Musik in bayern
ISSN: 0937-583x Volume 89, Issue 4 (April -2024)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2024-*251*

ENHANCED PRINICIPAL COMPONENT ANALYSIS BASED FEATURE EXTRACTION AND HYBRID DEEP LEARNING TECHNIQUES FOR PREDICTION OF HEART DISEASE DATA

Sugendran G, Research Scholar, Department of Computer Science, Dr.G.R. Damodaran College of Science, Coimbatore, & Assistant Professor, Department of Computer Applications, KSG College of Arts and Science, Coimbatore

Dr.S.Sujatha, Head & Assistant Professor, Department of Computer Science, Dr.G.R. Damodaran College of Science, Coimbatore

To Cite this Article

Sugendran G, Dr.G.R. Damodaran, Dr.S.Sujatha "ENHANCED PRINICIPAL COMPONENT ANALYSIS BASED FEATURE EXTRACTION AND HYBRID DEEP LEARNING TECHNIQUES FOR PREDICTION OF HEART *Musik Bayern, Vol. 89, Issue 4, April 2024, pp1-19*

Article Info

Received: 14-03-2024 Revised: 23-03-2024 Accepted: 01-04-2024 Published: 11-04-2024

Abstract: Cardiovascular and heart diseases constitute one of one of the most prevalent and fatal illnesses impacting people's health. Heart disease has increased as a result of the development of bad habits including smoking, overeating, and inactivity. Although angiography is often acknowledged as the most effective method for identifying coronary artery disease, it is costly and has a number of serious side effects. The ability to identify heart disease early on by recognizing its early signs is one of the most crucial things in the current world. Heart disease's fatal characteristic has given it the moniker "silent killer," as there are seldom any evident warning signs. In order to develop an effective strategy for recognizing the early stages of heart disease using fundamental knowledge and symptoms, further research is thus required. To overcome these challenges, our study effort produced a hybrid deep learning approach and efficient feature extraction method for the categorization of heart condition. The four primary stages of this study are extraction of feature, feature section, preprocessing, and classification. The K-Means Clustering (KMC) technique is used for preprocessing in the beginning with the goal of improving classifier accuracy. Subsequently, the decreased dimensional feature subset is extracted using the feature extraction technique with the pre-processed data. This paper suggests using an enhanced Principal Component Analysis to efficiently extract features.After that, the feature selection process is finished using the EGA method. In order to choose the most pertinent and significant characteristics from the cardiovascular illness dataset, it computes the optimal fitness value. The Enhanced Elman Neural Network (EENN) and a recurrent neural network are used in a hybrid deep learning model to identify heart disease data. According to the study results, machine learning techniques are not as effective in predicting heart disease data as the hybrid deep learning model that is suggested.

Keywords: K mean clustering, prediction of heart disease, modified Principal Component Analysis, EGA algorithm, Enhanced Elman Neural Network (EENN) and recurrent Neural Network (RNN)

1. Introduction

Based on projections form the WHO, heart issues cause more than twelve million fatalities globally each year. Cardiovascular illnesses are the primary cause of death in the US and other industrialized nations [1]. In many developing countries, it also serves as the primary cause of mortality. It is generally believed to be the primary factor in adult mortality. Heart ailments comprise all disorders that impact the heart in one broad category. Heart disease accounted for the majority of deaths in numerous nations, including India. Heart disease claims the

Musik in bayern
ISSN: 0937-583x Volume 89, Issue 4 (April -2024)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2024-*251*

life of one person per thirty-four seconds. Heart conditions include coronary cardiac disease, cardiomyopathy, and cardiovascular illness.

A collection of conditions collectively referred to as "cardiovascular disease" impact the cardiac muscle, blood vessels, and the body's circulatory system. Heart disease is a leading cause of illness, incapacity, and mortality [2]. Making medical diagnoses is a crucial and difficult task. Medical evaluation is seen as an important yet challenging process that must be completed promptly and precisely. Automating this system would substantially help it. However, few physicians are experts in every topic, and resource professionals are few in many specialties. Therefore, bringing them all together would be greatly aided by an automated medical diagnosing system [3]. You can achieve cost-effective clinical testing with the use of the right data collection as well as assisted decision-making technology. For automated solutions to be deployed precisely and efficiently, a comparative analysis of current methods is necessary.

The enormous scope for medical data mining makes it feasible to uncover hidden patterns among health records. Such patterns can be utilized applied to produce a clinically substantiated diagnosis. However, the initial medical information that are now available are many, diversified, and widely distributed [4]. This information must be gathered in a systematic way. The collected data may be integrated to form a database for hospitals. A convenient method for identifying unique and concealed patterns in data is provided by data mining tools. Machine learning is one useful method for training and testing that is used in testing.

Machine learning is a specialized topic of artificial intelligence (AI), a vast study that tries to mimic human skills in computers [5]. These technologies combine to form machine intelligence, which is another term for systems of machine learning that have been trained to understand when to comprehend and utilize information. In order to apply mathematical optimization to real-world corporate, social, industrial, and medical issues, the fields of computational statistics and machine learning work together. One may generally categorize it into two main groups: Learning can be supervised or unsupervised [6]. In supervised learning, an algorithm creates a mathematical model based on a set of inputs and projected outputs. Unsupervised learning, on the other hand, refers to the process of developing a model of mathematics using a technique that will not provide labeling as an output, but rather merely accepts inputs.

Considering the fact that our study attempts to predict the chance of cardiac disease by using physiological mechanisms [7]. The optimal choice is unquestionably supervised learning provided the resources and anticipated outcomes are known in advance.

To forecast the severity of cardiac disease in people, many data mining and neural network approaches were used. A variety of methods are employed to assess the severity of illness, such as the Naive Bayes (NB), Genetic Algorithm (GA), Decision Trees (DT), and K-Nearest Neighbor Algorithm (KNN) [8–9]. Because heart disease is so complicated, care must be taken with caution. If this isn't done, it might affect the heart and cause an early death. To identify various metabolic illness kinds, data mining and medical research perspectives are employed. Classification-based data mining is important for both data exploration and heart disease prediction. Decision trees have also been applied to the accurate prediction of heart disease-related events [10–11Various approaches have been used to apply known techniques for Data Mining to extract knowledge for the Predicting heart disease. Still, accuracy of the classifiers declines with increasing dataset complexity. Thus, this research effort generated a hybrid deep learning algorithm and an effective feature extraction for the categorization of heart condition.

The remainder of the research is organized as follows: Section 2 discusses some of the most modern strategies for recognizing cardiovascular problems in data. Section 3 describes an approach to the proposed strategy. Section 4 presents the findings and comments. Section 5 outlines the results and future research.

2. Literature Review

Musik in bayern
ISSN: 0937-583x Volume 89, Issue 4 (April -2024)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2024-*251*

The function and effectiveness of various are discussed in this section nature-inspired and feature extraction strategies used to diagnose the provided heart disease data.

Acharya et al [12] an eight-tier convolutional neural network (CNN) consisting of four optimum combination levels, three fully relevant coatings, and four convolutional tiers is presented using the two- and five-second ECG dossier sectors, separately, with the aim of detecting CAD. Using Deep CNN, 94.95% precision, 93.72% sensation, and 95.18% specificity are reached for Net 1 in two seconds; 95.11% accuracy, 91.13% sensation, and 95.88% precision are reached for Net 2 in five seconds. The proposed method makes use of an ECG dossier to help physicians get more accurate and trustworthy CAD judgments.

Liang et al [13] a novel method that combines A convolutional neural network (CNN) is combined with bidirectional long short-term memory (BiLSTM) to improve accuracy and save training time. To examine a practicable and viable heartbeat event classification approach, Database 1 a single-lead ECG and Database II with a twelve-lead ECG were used. Two approaches were identified and assessed for processing pulse occurrence classification: Method I, which employs a metamorphic effecting the whole animate nervous system, and Method II, which joins the CNN with the BiLSTM network using a deep knowledge approach. Overall, Method I performed a bit higher than Method II. In contrast to Method I, which required a mean of 28.3 hours, Method II only required one step to train the model. Method II permitted 82.6, 85%, and 80 truthfulness principles separately. Comparing these findings to the representation of modern methods used for the static job, they are better.

Vafaie et al [14] developed an appropriate algorithm capable of classifying ECG signals with unknown properties based on their similarities to ECG signals with known features must be created in order to automate the identification of cardiac disease. With the aim of improving the accuracy of ECG data identification, this study presents a novel classification technique relying upon a dynamical model of that signal. Empirical results show that this suggested approach incorporates a fuzzy classifier with a 93.34% accuracy rate in differentiating ECGs. To further enhance the classifier's performance and raise the prediction accuracy to 98.67%, a genetic algorithm is utilized. In order to more accurately diagnose arrhythmias, our suggested technique improves the precision of ECG categorization.

Chen et al [15] provide a unique two-step prophetic method for preparing ECG signals, whereby the data is compared using a global citation model by a worldwide classifier, which labels severe abnormalities (flame alarms). An accurate change research is made simpler by enhancing the percentage of signals for various abnormal class in the characteristic room through a managed nonlinear metamorphosis with optimal limitations. With a 96.6% classification accuracy, the suggested approach offers a special predictive analytic function by alerting users to the increased risk of cardiac problems and advising them to follow doctor's prescriptions for preventative measures.

Alarsan et al [16] suggested a quick and accurate method for classifying ECG data. The many specialized cardiac heart tissues create a mix of action impulse waveforms throughout each beating. Numerous classification approaches (Decision Trees, Random Forests, Gradient-Boosted Trees (GDB), and so on) admit that Spark-Scala machine intelligence can easily do this. The suggested method is evaluated and verified using the MIT-BIH Supraventricular Arrhythmia and baseline MIT-BIH Arrhythmia databases. According to the findings, our planning had an overall validity of 96.75% for twofold categorization when utilizing the GDB Tree technique and 97.98% while using haphazard Forest. Random Forest scored 98.03% validity for multiclass classification; slope pushing forests can only categorize in two ways.

Jahmunah et al [17] An automated system (AS) was created using convolutional neural networks (CNN) and unique GaborCNN models to identify ECG data as normal, CAD, myocardial infarction (MI), and congestive heart failure (CHF). Weight balancing was employed to fix the imbalances in the dataset. The CNN and Gabor CNN models achieve classification accuracy of exceeding 98.5 percent for regular, cardiovascular illness, myocardial infarction (MI), and acute heart failure, respectively.GaborCNN is the suggested model since it outperforms CNN while using less processing resources.

Musik in bayern
ISSN: 0937-583x Volume 89, Issue 4 (April -2024)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2024-*251*

Sharma et al [18] developed a computerized system using certain GaborCNN models and convolutional neural networks (CNN) to automatically classify ECG data into classifications for congestive heart failure (CHF), myocardial infarction (MI), CAD, and normal. To balance the unbalanced dataset, weight balancing was employed. The CNN and GaborCNN models performed well in the four-class categorization of regular, cardiovascular disease, myocardial infarction, and acute cardiac failure, with classification accuracies greater than 98.5%. More people use the GaborCNN model because it performs better and requires fewer processing resources than the CNN method.

Neagoe et al [19] To diagnose ischemic heart disease (IHD), a Fuzzy-Gaussian Neural Network (FGNN) for ECG signal recognition was created. The suggested ECG processing cascade comprises the following two primary steps: The QRST zone of ECG data may be used to extract features the FGNN, or Principal Component Analysis (PCA) or Discrete Cosine Transform (DCT)may be used to classify patterns and identify IHD. The proposed neuro-fuzzy model for diagnosing IHD has been tested, and the software has been developed. A 100% IHD recognition score produced the best results.

Dalal et al [20] presented an innovative, reliable method for identifying cardiac problems. The MLII, PTBDB, and UCI repository arrhythmia databases were used for the ECG experiments. Because each of these databases is unbalanced, resampling approaches assist to restore the databases' equilibrium. Multi-cumulants are used to acquire features, whereas discrete wavelet transforms (DWT) are used to eliminate noise. The basic foundation of this technique is provided by the characteristics that are retrieved from the ECG data about multi-cumulants. Multiple-cumulants feature-based methodologies Kernel extreme learning machine (KELM) is used to classify ECG data. Application of genetic algorithms (GAs) improves the KELM and multi-cumulants parameters. 99.57% accuracy in case classification is obtained on the PTBDB database, compared to 100% accuracy on the MLII and UCI repository arrhythmia datasets. A comparison of the proposed strategy with the latest methods now in use has also shown its efficacy.

Darmawahyuni et al [21] Recurrent neural networks (RNNs) and long short-term memory (LSTM) topologies are two deep learning approaches that have been presented for developing an automatic extraction of features and highly accurate diagnostic procedure with minimal input. The electrocardiogram (ECG) the study used data that came from the PhysioNet databases, which are accessible to the general public. To get the optimal model, we changed 24 LSTM models' hyperparameters. The portion of the ECG signal was one area that was looked at in the first fifteen minutes of the investigation. Of the 24_LSTM models, Model 1 (which was constructed using the first fifteen minutes of ECG data) yielded the highest results in terms of accuracy (99.86%), sensitivity (99.85%), specificity (99.85%), precision (99.87%), and F1-score (99.86%).

Shankar et al [22] ECG characteristics were classified using a range of classifiers, including K Nearest Neighbor (KNN), Fish Swarm Optimization (FSO), Whale Optimization Algorithm (WOA), Gray Wolf Optimization, and Particle Swarm Optimized (PSO) approaches. The average accuracy scores for the Adam and R-Adam classifiers were 72.32 and 85.63, respectively. Furthermore, classifier performance is improved by hyperparameter selection using the Adam technique and Random Adams (R-Adam) techniques.

Proposed Methodology

This work developed a deep learning system with a hybrid approach and an efficient feature extraction strategy for heart disease detection. This activity is divided into four fundamental phases: preliminary processing, categorization, extraction of features, and feature section. The ideal fitness value is calculated using the dataset on cardiovascular disease to find more relevant and crucial qualities. The Enhanced Elman Neural Network (EENN) model is used to categorize data relevant to cardiovascular illness. This paper suggests a useful strategy for extracting features that includes changing the Principal Component Analysis.

Musik in bayern
ISSN: 0937-583x Volume 89, Issue 4 (April -2024)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2024-251

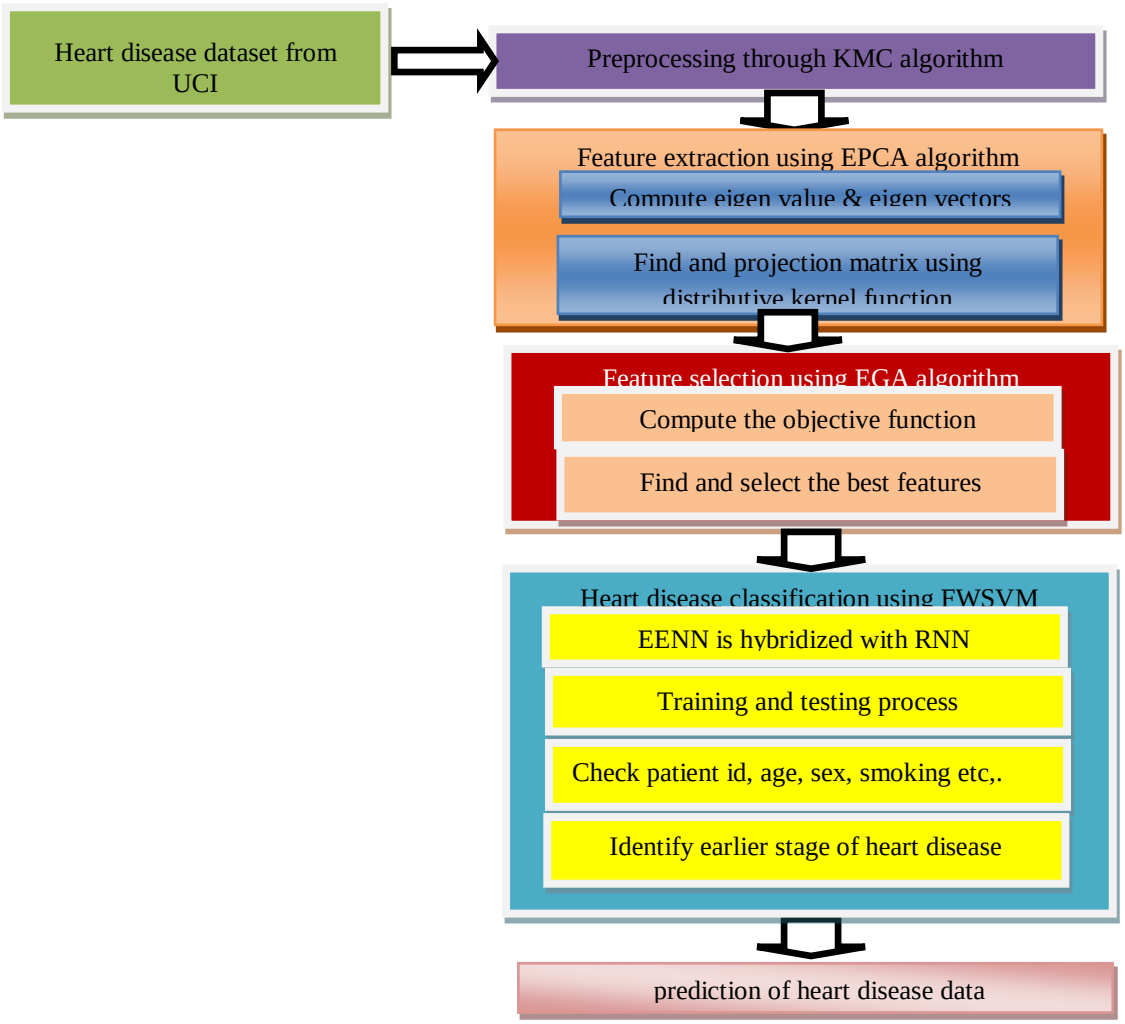


Figure 1. The methodology of the suggested heart disease prognosis model

3.1. Dataset description

Cleveland cardiovascular disease data were used to evaluate the suggested model. The dataset is believed to identify patients with cardiac ailment by allocating a number, ranging from 0 (lacking) or 1 (present). The online data extraction and machine learning (ML) library of the University of California, Irvine (UCI) provided the data used in the model. The patient's health status was determined by looking at merely the fourteenth and sixteenth features. The Cleveland dataset, which consisted of 303 examples with 76 variables, served as the basis for the model.

3.2. K-Means Clustering (KMC) Algorithm preprocessing

In the past, information preparation has been an essential initial step in the information mining process. Data must be formatted to facilitate quicker and more effective processing in order to be used in data exploration, machine learning, and other data science-related fields. Usually, data prep is used early in the pipeline for developing AI and machine learning. The methods are usually used early on in the creation of machine learning, AI, and software as well as data mining to ensure accurate results.

a) K-Means Clustering Algorithm

The optimal centroid is found via iteratively computing centroid values using the K-means clustering algorithm. Another term for it is the flat clustering approach. The symbol K stands for the greatest number of groups that the

Musik in bayern
ISSN: 0937-583x Volume 89, Issue 4 (April -2024)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2024-*251*

technique could identify from the data. This method groups data points in groups to ensure the total squared distance between the center and the information points is as near to zero as is practical. Less variation within clusters leads to more comparable data points within of them.

The processes listed below will aid in our comprehension of the operation of the K-Means clustering technique:

- The quantity of clusters must be provided in the first phase(K) that this computation should produce.
- Then next step, arbitrarily select K data points of interest and distribute each one to a group. Request information in the simplest words possible, based on the amount of data points of interest.
- In the third step, the centroids of the group are registered.
- Step 4: Repeat previous steps till the optimum centroid — the groupings where the data focuses are offered that remain constant — is identified.

  • 4.1 The primary number to be calculated represents the actual squared distance between both centroids and information centers.
  • 4.2 Next, assign each information highlight to the group (centroid) that is closest to the distinct groupings.
  • 4.3 Finally, calculate the group's centroids by averaging each element of data from all the groups.

The issue is addressed using K-implies and the Assumption Expansion approach. The Boost step is used to determine a group's centroid once data focus are assigned for the next group using the Assumption step. The E-step is preferred for information emphasis because to its proximity. The M-step approach involves computing the centroid of each bunch. If you want to skip it, here's an overview of the numerical layout.

The aim of the function is

$$J = \sum_{i=1}^{m} \sum_{k=1}^{K} w_{ik} \left\| x^i - \mu_k \right\|^2 \qquad (1)$$

where, if data point xi is a member of cluster k, $w_{ik} = 1$; if not, $w_{ik} = 0$. Furthermore, x^i's cluster centroid is μ_k.

3.3. Feature extraction using Enhanced Principal Component analysis (EPCA)

Consider an ensemble of s nodes V = {v_i , $1 \le i \le s$}, that may communicate with an overseer v0 in the distributed framework. Every node vi has a regional data matrices $P_i \in R^{ni \times d}$ with ni values in d dimension (n_i> d). The global data $P \in R^{n \times d}$ is made up of local data matrix combinations $n = \sum_{i=1}^{s} n_i$ and $P^T = [P_1^T, P_2^T, \dots \dots P_s^T]$. Pi indicates the i-th row of P. The data points are considered as being concentrated with zero mean, i.e.,$\sum_{i=1}^{s} p_i = 0$. Algorithms for uncentered information should be adjusted rank-one [24], via the costs of calculation and transmission carried mostly by those participating in the intermediate stages.

- **Constructing the distributive Kernel Matrix**

Assume the space representing the first-class D went through a nonlinear transformation $\phi(x)$,transforming into thespace comprising class M, wherein M≫D. The point $\phi(x_i)$ will appear for each of the reference points x_i. Conventional PCA is doable in the emerging element space, but it is often prohibitively costly and unworkable. Fortunately, the computation may be simplified with coin collation.

Initially, let's suppose that the mean of the anticipated new features is 0.

$$\frac{1}{N} \sum_{i=1}^{N} \phi(x_i) = 0. \qquad (2)$$

The covariance matrix for projected features is M × M, as defined by

$$C = \frac{1}{N} \sum_{i=1}^{N} \phi(x_i) \phi(x_i)^T. \qquad (3)$$

Provided by are its eigenvalues and eigenvectors.

Musik in bayern
ISSN: 0937-583x Volume 89, Issue 4 (April -2024)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2024-*251*

$$Cv_k = \lambda_k v_{k.} \quad (4)$$

In this case, k = 1, 2, · · ·, M. Drawing from Equations (3) and (4),

$$\frac{1}{N}\sum_{i=1}^{N} \phi(x_i)\{\phi(x_i)^T v_k\} = \lambda_k v_k, (5)$$

which can be rewritten as

$$v_k = a_{ki}\phi(x_i). (6)$$

Now, have changed vk in Eq. (5) to Eq. (6).

$$\frac{1}{N}\sum_{i=1}^{N} \phi(x_i)\phi(x_i)^T \sum_{j=1}^{N} a_{kj}\phi(x_j) = \lambda_k \sum_{i=1}^{N} a_{ki}\phi(x_i). \quad (7)$$

As defined by the kernel function,

$$k(x_i, x_j) = \phi(x_i)^T \phi(x_j), \quad (8)$$

and multiply $(x_l)^T$ on either side of Equation. (8), have

$$\frac{1}{N}\sum_{i=1}^{N} k(X_l, X_i) \sum_{j=1}^{N} a_{kj} k(X_i, X_j) = \lambda_k \sum_{i=1}^{N} a_{ki} k(X_l, X_i) \quad (9)$$

Use matrix notation as in this

$$K^2 a_k = \lambda_k N K a_k \quad (10)$$

where

$$K_{i,j} = k(X_i, X_j) \quad (11)$$

Moreover, a_k is a_{ki} N-dimensional column vector:

$$a_k = [a_{k1} a_{k2} \dots a_{kN}]^T \quad (12)$$

The solution to a_k is

$$Ka_k = \lambda_k N a_{k,} \quad (13)$$

and the kernel main components that are obtained may be computed using

$$y_k(X) = \phi(X)^T v_k = \sum_{i=1}^{N} a_{ki} k(X, X_i). \quad (14)$$

A benefit of using kernel techniques is that they do not require the direct calculation of $\phi(x_i)$. Utilizing the training dataset $\{x_i\}$, compute the kernel matrix directly. Two popular kinds of kernels are polynomial kernels and

$$k(x, y) = (x^T y)^d \quad (15)$$

or

$$k(x, y) = (x^T y + c)^d, \quad (16)$$

Musik in bayern
ISSN: 0937-583x Volume 89, Issue 4 (April -2024)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2024-*251*

where the Gaussian kernel and a constant c>0 are present.

$$k(x, y) = \exp(-\|x - y\|^2/2\sigma^2) \quad (17)$$

using the σ parameter. PCA approximation with lℓ_2- error fitting. Let A = [a_{ij}] be a matrix, and let $\sigma_i(A)$ be its ith singular value. Let$\|A\|_F^2 = \sum_{i,j} a_{ij}^2$. The matrix containing A's first t columns is indicated by the notation $A^{(t)}$. For each column of X, let L_X be the linear subspace that they span.Let $\pi_L(p)$ be the extension of an object p on subspace L, and $\pi_X(p)$ be the reduction of $\pi_L(p)$. Determine the square of the length that separates a point $p \in R^d$from an imaginary space $L \subseteq R^d$.

$$d^2(p, L) \coloneqq \min_{q \epsilon L} \|p - q\|_2^2 = \|p - \pi_L(p)\|_2^2. \quad (18)$$

The rSub_space k-Cluster on $P \in R^{n \times d}$, which can be linear or affine, is defined as

$$c_{\mathcal{L}}^{\min d^2(P,\mathcal{L}) \coloneqq \sum_{i=1}^{n} \min_{L \epsilon \mathcal{L}} d^2(p_i, L)} \quad (19)$$

Given a n × d matrix P with rows $p_1, \ldots, p_n$, and $\mathcal{L} = \{L_j\}_{j=1}^k$representing a collection of k centers, let's say that P is composed of affinal or direct subspaces in r dimensions.

PCA is a special instance of the rdimensional subspace with k = 1. The top r precisely single components of P, known as major components, which can be located via Singular Value Decomposition, span this ideal r-dimensional domain.In addition, k-means clustering with points as centers (r = 0) is a particular instance of the aforementioned. In particular, focus on relative-error approximation methods, where the goal is to produce a set $L'$of k centers such that $d^2(P, L') \leq (1 + \epsilon) \min_{\mathcal{L}} d^2(P, \mathcal{L})$.

Using the following protocol, each server calculates the top $O(r/\epsilon)$key elements Yi of Pi and sends those to the controller for approximated distributed PCA. A coordinator determines the highest r primary components of Y and transmits it to the servers after stacking the $O(r/\epsilon) \times d$matrices Yi on top of one another to create an $O(sr/\epsilon) \times d$matrix Y. This gives the PCA problem a relative-error approximation.

Improved Communication: To reduce the transmission cost while employing distributed PCA for ℓ_2-fitting issues akin to K-means clustering. Assuming a data matrix P, a $(1+)\epsilon$ -approximation may be obtained by projecting the k-means issue is solved in the subspace covered by the top $O(k/\epsilon^2)$ major components. To compute the $O(k/\epsilon^2)$ global principal components in a distributed context, this would need executing Algorithm IDK-PCA with parameter $r = O(k/\epsilon^2)$, first and requiring at least $O\left(\frac{skd}{\epsilon^3}\right)$of communication. In this subspace, a distributed k-means issue may be solved, and an α-estimatewithin equates to an aggregate α(1 + ϵ) estimate.

Improved Computation:Its point set Pi's Singular Value Decomposition (SVD), which requires $min(n_i d^2, n_i^2 d)$time, is a substantial bottleneck. Then modify Algorithm IDK-PCA such that each server runs the method on the point set provided by the rows of H_iP_i, rather than first sampling an Oblivious Subspace Embedding (OSE) matrix H_i. H_i can contain just one non-zero item per column by selecting known OSEs; as a result, H_iP_i can be calculated in $nnz(P_i)$ time. Also, H_i has $O(d^2/\epsilon^2)$ rows, which can be a lot fewer than the initial amount of rows. If you are prepared to invest $O(dlog^{o(1)}d/\epsilon^2)$time, you may further decrease this number of rows to $O(nnz(P_i)log^{O(1)}d/\epsilon)$ .Keep in mind that H_iP_i non-zero entries do not exceed P_i.

3.4. Feature Selection using Enhanced Genetic Algorithm (EGA)

Musik in bayern
ISSN: 0937-583x Volume 89, Issue 4 (April -2024)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2024-*251*

The process of eliminating duplicate, unnecessary, or noisy characteristics from the initial features set in order to choose the subset of those that are most relevant is known as feature selection. Other features in the dataset are redundant or irrelevant, and just a few variables are needed in order to construct a learning model.If the dataset includes all of this extraneous and worthless information, the model's overall efficiency and precision may suffer. As a result, it is critical to identify and choose the most important qualities from this data in order to eliminate any redundant or inconsequential information. This is performed through the selection of machine learning features.

3.4.1. Basics of Genetic Algorithm (GA)

John Holland initially presented the Genetic Algorithm (GA) [25] in the early 1970s. It is a potent stochastic algorithm that applies the laws of evolution and biological genetics to algorithms for learning and optimization issues. A genetic algorithm (GA) solves a problem by keeping track of a population of people, also known as strings or chromosomes, and using genetic operations like selection, crossovers, and mutations change the population in a probabilistic manner.

a) Initial population

A collection of chromosomes that are created at random makes up the starting population. Each chromosome is represented by one solution in the search space, representing a set of potential features subsections. The *n* binary vectors that make up each chromosome's characteristics are encoded with a value of "1" denoting that the relevant feature is picked and "0" denoting that it is not.

b) Fitness function

One important consideration for choosing which characteristic subsets to include in the subsequent generation of the new population is the function of fitness, which evaluates the efficacy of the subsets. Higher performing subsets will have a higher chance of being chosen and proliferating to establish a new population. In this study, the amount of data of the subset is utilized to determine the fitness-related feature and evaluate the subset appropriately. The fitness-related feature shown below will be applied in our method:

$$Fit - Fun(S_i) = Z * C(S_i) + (1 - Z)(1/size(S_i)) \qquad (20)$$

where Z is a parameter between 0 and 1, size (S_i) is the number of attributes in the subset, and $C(S_i)$ is the macro-average F-measure of the selected feature subset. A higher Z value suggests that classification ability is more important than feature subset size. S_i's feature subset fitness increases with categorizing efficiency and decreases with S_i size. Z's value in the current method was altered to 80%.

c) Selection

After the subsets have been evaluated, the selection process is carried out, and the subsets are chosen based on the chosen selection technique and their relative fitness. The GA is most commonly used with roulette wheel selection (RWS). Using a probabilistic being chosen value that is proportionate to the subset's fitness and inversely related to the health of the remaining sets in the sample under study, this approach selects the subset to reproduce. Here's how to calculate the probability that a subset S_i will be selected for replication:

$$P(S_i) = \frac{Fit-Fun(S_i)}{\sum_{i=1}^{n} Fit-Fun(S_i)} \qquad (21)$$

d) Crossover

Crossover and transmutation are examples of reproduction mechanisms that generate a new set of subsets (children) form chosen subsets (parents). In order to preserve population variety, these surgeries are crucial. By switching the characteristics of the two parent subsets, two new subsets are created using the most used crossover technique. The likelihood that both parents will switch their characteristics is determined by an interaction rate or probability, which is used to conduct the crossover. Use the one- point crossover technique, in which all features are exchanged once an arbitrary position *i* along the parent subsets' length is chosen. As a

Musik in bayern
ISSN: 0937-583x Volume 89, Issue 4 (April -2024)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2024-*251*

consequence, the first i attributes are chosen from one group, while the remaining characteristics are drawn from the next set.

The suggested enhancement to this crossover process divides each picked parent into two equal halves. The value of the weight all components is derived by adding the weighted measurements for the features inside that part, which are obtained using the characteristic frequency and record frequency approaches. The feature weights were previously calculated during the text preparation step. The traits in each category on each of the chromosomes (subset) are then ordered in order of weight. We get the weight of each characteristic and add them together to get the cumulative weight. The finest two parts from each parent are merged to generate an entirely novel subset (new child), whereas the remaining two parts constitute the second subset. This approach uses feature weighted data to direct the EGA search for the best subsets.

e) Mutation

Mutation is the process of altering a specified subset of traits. The mutation rate determines how many subgroups need to be modified. A random number of characteristics from the specified subset are chosen to have their values changed. Using this strategy, the best subsets from the previous generation are combined with a specified amount of new features to add to the mutant subset. Furthermore, this procedure involves calculating the cumulative correctness (F-measure) of both initial parents and examining the subset's origin of features (the child) that is to be modified. A weighted subset of characteristics is selected if the total accuracy is less than a predefined threshold. On top of these selected features are the most significant characteristics that are not included in the subset of the best-found characteristic subset. This process will provide fresh resources for future generations.

Through iterative application of stochastic operators like crossover, mutation, and selection, the genetic algorithm generates a set of randomly selected potential solutions to the problem at hand.The evaluation of the objective function, also known as fitness, is the foundation of this approach.The greater fitness value corresponds to the superior solution. To find the fitness value, just the function with the objective and the limitations must be evaluated; gradient information is not needed. Because of this derivative freeness method, the GA is more adaptable and can handle issues involving complex objective functions (nondifferentiable functions), where obtaining the derivative is challenging or impossible. The gradient-based optimization techniques' tendency to become stuck GA's stochastic and unpredictable character helps overcome the difficulty of becoming stuck in local optimal conditions.

3.4.2.Genetic algorithm using a Gaussian weight function basis

A unique method to gaussian mass function for crossover, deviant, and penalty function controllers is offered to improve GA behavior and relieve the user of the burden of judging delicate boundaries. This strategy, which employs several simple genetic algorithms (SGA), increases the likelihood of capturing the global best. In addition to a new origin population method, this design makes use of Gaussian weight functional methods for the crossover, deviant, and punishment function controllers. Gaussian Weight Function for the Consequence Function method is somewhat altered as shown in Eqs. (22), (23), (24) and (25).

$$\Phi(X) = F(X)(1 + penalty) \quad (22)$$

$$penalty = g_{ave}\left(\frac{(g_{max}+g(i))}{(g_{max}-g_{ave})}\right) g(i) \geq g_{ave} \quad (23)$$

$$penalty = g_{ave}\left(\frac{(g_{ave}+g(i))}{(g_{ave}-g_{min})}\right) g(i) < g_{ave} \quad (24)$$

$$penalty = 0 \qquad g(i) = 0 \quad i = 1, \ldots\ldots, m \quad (25)$$

Musik in bayern
ISSN: 0937-583x Volume 89, Issue 4 (April -2024)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2024-*251*

The design variables in Eqs. (22–25) are shown for a single heading *X*, the total amount of limitations is n, the primary objective for shortest capacity is *F(X)*, and g_{max}, g_{min}, and present, respectively, represent the mean, least, and highest defilement principles of the era. Changes are made to the objective function *U(X)*. Total breach advantage, *g(i)*, is shown together with the normalized dislocation g_{dj} (*x*), and stressed, g_{sr} (*x*), constraints. g_{dj} (*x*) and g_{sr} (*x*) are assumed to be

$$g_{dj} = d_j/d_{uj} - 1 \qquad j = 1 \dots \dots . m \qquad (26)$$

$$g_{sr} = g_r/g_{ar} - 1 \qquad r = 1 \dots \dots . m \qquad (27)$$

where d_{uj}represents the maximum deviation inside the jth node and gar is the maximum amount of stress allowed in the *rth* element. The stress limitations and the number of displacements is denoted by the letters *m* and *l*, respectively.Violating normalized constraints serves as the foundation for formulating the unconstrained optimization issue. There was not a formulation like this in the earlier edition. Because the probability of survival for creates with minor mistakes and smaller objective values is retained, including provided as an additional parameter makes sense. Designs with significant violations and higher objective values are compelled to be eliminated from generation by the algorithm. The mutation gaussian weight functions and crossover operator for the GA are shown below:

$$p_m = 0.5(f_{max} - f)/(f_{max} - f_{ave})\ f \geq f_{ave} \qquad (28)$$

$$p_m = (f_{ave} - f)/(f_{ave} - f_{min})\ f < f_{ave} \qquad (29)$$

$$p_c = (f_{max} - f')/(f_{max} - f_{ave})\ f' \geq f_{ave} \qquad (30)$$

$$p_c = 1.0 f' < f_{ave} \qquad (31)$$

This graph depicts the state's average relevance profit f_{ave} , the public's maximum and shortest relevance principles f_{max} and f_{min} , and the individual's appropriateness worth (*f*). The appropriateness advantage projected to be hampered is *f0*, which represents rude value. The criterion of constraint breach determines how an individual's variables of design are grouped according to the Gaussian burden function. The most usually abused design changing is so that the person specifically refreshed determined the mutation rate. Equations (28-31) manifold the series separation of every component during production, yielding the forecast of transformation and crossover. Adjusting crossover reduces the estimated number of crossover points from one to the resolution's series time. Different probabilities reflect the quantity of variables in the design that will alter for each individual based on their level of fitness. Consequently, over the course of development, the Gaussian weight function approach can self-correct. Consequently, the method does not require pre-defined parameters. Additionally, the crossover operator is capable of performing its gaussian weight function features, which lead to the flexible point crossover.

3.5. Classification using hybrid deep learning

To detect heart illness data, a network of Neural networks with recurrent properties and an altered improved Elman neural network are employed. High-dimensional huge sample data sets might need a lot of computing power, a considerable of space for storage, and may contain certain related or repeated elements across samples as well as between feature variables. Too many inputs affect productivity and accuracy in identifying when utilizing the Elman neural network to handle them; too many training samples being used simultaneously, together with the inability to create a suitable neural network model, limit recognition accuracy. Recurrent neural networks, which can lessen correlated and repeated information, can be added into the Elman neural network technique to allay these worries.

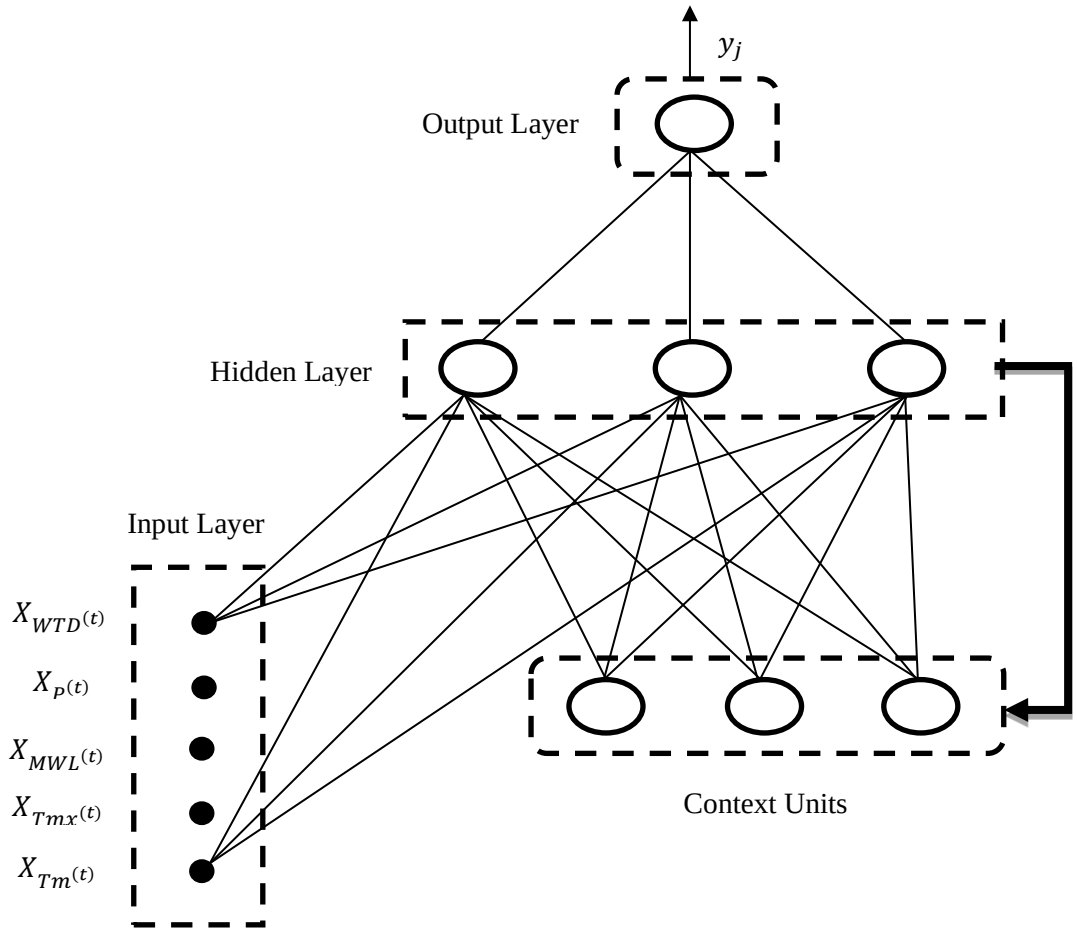
3.5.1. Enhanced Elman neural network

Musik in bayern
ISSN: 0937-583x Volume 89, Issue 4 (April -2024)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2024-*251*

The Elman neural network's basic architecture is shown in Figure 6. The immediately identifiable data input layer, hidden layer, contextual layer, and final layer comprise the four primary layers of the Elman network. Moveable weights link each of the two neighboring layers [26].It is generally regarded as a unique form of feed-forward neural network that incorporates local feedback and extra memory neurons. The Elman network's self-connections among context nodes enable it to be sensitive to input data history, which is particularly helpful for simulating dynamic systems. The following is the notation that was used in this section:

- A force vector called $w1_{ij}$ links node i in the suggested tier with bud j in the concealed layer.
- The relationship between node i in the observed coating and node j in the harvesting tier is represented by the weight$w2_{ij}$.
- $w3_{ij}$ is the pressure that brings together node j to conditions bud i in the covert coating.
- For each of the three responses—m, n, and r—the number of growths in the suggestion, productivity, and secret coatings is assigned.
- The inputs and outputs of Elman's animate nerve organ networks associated with i = 1, 2.....m and j = 1, 2,.....n are denoted as $u_i(k), y_i(k)$.
- Place i=l, 2, ...r corresponds to the secret node i's productivity, $x_i(k)$.
- The circumstances surrounding the node i's production, or the secret node i in the previous instance, is represented by$c_i(k)$.
- Delay is comprised of z^{-1}.

A unit known as an environment unit is created for each component in the hidden layer. Every concealed unit is completely and forwardly related to the context unit. This indicates that each context unit has a weight that is assigned to each concealed unit. Moreover, there exist repeating links between the context units and the hidden units. However, every hidden unit is merely linked to the context unit with which it is associated (Fig. 2).



Musik in bayern
ISSN: 0937-583x Volume 89, Issue 4 (April -2024)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2024-*251*

Fig. 2. Elman neural network model's structure.

The forward weights of recurrent connections are trained using back-propagation, with fixed connection weights.Similar to how the input units' function duringthe context units, as well as the forward phase. The computation of the result and hidden value units follows the same procedure as forward-looking networks. The present values are transmitted through the ongoing links once the outputs of the hidden units have been calculated—with a unit delay—to the associated context units. These are prerequisites for initializing data which will be utilized in the subsequent time step. Target values are used for the outputs in the reverse training phase, and back-propagation is used to adjust the forward weights. $u(k) \in R^m y(k) \in R^n$, $x(k) \in R^r$, are the network's inputs. The outputs of each layer may then be obtained by

$$x_j(k) = f(\sum_{i=1}^{m} w2_{i,j} u_i(k) + \sum_{i=1}^{r} w1_{i,j} c_i(k)) \quad (32)$$

$$c_i(k) = x_i(k-1) \quad (33)$$

$$y_j(k) = g(\sum_{i=1}^{r} w3_{i,j} x_i(k) \quad (34)$$

where f(.) and g(.) stand for the nonlinear or linear output functions, respectively, for the layer that is hidden and outcome layer. Elman networks don't require the usage of states both input and training signals as their dynamic properties are only supplied by internal connections. This is how the Elman network differs from the static feed-forward network.

3.5.2. Recurrent Neural Network (RNN)

An artificial neuron containing one or more feedback loops makes up an RNN, a class of supervised machine learning models. Figure 8 illustrates The loops that repeat over time or in a sequence are known as feedback loops. An input-target pair training dataset is necessary for supervised RNN training [27]. By optimizing the network's weights, the goal is to minimize the variation (i.e., the loss value) among the outcome and goal pairings.

A. Model Architecture

A natural RNN has three coatings, as shown in Figure 3: suggestion, unseen repeating, and quantity. The coating has N recommended holes. A set of headers spanning {..., xt−1, xt, xt+1...}, where xt = (x1, x2,..., xN). These are the parameters for this tier. The weight source W_{IH} identifies the links between recommendation and unseen portions of a fully associated RNN's hidden coating. As shown in Fig. 1b, the layer that is hidden consists of M units that are concealed, ht = (h1, h2..., hM), which are linked together over time via recurrent connections. Hidden units can be initialized with minor non-zero components to improve the network's overall performance and stability. The buried layer defines the system's "memory," or state space, as

$$h_t = f_H(o_t) \quad (35)$$

$$o_t = W_{IH} x_t + W_{HH} h_{t-1} + b_h \quad (36)$$

bh is the hidden units' bias vector, while fH(·) represents the hidden layer's activation function. The weighted link between the output layer and the hidden units is denoted as W_{HO}. The output layer calculates the p units, yt = y1..., yP, as

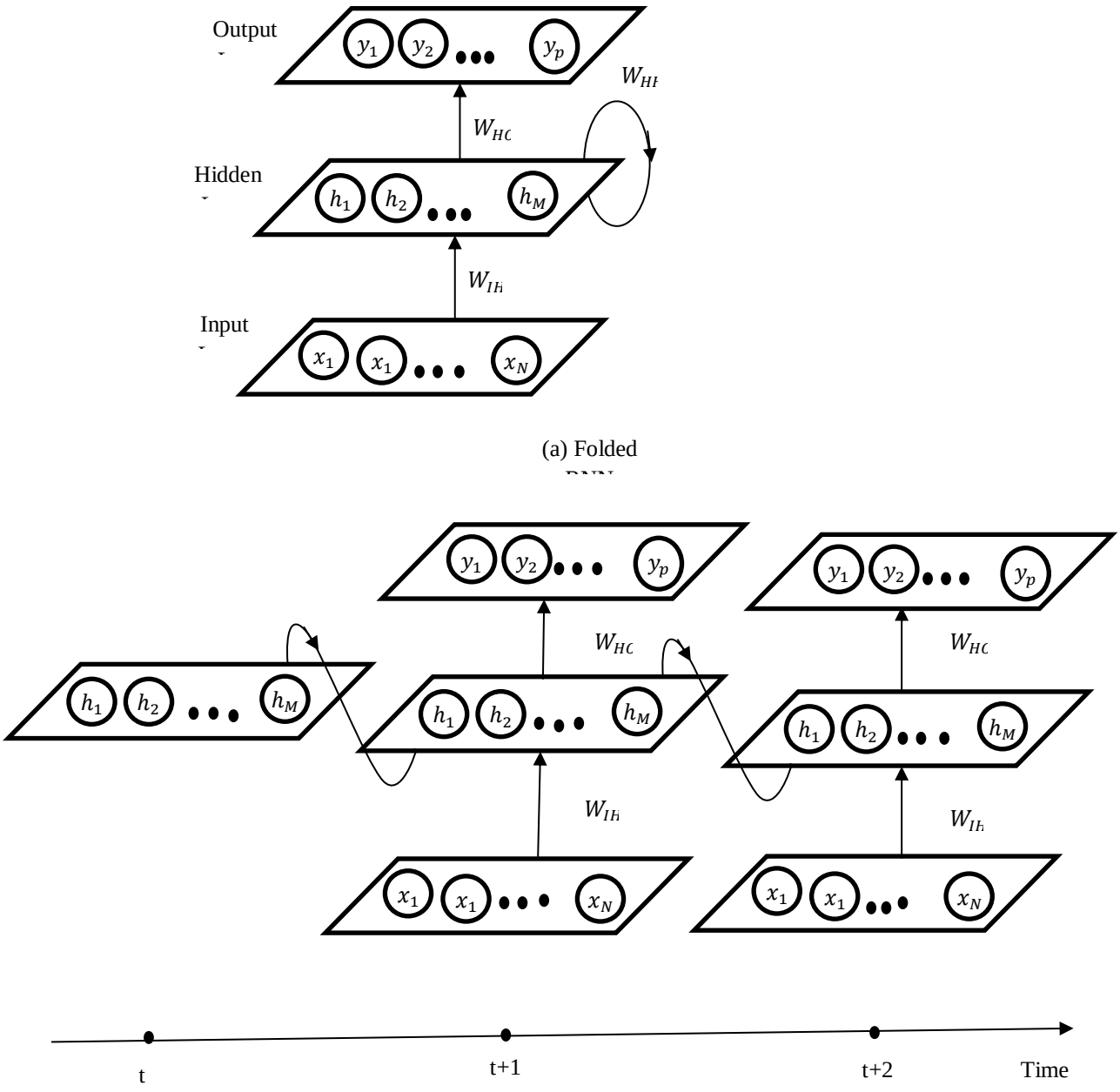
$$y_t = f_0(W_{HO} h_t + b_o) \quad (37)$$

where the output layer's bias vector is *bo* and the activation functions are represented by fO(·). The aforementioned stages are then repeated throughout time t = (1,..., T) Because the target and input historical pairs occur sequentially. Equations (1) and (3) exhibit the nonlinear state equations that make up an RNN over time. At each timestep, the buried states create an output layer prediction based on the input data. A collection of values, unaffected by external effects, contains every unique piece of data required to reconstruct the network's prior states over a range of timesteps. A collection of values reflects an RNN's hidden state. Accurate predictions may

Musik in bayern
ISSN: 0937-583x Volume 89, Issue 4 (April -2024)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2024-*251*

be made at the output layer and the future behavior of the network defined by this combined knowledge. Every unit of an RNN employs a basic nonlinear activation function. However, if properly taught using timesteps, even a basic structure may simulate rich dynamics.

.



(a) Folded

Figure 3. The RNN's Structure

B. Activation Function

In the case of linear networks, a number of linear hidden layers function as one. The ability to define nonlinear boundaries makes nonlinear functions more potent than linear functions. An RNN's hidden levels can include discontinuities, which allow for the identification of input-target links. A selection of the most often utilized stimulation functions are displayed in Find 3. Though not as well-known as the remaining functions of activation, the "sigmoid," "tanh," and corrected linear unit (ReLU) have garnered attention lately. The "sigmoid calculation" is a widely used approach that entails entering a real integer into the range [0, 1]. Utilizing the Cross

Musik in bayern
ISSN: 0937-583x Volume 89, Issue 4 (April -2024)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2024-*251*

Entropy functions for loss in the output layer, the stimulation function is widely used to train classification models. The following is the definition of the words "sigmoid" and "tanh" activation functions:

$$\tanh(x) = \frac{e^{2x}-1}{e^{2x}+1} \qquad (38)$$

$$\sigma(x) = \frac{1}{1+e^{-x}} \qquad (39)$$

in that order. The "sigmoid" function of activation and the "tanh" stimulation function are, as it turns out, scaled.

$$\sigma(x) = \frac{\tanh\left(\frac{x}{2}\right)+1}{2} \qquad (40)$$

For inputs that have positive values, the widely used ReLU activation function is open-ended and is described as

$$y(x) = \max(x, 0) \qquad (41)$$

Most of the time, the problem and the type of data determine which activation function is best. For networks with an output in the range [0, 1], for instance, "sigmoid" is appropriate. The neuron is rapidly saturated by the "tanh" and "sigmoid" activation functions, which have the ability to dissolve gradients. Ignoring that might lead to unstable dynamics in the weights' gradient updates due to the non-zero centered output of the "sigmoid ". Stochastic gradient descent (SGD) produces quicker convergence and sparser gradients than "sigmoid" or "tanh" activation functions. ReLU may be implemented with a border activation value of zero, making it computationally inexpensive.The neuron may stay dormant during training as the weight matrix expands since ReLU is not robust to strong gradient flows.

C. Loss Function

By comparing the corresponding target (zt) with the outcome (yt), that is known as

$$\mathcal{L}(y, z) = \sum_{t=1}^{T} \mathcal{L}_t(y_t, z_t) \qquad (42)$$

It is the total of all losses for every timestep. The particular problem must be considered while selecting a loss function. Popular loss functions for real-value forecasting include the Euclidean and Hamming distances, and for classification issues, cross-entropy over the probability distribution of the outputs.

Results and Discussion

This field investigation demonstrates how well the proposed model predicts cardiac disease. K-Nearest Neighbor (KNN), Random Forest (RF), and the proposed Enhanced Genetic Algorithm were compared to the performance of the recommended Enhanced Genetic Algorithm with Fuzzy Weight Updated Support Vector Machine (EGA-FWSVM). The suitability of the previously described machine learning models was evaluated through the use of many efficiency criteria.The best way to describe the overall predicting capabilities of the proposed deep learning model is accuracy. True positive (TP) and true negative (TN) scores are used to assess how well the classifier models predict whether or not a person has cardiac disease. The terms "false positives" (FP) and "false negatives" (FN) refer to the statistics that show how many incorrect forecasts the models produce. The cardiovascular diagnostic model's sensitivity and efficacy are evaluated using two metrics: precision and recall.
The precision is expressed as the proportion of all expected positive annotations to correctly acquired optimistic findings.

$$\text{Precision} = \text{TP/TP+FP} \qquad (43)$$

The ratio of accurately recognized positive findings compared to every finding within the actual group is referred to as sensitivity.

$$\text{Recall} = \text{TP/TP+FN} \qquad (44)$$

Musik in bayern
ISSN: 0937-583x Volume 89, Issue 4 (April -2024)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2024-*251*

The F1 score represents the accuracy and recall weighted average. It accommodates the two types of results: false positives and false negatives.

F1 score = 2*(Recall * Precision)/ (Recall + Precision) (45)

The following formula calculates accuracy in terms of positives and negatives:

Accuracy = (TP+FP)/(TP+TN+FP+FN) (46)

True Positive, False Positive, True Negative, and False Negative are defined as TP, FP, TN, and FN, respectively.

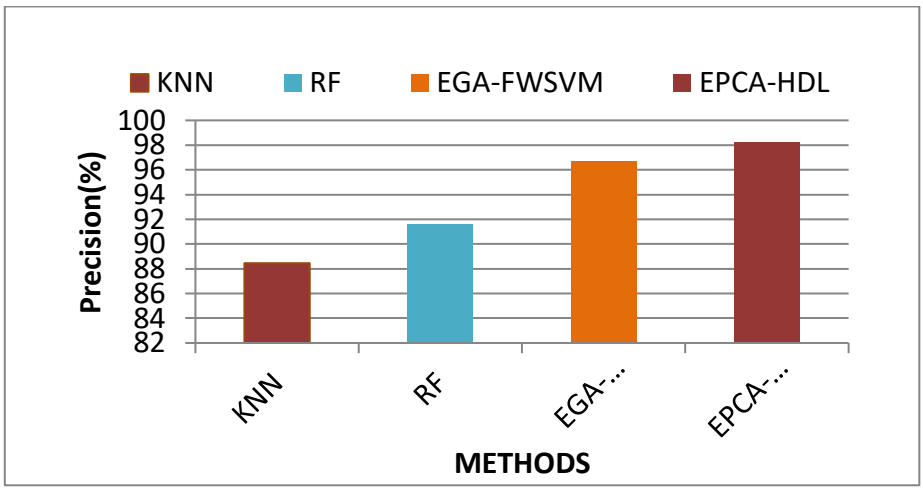


Figure 4. precise Comparison between the Suggested and Current Approaches

The results show that, in the event that adequate data is provided to the classifier modelthe suggested model may correctly predict heart disease. By employing improved feature extraction approaches, the suggested model performs better. These outcomes also show that the suggested model is capable of outperforming other models. Figure 4 shows that the RF method metric is 91.57%, the KNN method metric is 88.45%, the EGA-FWSVM technique has 96.7%, and the suggested EPCA-HDL method has an accuracy rate of 98.26%.

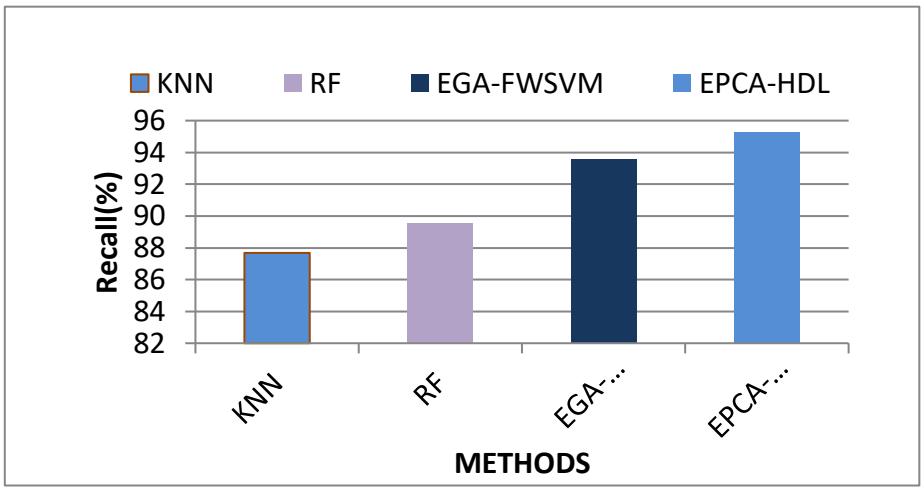


Figure 5. Recallthe distinction between the proposed and current approaches.

The proposed EPCA-HDL performs better in recall than earlier methods, as shown in Figure 5. With a recall of 95.27%, the EPCA-HDL approach is more effective than the RF method (89.54%), KNN method

Musik in bayern
ISSN: 0937-583x Volume 89, Issue 4 (April -2024)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2024-*251*

(87.68%), and EGA-FWSVM methodology (93.57%). Recall increases rapidly in the beginning of training, and using the recommended EGA-based feature selection, it is evident that this reduces the distance between the points, stabilizing the output.

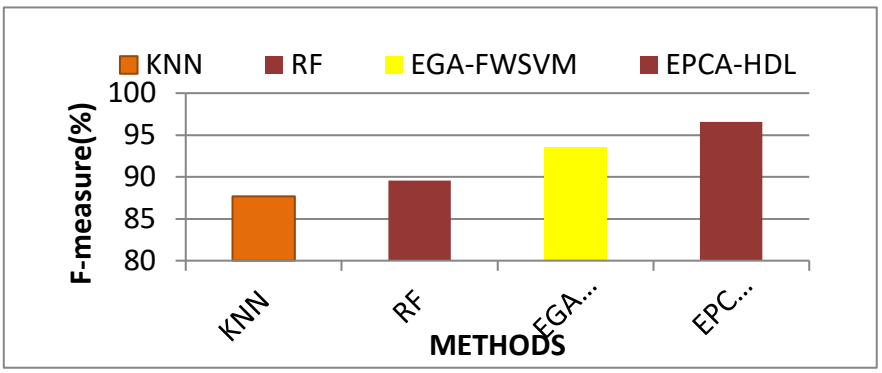


Figure 6. F-value review between the suggested and current approaches

As seen in Figure 6, the recommended EPCA-HDL outperforms previous approaches in terms of F-measure. The suggested system's performance was further enhanced by combining the feature selection approach with the Gaussian weight function methodology. After selecting the characteristics, we used both general and particular feature selection methods to assess their relevance.

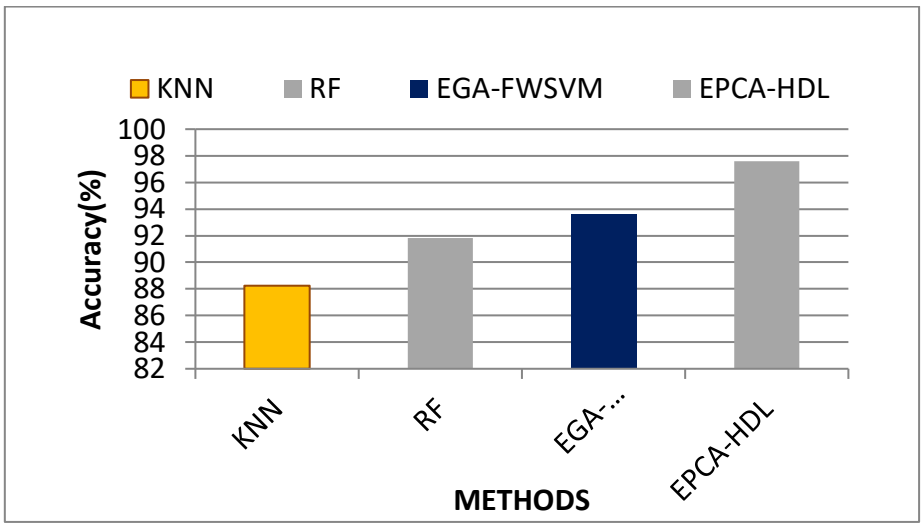


Figure 7. Evaluation of the Suggested and Current Methods' Accuracy

The suggested HDL model has the best accuracy (97.58%), as seen in Figure 7. This classifier has better accuracy than others. The obtained findings demonstrate that the suggested model may reliably predict cardiac disease if sufficient information is supplied to the classifier model. These findings also show that the model that was suggested outperforms existing models in data categorization with more characteristics. All baseline models' accuracy improved significantly after utilizing the suggested feature extraction and selection technique.

3. Conclusion

Understanding how to interpret unprocessed medical data about the heart can save lives in the long run and helpwithin the early recognizable proof of abnormalities in cardiac illnesses. In this work, rawinformation was handledutilizing machine learning strategies to create a special and uniqueconclusion of heart infection. The K-Means Clustering (KMC) technique is used for preprocessing in the beginning with the goal of improving classifier accuracy. Subsequently, the reduced dimensional feature subset is extracted using the feature extraction technique with the pre-processed data. This paper suggests using a modified Principal Component Analysis to

DOI https://doi.org/10.15463/gfbm-mib-2024-*251*

efficiently extract features.then after that comes the feature selection stage, which uses the EGA algorithm. Using data from the heart disease dataset, it determines the top wellness index, which is then used to identify additional pertinent and important characteristics. Ultimately, the heart disease data is classified using a half-breed profound learning model, such the Enhanced Elman Neural Network (EENN), which integrates elements that comprise the repetitive neural network (RNN) with hybridization. Affectability, Specificity, and Precision were used to compare the outcomes of the models developed using HDL techniques after features from the Cleveland heart sickness dataset were chosen. The experiment's findings show that crossover deep learning—which has been suggested—performs better than machine learning techniques in terms of anticipating heart disease information.

REFERENCES

1. Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Predictive data mining for medical diagnosis: An overview of heart disease prediction. *International Journal of Computer Applications*, *17*(8), 43-48.
2. Shah, D., Patel, S., & Bharti, S. K. (2020). Heart disease prediction using machine learning techniques. *SN Computer Science*, *1*, 1-6.
3. Palaniappan, S., & Awang, R. (2008, March). Intelligent heart disease prediction system using data mining techniques. In *2008 IEEE/ACS international conference on computer systems and applications* (pp. 108-115). IEEE.
4. Saxena, K., & Sharma, R. (2016). Efficient heart disease prediction system. *Procedia Computer Science*, *85*, 962-969.
5. Rajdhan, A., Agarwal, A., Sai, M., Ravi, D., &Ghuli, P. (2020). Heart disease prediction using machine learning. *INTERNATIONAL JOURNAL OF ENGINEERINGRESEARCH & TECHNOLOGY (IJERT)*, *9*(O4).
6. Katarya, R., & Meena, S. K. (2021). Machine learning techniques for heart disease prediction: a comparative study and analysis. *Health and Technology*, *11*, 87-97.
7. Kumar, M. N., Koushik, K. V. S., & Deepak, K. (2018). Prediction of heart diseases using data mining and machine learning algorithms and tools. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *3*(3), 887-898.
8. Rani, P., Kumar, R., Ahmed, N. M. S., & Jain, A. (2021). A decision support system for heart disease prediction based upon machine learning. *Journal of Reliable Intelligent Environments*, *7*(3), 263-275.
9. Maheswari, S., &Pitchai, R. (2019). Heart disease prediction system using decision tree and naive Bayes algorithm. *Current Medical Imaging*, *15*(8), 712-717.
10. Pattekari, S. A., & Parveen, A. (2012). Prediction system for heart disease using Naïve Bayes. *International journal of advanced computer and mathematical sciences*, *3*(3), 290-294.
11. Mythili, T., Mukherji, D., Padalia, N., & Naidu, A. (2013). A heart disease prediction model using SVM-decision trees-logistic regression (SDL). *International Journal of Computer Applications*, *68*(16).
12. Acharya, U. R., Fujita, H., Lih, O. S., Adam, M., Tan, J. H., & Chua, C. K. (2017). Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network. *Knowledge-Based Systems*, *132*, 62-71.
13. Liang, Y., Yin, S., Tang, Q., Zheng, Z., Elgendi, M., & Chen, Z. (2020). Deep learning algorithm classifies heartbeat events based on electrocardiogram signals. *Frontiers in Physiology*, *11*, 569050.
14. Vafaie, M. H., Ataei, M., &Koofigar, H. R. (2014). Heart diseases prediction based on ECG signals' classification using a genetic-fuzzy system and dynamical model of ECG signals. *Biomedical Signal Processing and Control*, *14*, 291-296.
15. Chen, J., Valehi, A., & Razi, A. (2019). Smart heart monitoring: Early prediction of heart problems through predictive analysis of ECG signals. *Ieee Access*, *7*, 120831-120839.
16. Alarsan, F. I., & Younes, M. (2019). Analysis and classification of heart diseases using heartbeat features and machine learning algorithms. *Journal of big data*, *6*(1), 1-15.
17. Jahmunah, V., Ng, E. Y. K., San, T. R., & Acharya, U. R. (2021). Automated detection of coronary artery disease, myocardial infarction and congestive heart failure using GaborCNN model with ECG signals. *Computers in biology and medicine*, *134*, 104457.
18. Sharma, P., Dinkar, S. K., & Gupta, D. V. (2021). A novel hybrid deep learning method with cuckoo search algorithm for classification of arrhythmia disease using ECG signals. *Neural Computing and Applications*, *33*, 13123-13143.
19. Neagoe, V. E., Iatan, I. F., & Grunwald, S. (2003). A neuro-fuzzy approach to classification of ECG signals for ischemic heart disease diagnosis. In *AMIA Annual Symposium Proceedings* (Vol. 2003, p. 494). American Medical Informatics Association.

DOI https://doi.org/10.15463/gfbm-mib-2024-*251*

20. Dalal, S., & Vishwakarma, V. P. (2021). Classification of ECG signals using multi-cumulants based evolutionary hybrid classifier. *Scientific Reports*, *11*(1), 15092.
21. Darmawahyuni, A., Nurmaini, S., Yuwandini, M., Rachmatullah, M. N., Firdaus, F., &Tutuko, B. (2020). Congestive heart failure waveform classification based on short time-step analysis with recurrent network. *Informatics in Medicine Unlocked*, *21*, 100441.
22. Shankar, M. G., Babu, C. G., & Rajaguru, H. (2022). Classification of cardiac diseases from ECG signals through bio inspired classifiers with Adam and R-Adam approaches for hyperparameters updation. *Measurement*, *194*, 111048.
23. Wagstaff, K., Cardie, C., Rogers, S., &Schrödl, S. (2001, June). Constrained k-means clustering with background knowledge. In *Icml* (Vol. 1, pp. 577-584).
24. Labrín, C., &Urdinez, F. (2020). Principal component analysis. In *R for political data science* (pp. 375-393). Chapman and Hall/CRC.
25. Mirjalili, S. (2019). Genetic algorithm. In *Evolutionary algorithms and neural networks* (pp. 43-55). Springer, Cham.
26. Cheng, Y. C., Qi, W. M., & Cai, W. Y. (2002, November). Dynamic properties of Elman and modified Elman neural network. In *Proceedings. International Conference on Machine Learning and Cybernetics* (Vol. 2, pp. 637-640). IEEE.
27. Salehinejad, H., Sankar, S., Barfett, J., Colak, E., &Valaee, S. (2017). Recent advances in recurrent neural networks. *arXiv preprint arXiv:1801.01078*.