IMPROVING CARDIOVASCULAR DISEASE DETECTION WITH AUTOE-SOM THROUGH DEEP LEARNING AND FEATURE SELECTION

B. Dhiyanesh[1], Parveen Begam Abdul Kareem[2], P. Shanmugaraja[3*], V. Anusuya[4], A. Baseera[5]

[1]Associate Professor,Department of Computer Science Engineering, SRM Institute of Science and Technology, Chennai,Tamil Nadu 600026,India.**Email: dhiyanu87@gmail.com**

[2]Professor, Department of Computer Science Engineering, Taibah University, Saudi Arabia, Yanbu 46422, Saudi Arabia.**Email: parveenkareem@gmail.com**

[3*]Associate Professor, Department of Information Technology, Sona College of Technology, Salem, Tamil Nadu636005, India.**Email: shanmugarajap@gmail.com***

[4]Associate Professor, Department of Computer Science Engineering, Ramco Institute of Technology, Rajapalayam, Tamil Nadu626117, India. **Email: pgkrishanu@gmail.com**

[5]Assistant Professor/School of,Department of Computer Science Engineering, VIT Bhopal University, Bhopal,Madhya Pradesh – 466114, India.**Email: basheera.zain@gmail.com**

Abstract

CVD is responsible for numerous deaths worldwide and impacts millions of individuals annually. Its impact spans all demographics, making it a major public health concern worldwide. In CVD, the heart does not pump sufficient blood to the rest of the body. A precise analysis of CVD is critical for its prevention and treatment. CVD encompasses four categories, including coronary heart disease, transient ischemic attack, peripheral artery disease, and aortic disease. These symptoms often manifest in the elderly and can be mistaken for other diseases, making an accurate diagnosis challenging and potentially leading to fatal outcomes. ML and DL algorithms play a crucial role in disease analysis, as they can categorize or predict outcomes. However, a common challenge in ML is the high dimensionality of data, which complicates feature selection. Thus, operations performed on this data require more training, leading to training loss, low precision, and recall rate to degrade the detection accuracy. An Autoencoder Methodology Combined with the Self Organizing Maps (AutoE-SOM) algorithm was proposed to resolve the issue. Initially, the CVD dataset at the Kaggle repository includes the healthcare margins. The proposed system works under preprocessing, feature selection, and classification phases. The

CVD dataset will be preprocessed based on the normalization method in the starting phase. It converts features in CVD datasets to a standard scale, improving the efficiency and accuracy of output values. The primary purpose of normalization is to remove possible biases and distortions that may occur due to differences in feature sizes. The second phase is featuring selection; the relevant features related to CVD are extracted from the dataset using the SVM method. The third phase is classification. The AutoE-SOM assists in comprehending the fundamental structure of the dataset, recognizing patterns, and enhancing classification accuracy through prioritizing the most significant features. This novel aims to develop a reliable and accurate CVD detection model, which could contribute to healthcare specialists in initial diagnosis and interference, eventually leading to improved patient outcomes. Investigational outcomes determine that our method beats the previous method in precision, recall, F1 score, accuracy, and time complexity.

Keyword: Cardiovascular Disease, Autoencoder, Self-Organizing Maps, Feature Selection, Normalization, Classification, Healthcare

I. INTRODUCTION

The proposed method effectively classifies the cardiovascular disease dataset by integrating dimensionality reduction techniques with clustering techniques. By combining an Autoencoder and a self-organizing map (SOM), we not only solve the complexity of high-dimensional data but also enhance the efficiency of disease detection. This reassures the audience about the practicality and applicability of the methodology in real-world scenarios.

Auto encoders, a class of neural networks, are pivotal in the proposed method. Their crucial role lies in learning well-organized encodings of input information. By guiding input data through a bottleneck layer, auto-encoders proficiently reduce noise and sift out irrelevant data. This crucial process enables extracting meaningful and relevant features from the dataset, thereby significantly enhancing the accuracy of disease detection.

However, SOMs complement this by organizing data to make them useful for identifying patterns and clusters in complex and high-dimensional datasets. Therefore, SOMs are very useful in visualizing relationships among the different data points and give an overview of how the CVD data have been structured. The auto-encoder first compresses data, then SOMs cluster this reduced representation, making classification manageable and precise.

The potential impact of the proposed methodology on patient care and healthcare costs is significant and promising. By providing a dependable and accurate method of detecting CVD, the new methodology can enable earlier diagnosis and more effective interventions, leading to better patient outcomes. Moreover, the high accuracy and efficiency of the method can help reduce healthcare costs by minimizing misdiagnosis and related therapies. This dual advantage underscores the significant potential impact of the AutoE-SOM methodology in medicine, offering hope for improved patient care and cost-effective healthcare.

II. Literature Survey

The author [1]discussed that compared to other diseases in the world; heart syndrome is the most noxious disease that kills most people. However, with the potential of the GAN-LSTM methodology, we have a promising tool to improve the prediction and diagnosis of this deadly disease. The main disadvantage of GAN-LSTM, creating data imbalance, is a challenge we are working to overcome.

Introduced [2]the Cardio-Help methodology, based on Convolutional Neural Network (CNN), as a potential game-changer in estimating a patient's CVD odds. The method's requirement for high accuracy in image classification underscores its potential to significantly improve patient care.Underscored the urgency of predicting CVD before it becomes critical[3]. They introduced an ensemble framework for this purpose, despite its high time complexity. The inability of the proposed framework to perform premature patient analyses and its potential impact on the initial prediction of cardiovascular diseases were also highlighted.

In ref [4]discussed that, given the global situation, the initial detection of heart syndrome is a significant challenge. If not diagnosed immediately, it can cause death. However, with the deployment of a hybridized feature selection methodology, we have a tool that can provide highly accurate results in image classification, giving us confidence in our ability to tackle this significant challenge. [5]Discussed that predicting heart syndrome is a well-thought-out, uttermost medical testing mission. However, with the deployment of a GridSearchCV methodology, we have a method that efficiently matches the Extreme Gradient Boosting Classifier methodology to generate optimal hyperparameters, making our testing process more accurate and effective.

The author [6]carried out artificial intelligence software that that clearly categorizes data to analyze it effectively and make better predictions. To analyze different heart diseases, use a baseline method that can run different data processing protocols. However, it could be a manageable crossing point for ease of use.

[1]Mehmood, et al., Prediction of Heart Disease Using Deep Convolutional Neural Networks.

[2] Tiwari, A., et al.,Ensemble framework for cardiovascular disease prediction. Computers in Biology and Medicine.

[3] Rani, et al., A decision support system for heart disease prediction based upon machine learning. https://doi.org/10.1007/s40860-021-00133-6.

[4]G. N. Ahmad et al.,"Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques With and Without GridSearchCV.

[5]A. L. Yadavet et al, "Heart Diseases Prediction using Machine Learning," (ICCCNT), Delhi, India, 2023, pp. 1-7, doi: 10.1109/ICCCNT56998.2023.10306469.

[6]Slart, R.H.J.A., et al. Position paper of the EACVI and EANM on artificial intelligence applications in multimodality cardiovascular imaging using SPECT/CT, PET/CT, and cardiac CT.

The author [7] carried out nuclear cardiology and CT protocols to sustain the clinical function of ML in dealing with the CVD dataset. CVD dataset technology can further improve the analysis and forecast of patients with cardiovascular disease. However, the proposed methods for interpreting, quantifying, and integrating these data sets have limitations.

The selected features were compared with a given heart disease database using a Back-Propagation Artificial Neural Network (BP-ANN). The proposed method achieves higher accuracy in experimental results than the previous technique [8].To resolve the issue, a Multi-Layer Perceptron (MLP) methodology was deployed. This advanced method uses artificial neural network DL methodology. The deployed method uses DL and data mining to obtain reliable and error-free results. The author [9]carried out an ML analytical methodology for coronary artery disease that has been technologically advanced with various ML methodologies. The proposed method can be translated into construction instructions and castoff to progress specialist systems to diagnose CAD patients. The limitation is that the proposed method does not identify the hidden patterns in the large dataset. The author [10]suggested Hyper Parameter Optimization (HPO) methodology, which was deployed to execute better the XG Boost and Random Forest (RF) classifier methodology. The major drawback is that machine learning models combined with optimization techniques have very low predictive accuracy for heart disease. The author [11]proposes an Advanced Attention Capsule Network (AACNet) architecture to manage the complexity of diverse medical images. AACNet integrates a multifunctional extractor with SPP layers, a multi-level capsule network, and a dynamic channel attention module. The IBK classifier was adjusted by the author [12] to change the number of neighboring hyper parameters. IBK's drawback was its low accuracy when "k" hyperparameter was modified with the chi-squared attribute set.

[7] M, Y. B, S. S and S. M, "Cardiovascular Disease Prediction using Deep Learning," 2022 6th International Conference on Trends in Electronics and Informatics .

[8]Gopala Krishnan, et al., "Skin Cancer Detection and Classification using BP-ANN and SGLD." Turkish Journal of Computer and Mathematics Education .

[9] R., & Sheela, et al., Heart disease prediction using hyper parameter optimization (HPO) tuning. Biomedical Signal Processing and Control.

[10]Sekar, J., Aruchamy, et al., An efficient clinical support system for heart disease prediction using TANFIS classifier. Computational Intelligence, Vol 38(2).

[11]G. Maheswari et al., "Dynamic Channel Attention for Enhanced Spatial Feature Extraction in Medical Image Analysis using Advanced Attention Capsule Network.

[12]Rai, H.M., Chatterjee, et al., CNN-LSTM deep learning model and ensemble technique for automatic detection of myocardial infarction using big ECG datahttps://doi.org/10.1007/s10489-021-02696-6

The author [13] proposed it is challenging to plan and exactly predict myocardial infarction using ECG data for heart disease diagnosis and treatment. To resolve the issue, a hybrid CNN—LSTM network was deployed. The deployed network not only resolved the information imbalance issue but also significantly improved the accuracy of the minority classifier.

The author [14]discussed that predicting cardiovascular disease is a medical ML challenge. To prevent the development of the disease, it is necessary to identify those at risk early. A Stacked Sparse Auto-encoder (SSAE) methodology was deployed to resolve the issue. The main issue is that the proposed method parameters affect the network's generalization ability.The proposed CNN system processes the input dataset for disease prediction, allows the data to be analyzed, and provides relevant findings [15].The author [16]carried out a Gradient Descent Optimization with CNN (GDO-CNN) methodology, which was deployed to automatically analyze and sort heart circumstances using ECG to discourse the growing death amount from CVD. The proposed method is a low-precision, inefficient diagnostic tool that fails to address pressing cardiovascular health challenges. The author [17] carried out InceptionV3 collated with VGG16 (IV3-VGG16) methodology to estimate CVD hazard levels via willingly obtainable and non-invasive fundus images. It plays an important castoff in obtaining complex outlines and features from data, enabling exact guesses of CVD hazards. By combining these features with clinical data, models are trained to predict cardiovascular hazard rates[18].

[13] Ammal, S.G., Saranya, K. et al. Advanced Cloud-Based Prediction Models for Cardiovascular Disease: Integrating Machine Learning and Feature Selection Techniques.
[14] Mienye, I. D., et al., Improved Heart Disease Prediction Using Particle Swarm Optimization Based Stacked Sparse Autoencoder. Electronics, 10(19), 2347.
[15] Gopalakrishnan, S., et al. "A Novel Deep Learning-Based Heart Disease Prediction System Using Convolutional Neural Networks Algorithm." International Journal of Intelligent Systems and Applications in Engineering .
[16] Gopinath P, Shivakumar R, "Classification of vein pattern recognition using hybrid deep learning", Journal of Intelligent & Fuzzy Systems.
[17] Singh, N., Gunjan, V.K., Shaik, F. et al. Detection of Cardio Vascular abnormalities using gradient descent optimization and CNN.
[18] Jeba Sheela, A., & Krishnamurthy, Revolutionizing cardiovascular risk prediction: A novel image-based approach using fundus analysis and deep learning. Biomedical Signal Processing and Control, 90.

This review illustrates the various Machine Learning (ML) and Deep Learning (DL) -based approaches introduced for CVD detection. Yet, CVD detection has some issues, such as high time complexity, failure to provide accurate results due to the high dimensionality of the dataset, and poor detection performance. Furthermore, the existing methods didn't focus on feature selection and normalization processes. Thus, the proposed method's objective is precision and reliable CVD detection from the collective dataset. The proposed proficiently reduces the dimensionality and normalizes the collective dataset. In addition, the proposed method improves classification accuracy and low time complexity for CVD detection.

III. IMPLEMENTATION OF PROPOSED METHOD

This section defines the inclusive process of the proposed method. Figure 1 shows the details of the deployed method. In this novel, we get the CVD dataset from the Kaggle Website. Using the Normalization method, the CVD dataset recognizes patterns and makes predictions based on labeled data. After pre-processing, the features related to CVD are extracted from the dataset using the SVM method. Our method develops a reliable and accurate model for CVD detection, which could support healthcare experts in premature diagnosis and interference, eventually getting better patient consequences and reducing healthcare costs.

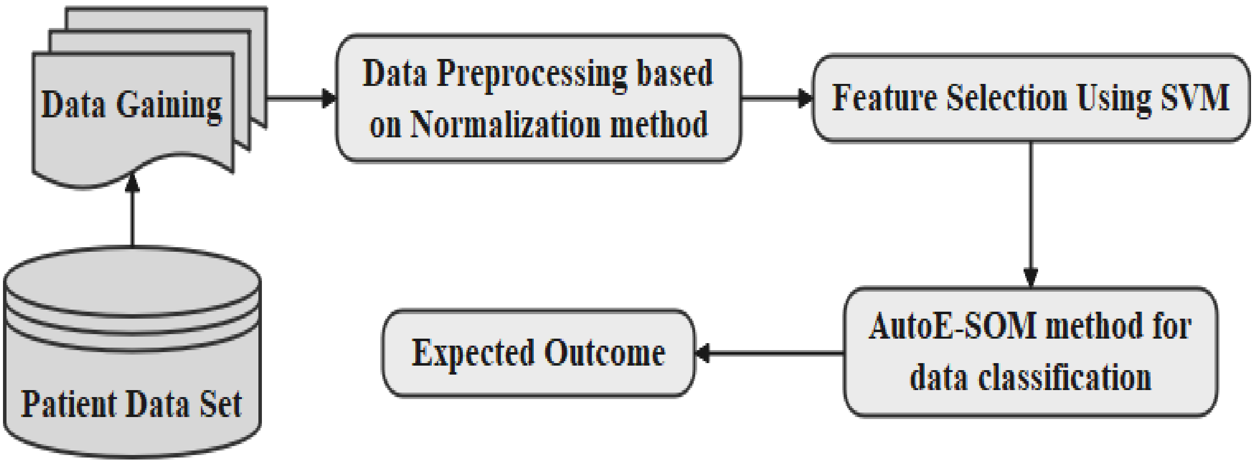

Fig. 1. The Proposed AutoE-SOM Architecture Diagram

A. Dataset Collection

The Cardiovascular_Disease_Dataset is used in this novel, taken at the Kaggle website. As described, the dataset castoff in this novel consisted of 1000 patient records with 14 different characteristics listed in the CVD dataset. These characteristics include age, gender, chest pain, restingBP, etc., The target category indicates whether the patient has CVD or is healthy. Figure 2 shows the Cardiovascular_Disease_Dataset description.

S. No	Attribute	Allocated Code	Range
1	Patient Identification Number	patientid	Number
2	Age	age	In Years
3	Gender	gender	1,0(0=female, 1=male)
4	Chest Pain Type	chestpain	0,1,2,3 0=typical angina 1=atypical angina 2=non-anginal pain 3=asymptomatic
5	Resting Blood Pressure	resting BP	94-200 (in mm HG)
6	Serum Cholesterol	serumcholesterol	126-564 (in mg/dl)
7	Fasting Blood Sugar	fastingbloodsugar	0,1 > 120 mg/dl (0 = false, 1 = true)
8	Resting electrocardiogram results	restingrelectro	0,1,2 (Value 0: normal, Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV), Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria)
9	Maximum heart rate achieved	maxheartrate	71-202
10	Exercise induced angina	exerciseangia	0,1 (0 = no, 1 = yes)
11	Oldpeak =ST	oldpeak	0-6.2
12	Slope of the peak exercise ST segment	slope	1,2,3 (1-upsloping, 2-flat, 3-downsloping)
13	Number of major vessels	noofmajorvessels	0,1,2,3
14	Classification	target	0,1 (0= Absence of Heart Disease, 1= Presence of Heart Disease)

Fig. 2. Cardiovascular Disease Dataset description

B. Z-score Normalization method

In this section, the CVD dataset is preprocessed using the Normalization method to transform the columns in a dataset to the same scale referred to as normalization. Normalization converts values in a data set to a standard quantity in data mining. The deployed methodology is cast to normalize a dataset's characteristics to a certain predefined standard. This allows you to remove unnecessary or noisy material and use valid and reliable data. This allows you to influence and improve the exactness of deployed decisions. Normalization is essential as it improves the exactness of the consequences gained through clustering. To get better results, various methodologies can be used to perform data preparation, including data preprocessing and normalization. Three methods for data normalization include decimal scaling, Z-Score, and min-max normalization.

The min-max methodology requires a linear transformation of the fundamental attribute. MinX and MaxX are the least and most determined assesses of characteristic X, respectively. This methodology maps the X characteristic's assessment in the range [0, 1]. Equ 1 demonstrates the calculation of the min-max regularization method.

$$m' = \frac{m\_MinX}{MaxX\_MinX} \quad (1)$$

When the authentic least and determined values of a characteristic X are unidentified, the Z-score normalization technique for characteristic values using X's mean and standard deviation attributes is proper. Equ 2 shows the total of the Z-score normalization method. $\bar{X}, \sigma X$ in Equ2and m are the mean, standard deviation, and value of characteristic A, respectively,

$$m' = \frac{m - \bar{X}}{\sigma X} \quad (2)$$

Decimal normalization methodology normalizes the value by shifting the decimal point of the attribute value X. The number of places to move the decimal point is contingent on the maximum integer of X value. Equ 3 illustrates the calculation of the normalization method of the decimal place. In Equ 3, s is the value of characteristic X and v is the lowest integer where Max (|s'|) <1.

$$s' = \frac{s}{10^v} \quad (3)$$

Normalizing the CVD dataset is essential for accurate and efficient model training. Normalization prevents specific features from dominating the learning process by standardizing feature scales, guaranteeing that the approach can efficiently acquire all features.

C. Support Vector Machine (SVM) Method

After preprocessing, the dataset is ready to select the features based on the SVM method. SVM makes better decisions on data points outside the training set. Feature selection provides a small but unique subset of the initial data, selecting unique features from the feature set and removing irrelevant features. Identifying small but essential features aims to diminish the dimensionality of the information. This both reduces the processing time and improves the classification accuracy. The closest points to the separating hyperplane are called support vectors. Kernels can be used to scale SVM to large data sets. The kernel performs mapping from one feature space to another. The SVM implementation responds to the resource of parameters and provisions of the core being cast. The use of SVM in pattern distinguishing is discussed below.

A n-dimensional pattern $i$has $n$coordinates, each of $i$which $i = i_1, i_2, \ldots, i_n$is aactualdigit, $i_x \in R$for $x = 1,2,\ldots,n$toeachpattern$i_y$have its place to a class$j_y \in \{-1,+1\}$. We also deliberate a training set $G$of $r$patterns in concert with their classes$G = \{(i_1, j_1), (i_2, j_2), \ldots, (i_r, j_r)\}$, and the dot product space Din which the patterns $i$are entrenched$i_1, i_2, \ldots, i_r \epsilon D$. Any hyperplane in space can be written this way.,

$$\{i \in D | u_. i + c = 0\}, u\epsilon D, c\epsilon R \quad (4)$$

The $u_. i$dot product is definite by,

$$u_. i = \sum_{x=1}^{n} u_x . i_x \tag{5}$$

A set of training patterns can be partitioned linearly if there is at least one representation by a pair(u, c) of linear classifiers that suitablycategorizesentire training patterns. This linear classifier is signified by a hyperplane $P(u_. i + c = 0)$defining a region of +1 patterns $(u_. i + c > 0)$ and another region of -1 patterns $(u_. i + c < 0)$.

After the training process is completed, the classifier can predict the class members of new patterns. The type of the model is determined by i_q the following equation:

$$class(x_q) = \begin{cases} +1 \; if u_. i_q + c > 0 \\ -1 \; if u_. i_q + c < 0 \end{cases} \tag{6}$$

The sorting of update patterns countson only the symbol of countenance$u_. i + c$.

The training phase of SVM uses Sequential Minimal Optimization (SMO). The SMO algorithm is a common optimization technique used for training SVM. A dual representation of the SVM master optimization problem is shown in (7).

$$\max_{\alpha} \varphi(\alpha) = \sum_{x=1}^{N} \alpha_x - \frac{1}{2} \sum_{x=1}^{N} \sum_{y=1}^{N} j_x j_y q(i_x, i_y) \alpha_x \alpha_y \\ subject to \sum_{x=1}^{N} j_x \alpha_x = 0, \quad 0 \le \alpha_x \le F, x = 1, \ldots, n, \tag{7}$$

i_x is a training example, $j_x \in \{-1, +1\}$ is the matchinggoalassessment, α_x is the lagrange multiplier, and F is unactual assessment cost parameter.

The dataset analysis using this approach can reveal the connections between features and the target variable, offering valuable insights for deeper analysis or domain comprehension.

D. AutoE-SOM Method

At last, the CVD dataset is classified using the AutoE-SOM method. In the proposed

 ISSN: 0012-0413

2025 123(4)

method, we can understand the high dimensional data. It initializes weights for each characteristic and selects an attribute from the training data. It examines every characteristic to find which weights are more likely to input attributes. The proposed method is structured to convert an input into different representations. Attempt to renovate the actual input as accurately as possible—the AutoE-SOM original input to the hidden layer. Latent layers are treated as latent spatial representations. The potential representation is reconstructed by the decoder using the progressions for encoding and decoding as outlined in Equations 8 and 9.

$$l = \sigma(Ga + c), \tag{8}$$

$$\hat{a} = \sigma(Gl + c), \tag{9}$$

Here, $a = (a_1, a_2, \dots, a_n)$is the input data vector and $l = (l_1, l_2, \dots, l_n)$is the truncated dimensional vector derived from the unseen layer. $\hat{a} = (\hat{a}_1, \hat{a}_2, \dots, \hat{a}_n)$is the reassembled entry. GAnd$G^1$are the weight matrices, $c$and $c^1$are the bias vectors, and σsignifies the S-shaped activation function. i.e., $\sigma = \frac{1}{1+e^{-a}}$. Use the mean-square error process as the renovation error process among l and $\hat{a}$.

$$T = \frac{1}{Q}\sum_{x=1}^{Q} ||\hat{a}_x - a_x||^2 \tag{10}$$

A common challenge when training an auto-encoder network is overfitting. A better solution would be to place a weight penalty on the cost function.

$$T = \frac{1}{Q}\sum_{x=1}^{Q}\frac{1}{2}||\hat{a}_x - a_x||^2 + \frac{\lambda}{2}\left(||G||^2 + ||G'||^2\right), \tag{11}$$

coefficient of weight decay is represented as λ. Also, a sparsity consequences pan is presented in the unseen layer of the autoencoder to accomplish improvedFL further down sparsity limitations and evade the circumstances where the autoencoder replicas input value to the output. $\hat{z}$represents the regular activity of the unseen layer characteristics, definite as $\hat{z}_y = \frac{1}{Q}\sum_{x=1}^{Q} p_y\,(a_x)$, and z is the sparsity fraction, which is usually a positive assessment adjacent to 0. To reach the interval, $\hat{z}_y = z$is the thresholdand the Kullback-Leibler (KL) difference is introduced as a regularization term in the loss function.

$$KL(\hat{z}||z) = \sum_{y=1}^{k} z\log\left(\frac{z}{\hat{z}_y}\right) + (1 - z)\log\left(\frac{1-z}{1-\hat{z}_y}\right) \tag{12}$$ 1

where K is the numeric of unseen characteristic. Therefore, the sparse autoencoder's loss process now includes three portions: mean squared error, weight decay, and sparse regularization.

$$T = \frac{1}{Q}\sum_{x=1}^{Q}\frac{1}{2}||\hat{a}_x - a_i||^2 + \frac{\lambda}{2}(||G||^2 + ||G'||^2) + \beta KL(\hat{z}||z), \tag{13}$$

In this case, β represents a sparse regularization parameter, and multiple sparse autoencoders are arranged in a stack in order to achieve enhanced FL. It is essential to combine the encoding layer with the input layer of the subsequent SOM autoencoder in this structural configuration to facilitate improved representational learning.

In AutoE-SOM technology, the previous sparse autoencoder's hidden layer is used as the input for the following sparse autoencoder. After training different sparse autoencoders, it's crucial to remove the decoder layers since the learned features are stored in the hidden layer. The last hidden layer is combined with a softmax classifier for the purpose of conducting classification. The suggested AutoE-SOM technique comprises a trained sparse autoencoder and a SoftMax classifier.

$$T = -\frac{1}{h}\left[\sum_{x=1}^{h}\sum_{y=1}^{Q} 1\{j^x = y\} log \frac{e^{\theta_y^\tau a^x}}{\sum_{p=1}^{Q} e^{\theta_y^\tau a^x}}\right], \quad (14)$$

here, $1\{\cdot\}$signifies the indicator process, i.e., $1\{j^x = y\} = 1$ if$j = 1$. $1\{j^x = y\} = 0\ if j \neq y$,$Q$ denotes the numeric type, and $\theta_x$symbolizes the weight matrix that associates the ithoutput units.

In the proposed method, all particles $z_x(t)$are arbitrarily adjusted to different situations$a_x(t)$in the examination space with speed$s_x(t)$. Then, AutoE-SOM is constructed for training by using the values of$a_x(t)$. The proposed method intentions to examination the space for the least fitness assessment. Let's assume $l_x(t)$ as the fitness assessment. At each iteration, $l_x$is calculated viaEqu 13, and the optimal state f is set if the constant $a_x(t)$is the least fitness assessment present in the group. If$l_x(t)$ is the least fitness assessment reckoned by the $z_x(t)$, then $a_x(t)$ is castoff as the best position $s_x(t)$. The speed of all maps is modernized on the basis of their position, present speed, and global population information. Map speed and position updates are implemented using Equations (15) and (16), respectively.

$$s(t+1) = \omega s(t) + u_1 h_1\big(s_x - a(t)\big) + u_2 h_2\big(f - a(t)\big), \quad (15)$$

$$a(t+1) = a(t) + s(t+1), \quad (16)$$

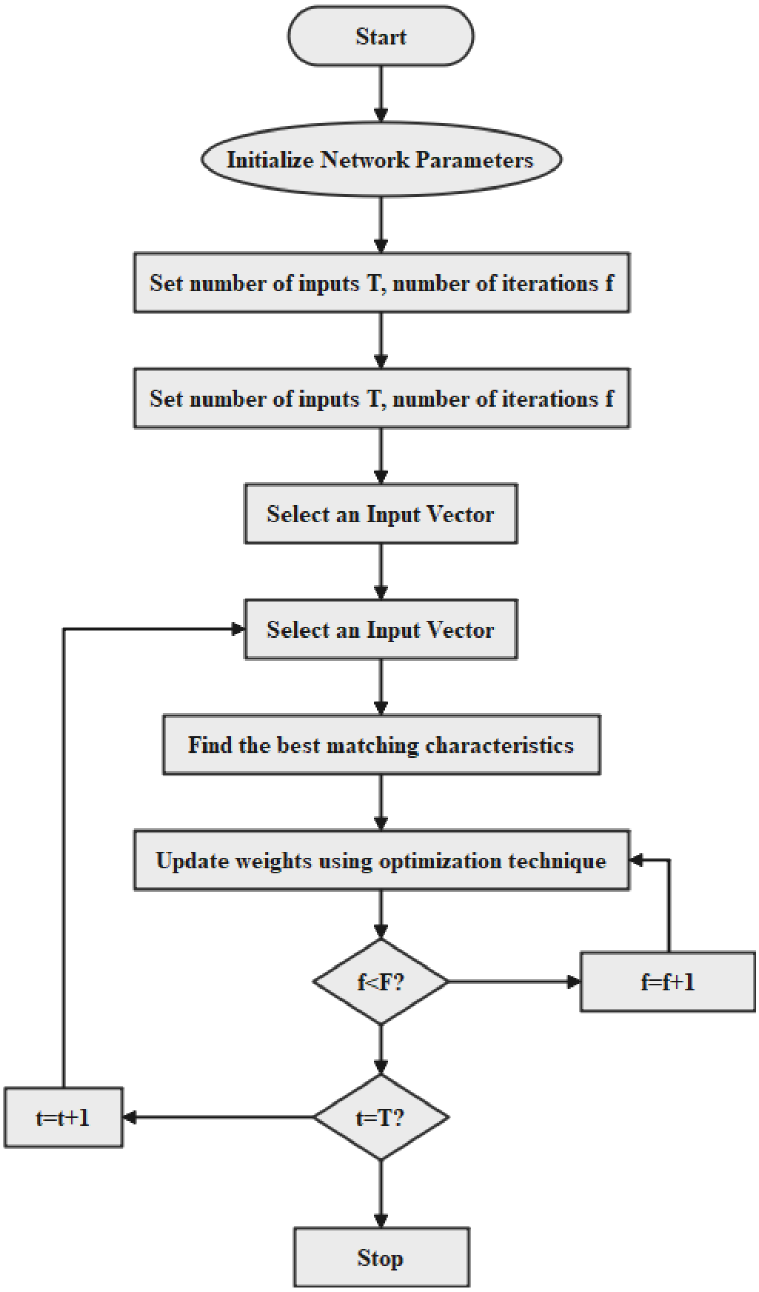

Fig 3. Flow Chart Diagram of AutoE-SOM method

here, $s(t+1)$denotes the speed of the maps at interval$t+1$, ω denotes the inertial weight, $v(t)$ is the present speed, c_1 and c_2 are the acceleration coefficients, whose assessments are 1.5 and 2, and correspondingly, where r_1 and r_2 are arbitrary statistics amongst 0 and 1. After a certain statistics of iterations, f has the optimal way out, which is used to adjust the parameters of AutoE-SOM.

AutoE-SOM operates by converting high-dimensional data into a low-dimensional grid, with each cell representing a unique feature or pattern in the dataset. This method can be utilized to analyze complex datasets like clinical imaging, patient medical history, and genetic data to identify patient subgroups or individuals with specific illnesses. It evaluates individuals based on their symptoms and medical history, pinpointing subgroups that may benefit from specialized treatment. AutoE-SOM is a crucial tool for clinical diagnostics due to its ability to handle intricate datasets and reveal correlations and patterns that are challenging to identify using traditional statistical methods. Figure 3 illustrates a flowchart of this method, providing an overview of the entire process.

The AutoE-SOM method shows promise for CVD datasets. The proposed method can effectively classify CVD datasets by leveraging the unsupervised learning capability of autoencoders to extract meaningful features and the clustering ability of SOMs to organize these features. This method can improve classification accuracy and robustness, making it a valuable tool for CVD detection and diagnosis. However, further research and experimentation are needed to validate its effectiveness compared to other methods and to optimize its parameters for different datasets and scenarios.

IV. RESULT & DISCUSSION

The deployed method performance is evaluated for investigation using precision, recall, F1 score, and time complexity. This assessment is determining a reliable and accurate method for CVD detection using the deployed method with GDO-CNN, SSAE, and HPO. Table 1 shows the simulations and parameters of the paper. The CVD dataset is used to classify the patient history and contains 30855 datasets. The implementation method is run using Python and the Anaconda Tool.

Table 1. Modeling Parameters

Parameters	Values
Dataset Name	Cardiovascular_Disease_Dataset
No. of Records	1000
Language	Python
Tool	Anaconda

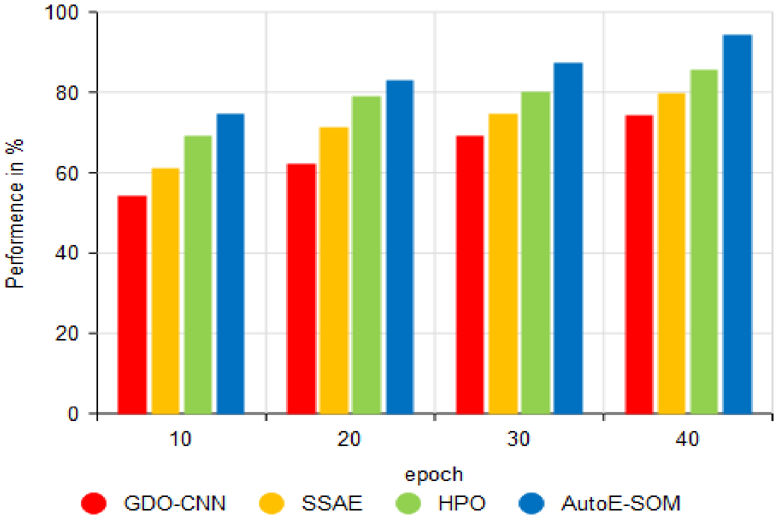

Fig.4. The precision's performance

Figure 4 shows that the precision execution of the GDO-CNN is 74.23%, SSAE is 79.64%, HPO is 85.63%, and the precision performance of the AutoE-SOM is 94.32%. The deployed methodology has better precision than the previous methodology. It promotes better generalization, allowing the network to accomplish well on unseen information. The precision is essential for datasets that need high accuracy, as it guarantees that the model can make precise predictions for new, unseen instances.

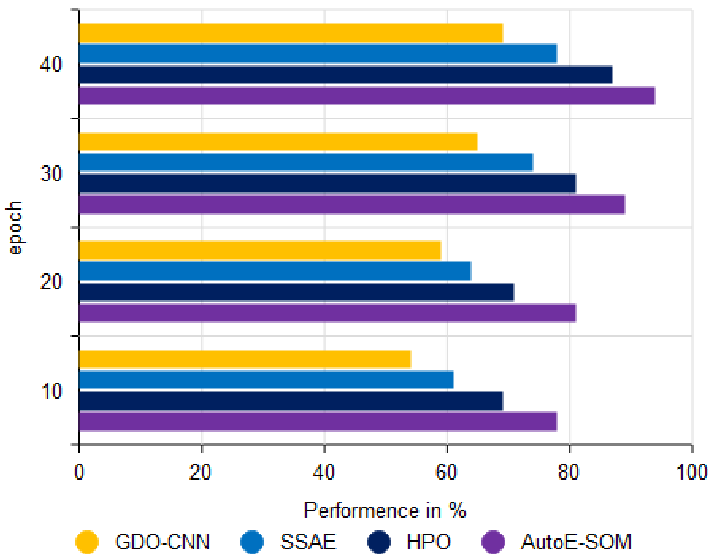

Fig.5. The performance of the recall in %

Figure 5 expresses that the recall performance of the GDO-CNN is 69%, SSAE is 78%, HPO is 87%, and the recall performance of the AutoE-SOM is 94 %. The deployed methodology has a better recall value than the previous methodology. This is advantageous in tasks where it is decisive to diminish incorrect rejections, for instance, in medical examiner, where the absence of an optimistic case can have serious consequences.

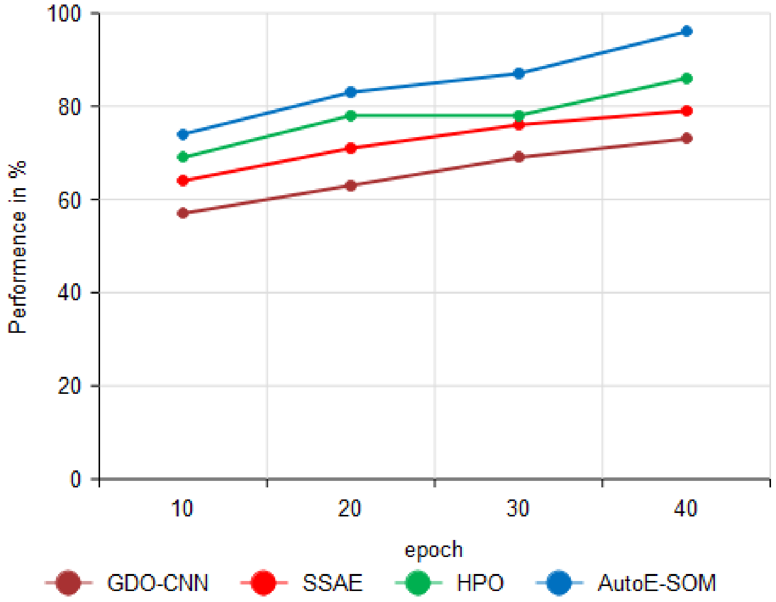

Fig.6. The performance of the accuracy in %

Figure 6 illustrates that the accuracy performance of the GDO-CNN is 73%, SSAE is 79%, HPO is 86%, and the recall performance of the AutoE-SOM is 96 %. The deployed methodology is more accurate than the previous methodology. It achieves parameter efficiency, ultimately leading to better performance in various tasks, including classification tasks of CVD detection.

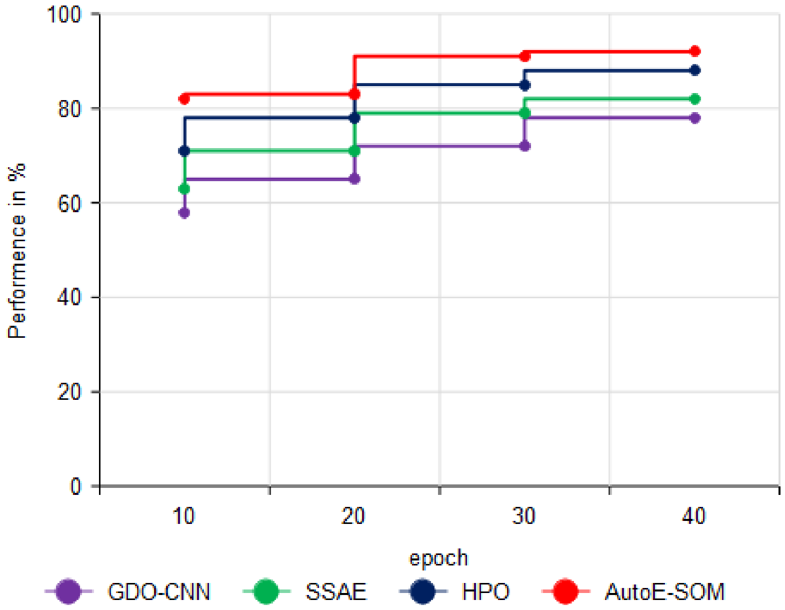

Fig.7. The performance of the F1 Score in %

Figure 7 illustrates that the recall execution of the GDO-CNN is 78%, SSAE is 82%, HPO is 88%, and the recall performance of the AutoE-SOM is 92 %. The deployed methodology has a better F1 score value than the previous methodology. The F1 score is intuitive to interpret, representing the harmonic mean of recall and precision. An elevated F1 score specifies improved execution in both recall and precision relationships. It is a valuable metric for evaluating AutoE-SOM models on CVD datasets, so long as a composed assessment of their execution that is sensitive to together false positives and false negatives.

Table 2. Comparative Analysis of the Deep Learning Method

Model	Precision	Recall	Accuracy	F1 Score
GDO-CNN	74.23%	69%	73%	78%
SSAE	79.64%	78%	79%	82%
HPO	85.63%	87%	86%	88%
AutoE-SOM	94.32%	94%	96%	92%

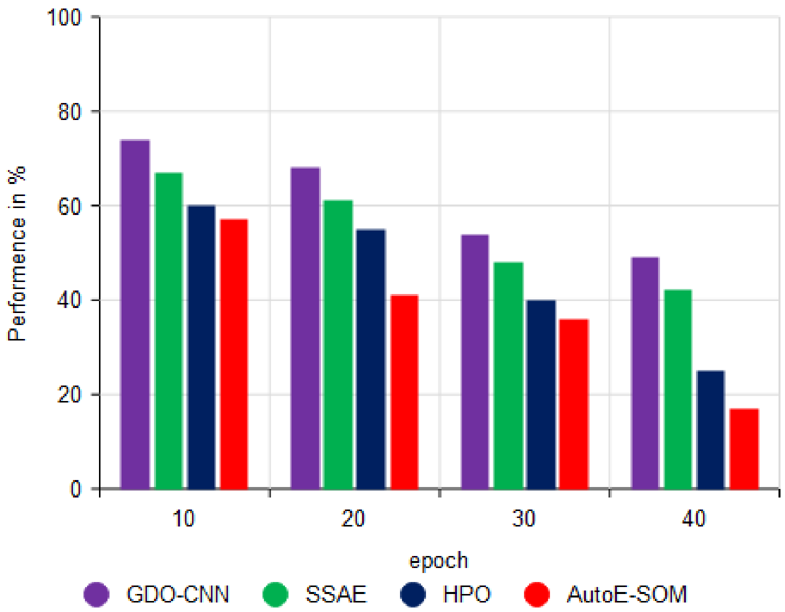

Fig.8. The performance of the time complexity in %

Figure 8 illustrates that the time complexity performance of the GDO-CNN is 49%, SSAE is 42%, HPO is 25%, and the time complexity performance of the AutoE-SOM is 17 %. In the deployed methodology, the time complexity is much lower than the previous methodology. It can classify the dataset faster, crucial for real-time or large-scale applications. This efficiency can lead to quicker analysis and decision-making in medical contexts, potentially improving patient care and outcomes.

V. Conclusion

Our research novel deployed an AutoE-SOM method for the CVD dataset classification. Similarly, this paper studies the CVD dataset to evaluate its narrative nature and the comprehensiveness of the normalization method. We used the normalization method of castoff to normalize a dataset's attributes to a certain predefined standard. This allows you to remove unnecessary or noisy material and use valid and reliable data. The preprocessed dataset then uses the SVM algorithm to choose features. The AutoE-SOM system is created to group patients based on symptoms and medical information to identify specific subgroups that may require specialized treatment. The methods used are valuable for clinical diagnosis as they can analyze large and complex datasets and uncover connections and patterns that are challenging to identify using traditional statistical methods. Additionally, the proposed method achieved strong results on the CVD dataset, with a precision of 94.32%, recall of 94%, accuracy of 96%, F1 score of 92%, and time complexity of 17%. The potential of this groundbreaking approach is to help healthcare providers in identifying and intervening early, which can ultimately result in better patient outcomes and lower healthcare expenses.