac{\mathrm{FP}}{\mathrm{TN+FP}} \times 100 \quad (12)$$

$$\mathrm{Sensitivity}(\%) = \frac{\mathrm{PC}}{\mathrm{PC+FA}} \times 100 \quad (13)$$

$$\mathrm{Specificity}(\%) = \frac{\mathrm{PC}}{\mathrm{PC+MC}} \times 100 \quad (14)$$

$$\mathrm{PI}(\%) = \frac{\mathrm{PC-MC-FA}}{\mathrm{PC}} \times 100 \quad (15)$$

$$\mathrm{Accuracy}\ (\%) = \frac{\mathrm{Sensitivity+Specificity}}{2} \times 100 \quad (16)$$

Results and Discussions

The performance evaluation for GMM, CSA, DFA, LR-GMM classifiers are computed and tabulated. Table 4 shows a comprehensive analysis of the effectiveness of different types of classifiers.

From Table 4, it is inferred that the accuracy of GMM is highest at 95.9% when compared with the other types of post-classifiers. The accuracy of cuckoo algorithm comes next to the GMM and it is found to be 92.73%. If the false alarm rate parameter is considered, GMM provides less false alarm of 7.2%. This is followed by Cuckoo search algorithm having an average FA of about 12.6%. Thus, it is concluded that the GMM stands as the optimal choice for discerning cardiac risk levels based on PPG signals. Figure 5 depicts the average classification accuracy and performance index of all post-classifiers.

Tables 5 and 6 showcase the evaluation of LR-GMM, DFA, Cuckoo and GMM classifiers in detecting various risk categories related to CVD. Figure 6 provides a summary of the performance measures based on accuracy and performance index of various types of classifiers at different risk levels.

As from Tables 5 and 6, it is obvious that GMM attains better classification accuracy of 97.29, 95.92 and 98.22% in the classification of P, Q and R, respectively, when compared to other classifiers. However, the accuracy of GMM falls by 1% in the classification of normal (S) than Cuckoo algorithm. Even though accuracy for GMM decreases for risk category S, the overall classification performance is better for GMM than the other classifiers.

Analysis using error rate

Mean square error (MSE) is a measure that provides insight into the overall accuracy of a model by evaluating its ability to predict target values (T_i) based on observed values (O_i).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(O_i - T_i)^2 \quad (16)$$

Table 7 gives average MSE as well as MSE at each risk phases obtained by LR-GMM, DFA, Cuckoo and GMM classifier. As by rule of thumb, MSE estimation of 0.001 will be accepted level for a specific classification of the classes.

From Table 7, it is observed that MSE value is found to less for GMM classifier of 6.2E-07 and 2.82E-06 for the signals with risk stages of P and Q, respectively. In contrast, MSE value is found to be very low for the cuckoo search algorithm for normal and signals with respiratory disorders. But it is observed that the average MSE for GMM classifier reached the appreciable MSE value of 2.82E-06 among other classifiers. Hence the classification accuracy is better for GMM classifier.

Conclusion

It's crucial to formulate efficient methods for predicting the risk of CVD in individuals before noticeable symptoms manifest. The paper aims to categorize the risk level of CVD patients based on PPG signals striving for a high classification accuracy, minimal false alarms and less misclassification rate. The SVD feature was extracted from the PPG signal and optimized using a min-max decision model. The output from the min-max decision model is applied to LR-GMM, DFA, CSA and GMM classifiers. From the results, it is found that GMM classifier was resulted in an improvement of accuracy by 3.17%, a reduction in miss classification rate by 0.88% and a reduction in false alarm rate by 5.5% compared to cuckoo search algorithm. In terms of high-risk category namely respiratory disorder and cardiac risk level 2, the GMM classifier outperformed the classifier utilizing CSA in accuracy and perfect classification rate. With respect to missed classification rate, the DFA-based classifier surpasses all other classifiers and it is considered the best. The performance of the classifier using LR-GMM resulted in a very high value of false alarm, whereas GMM was found to have a low false alarm. As an extension of this work, diabetes could be detected using the PPG signal and by applying deep learning algorithms like convolution neural networks (CNN).

Acknowledgment

The authors acknowledge Management and Principal for supporting the conduction of research work.

References

Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, **28(3)**, 1-39. https://doi.org/10.1088/0967-3334/28/3/R01.

Allen, J., Overbeck, K., Stansby, G., & Murray, A. (2006). Photoplethysmography assessments in cardiovascular disease. *Measurement and Control*, **39(3)**, 80-83. https://doi.org/10.1177/0020294006039003.

Bhattacharya, J., Kanjilal, P. P., & Muralidhar, V. (2001). Analysis and characterization of photo-plethysmographic signal. *IEEE Transactions on Biomedical Engineering,* **48(1)**, 5-11. https://doi.org/10.1109/10.900243

Divya, R., & Vanathi, P. T. (2019). Cardiovascular Disease Classification using Photoplethysmography Signals—Survey. *International Journal of Computer Applications*, **182(43)**, 10-15.

Divya, R., Shadrach, F.D. & Padmaja, S. (2023). Cardiovascular risk detection using Harris Hawks optimization with ensemble learning model on PPG signals. *Signal, Image and Video Processing*, **17(8)**, 4503-12. https://doi.org/10.1007/s11760-023-02684-y

Elgendi, M. (2012). On the analysis of fingertip photoplethysmogram signals. *Current Cardiology Reviews*, **8(1)**, 14-25. https://doi.org/10.2174/157340312801215782

Fister Jr, I., Fister, D. & Fister, I. (2013). A comprehensive review of cuckoo search: variants and hybrids. *International Journal of Mathematical Modelling and Numerical Optimisation*, **4(4)**, 387-409. https://doi.org/10.1504/IJMMNO.2013.059205

Gupta, S., Gudapati, R., Gaurav, K. & Bhise, M. (2013). Emerging risk factors for cardiovascular diseases: Indian context. *Indian Journal of Endocrinology and Metabolism,* **17(5)**, 806-14. https://doi.org/10.4103/2230-8210.117212

Harikumar, R. & Vijayakumar, T. (2015). A real time experimental setup for classification of epilepsy risk levels. *Applied Soft Computing*, **35**, 493-501. https://doi.org/10.1016/j.asoc.2015.05.039

Harikumar, R., Balasubramani, M. & Vijayakumar, T. (2012). Performance analysis of patient specific epilepsy risk level classifications from EEG signals using two tier hybrid (fuzzy, soft decision trees models and MLP neural networks) classifiers. *International Journal of Soft Computing and Engineering*, **2(2)**, 541-549.

Hosseini, Z. S., Zahedi, E., Movahedian Attar, H., Fakhrzadeh, H. & Parsafar M. H. (2015). Discrimination between normal-to-mild and severe coronary artery disease using time-domain features of the finger photoplethysmogram response to reactive Hyperemia. *Biomedical Signal Processing and Control*, **18(2)**, 282-292. https://doi.org/10.1016/j.bspc.2014.12.011

Karlen, W., Raman, S., Ansermino, J. M., & Dumont, G. A. (2013). Multiparameter respiratory rate estimation from the photoplethysmogram. *IEEE Transactions on Biomedical Engineering*, **60(7)**, 1946-1953. https://doi.org/10.1109/TBME.2013.2246160.

Karlen, W., Turner, M., Cooke, E., Dumont, G., & Ansermino, J. M. (2010). CapnoBase: Signal database and tools to collect, share

and annotate respiratory signals. In *2010 Annual meeting of the society for technology in anesthesia*, 27.

Klema, V., & Laub, A. (1980). The singular value decomposition: Its computation and some applications. *IEEE Transactions on Automatic Control*, **25(2)**, 164-176. https://doi.org/10.1109/TAC.1980.1102314

Lee S., & Hayes M. (2004). Properties of the singular valued for efficient data clustering. *IEEE Signal Processing Letters*, **11(11)**, 862-866. https://doi.org/10.1109/LSP.2004.833513

Lin, W.H., Zhang, H., & Zhang, Y.T. (2013). Investigation on cardiovascular risk prediction using physiological parameters. *Computational and Mathematical Methods in Medicine*, **2013(1)**, 1-21. https://doi.org/10.1155/2013/272691

Manimegalai, P., Jacob, D., & Thanushkodi, K. (2012). An early prediction of cardiac risk using augmentation index developed based on a comparative study. *International Journal of Computer Applications*, **49(15)**, 27-32. https://doi.org/10.5120/7705-1068

Montuelle, L., & Le Pennec, E. (2014). Mixture of Gaussian regressions model with logistic weights, a penalized maximum likelihood approach. *Electronic Journal of Statistics*, **8(1)**, 1661–1695. https://doi.org/10.1214/14-EJS939.

Permuter, H., Francos, J., & Jermyn, I. (2006). A study of Gaussian mixture models of color and texture features for image classification and segmentation. *Pattern Recognition*. **39(4)**, 695-706. https://doi.org/10.1016/j.patcog.2005.10.028.

Rahmad, C., Ariyanto, R., & Yunianto, D. R. (2018). Brain signal classification using genetic algorithm for right-left motion pattern. *International Journal of Advanced Computer Science and Applications*, **9(11)**, 247-251.

Rajaguru, H., & Prabhakar, S. K. (2016). An exhaustive analysis of code converters as pre-classifiers and K means, SVD, PCA, EM, MEM, PSO, HPSO and MRE as post classifiers for classification of epilepsy from EEG signals. *Journal of Chemical and Pharmaceutical Sciences*, **9(2)**, 818-822.

Ramachandran, D., Thangapandian, V.P., & Rajaguru, H. (2020). Computerized approach for cardiovascular risk level detection using photoplethysmography signals. *Measurement*, **150**, 107048. https://doi.org/10.1016/j.measurement.2019.107048

Sadasivan, P. K., & Dutt, D. N. (1996). SVD based technique for noise reduction in electroencephalographic signals. *Signal Processing*, **55(2)**, 179-189. https://doi.org/10.1016/S0165-1684(96)00129-6

Safavian, S. R., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man, and Cybernetics*, **21(3)**, 660-674. https://doi.org/10.1109/21.97458

Salembier, P., & Garrido, L. (2000). Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval. *IEEE Transactions on Image Processing*, **9(4)**, 561-576. https://doi.org/10.1109/83.841934

Selvaraj, N., Jaryal, A., Santhosh, J., Deepak, K. K., & Anand, S. (2008). Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography. *Journal of Medical Engineering & Technology*, **32(6)**, 479-484. https://doi.org/10.1080/03091900701781317

Sološenko, A., Petrėnas, A., & Marozas, V. (2015). Photoplethysmography-based method for automatic detection of premature ventricular contractions. *IEEE Transactions on Biomedical Circuits and Systems*, **9(5)**, 662-669. https://doi.orf/10.1109/TBCAS.2015.2477437.

Sowmiya, M., Banu Rekha, B., & Malar, E. (2023). Ensemble classifiers with hybrid feature selection approach for diagnosis of coronary artery disease. *The Scientific Temper*, **14(3)**, 726-734. https://doi.org/10.58414/ SCIENTIFICTEMPER.2023.14.3.24.

Sukor, J. A., Redmond, S. J., & Lovell, N. H. (2011). Signal quality measures for pulse oximetry through waveform morphology analysis. *Physiological Measurement*, **32(3),** 369–384. https://doi.org/10.1088/0967-3334/32/3/008

Telesca, L., Lovallo, M., Ramirez-Rojas, A., & Angulo-Brown, F. (2008). Non-uniform scaling behavior in self-potential earthquake related signals. *Fluctuation and Noise Letters*, **8(3-4)**, 261-267. https: doi.org/10.1142/S0219477508005100.

Umadi, L. F., Azam, S. N. A. M., & Sidek, K. A. (2016). Heart Abnormality Detection Technique using PPG Signal. *Journal of Telecommunication, Electronic and Computer Engineering,* **8(12)**, 73-77.