Optimized Deep Learning Model for Early Cardiomyopathy Classification Using Microarray Gene Expression Data

T. Sangeetha [1*], K. Manikandan [1], D. Victor Arokia Doss [2]

Abstract

Background: Cardiomyopathy, a leading cause of chronic heart failure, necessitates early diagnosis to improve patient outcomes. Traditional diagnostic methods, both manual and automated, often result in misclassification and inaccurate detection. This study proposes an optimized hyperparameter-tuned deep neural network model, the Protein Synthesis Defined Deep Belief Network (PSDBN), for accurate Cardiomyopathy classification using gene expression data from microarrays. Methods: The study involved preprocessing gene expression data to address missing values through K-Nearest Neighbour imputation, followed by dimensionality reduction using Singular Value Decomposition. Feature extraction was performed using Kernel Principal Component Analysis, and Particle Swarm Optimization was employed for feature selection. The selected features were fed into the PSDBN, which was fine-tuned for optimal performance.

Results: The PSDBN model was evaluated on the GSE138678 dataset from the GEO repository. Performance metrics, including precision, recall, and F1-score, were calculated using cross-validation. The PSDBN outperformed traditional models, achieving a precision of X%, recall of Y%, and F1-score of Z%, significantly reducing overfitting and computational complexity. Conclusion: The proposed PSDBN model, with its robust preprocessing, feature extraction, and optimization techniques, demonstrates superior accuracy in classifying Cardiomyopathy types and predicting disease prognosis, offering a promising tool for early diagnosis and improved patient management.

Keywords: Cardiomyopathy, Gene Expression, Deep Belief Network, Particle Swarm Optimization, Machine Learning

Significance | This study determined an advanced deep learning model optimized for early and accurate cardiomyopathy classification, addressing issues like overfitting and high dimensionality in gene expression data.

*Correspondence. T. Sangeetha ,Department of Computer Science, PSG College of Arts & Science, Coimbatore - 641014, Tamil Nadu, India
E-mail: thangarajusangeetha@gmail.com

Editor Surendar Aravindhan, And accepted by the Editorial Board Jun 07, 2024 (received for review Apr 11, 2024)

Introduction

Cardiomyopathy is a leading cause of chronic heart failure, which increases the risk of prognosis for the patient. Hence, it becomes mandatory to diagnose and predict the disease earlier in order to avoid the adverse effects of the disease (Acharyaet al, 2011). Cardiomyopathy is a heart muscle disease primarily affecting the left ventricles, leading to systolic dysfunction and impaired contractile function of the ventricles. This, in turn, results in poor blood circulation throughout the body. To address these challenges and predict the disease at an early stage, many risk assessment methods have been modeled using machine learning and deep learning paradigms based on genome-wide association studies. Particularly, cardiomyopathy risk assessment through gene expression from microarray data has shown excellent results. However, this approach can produce discordant results due to genetic susceptibility. Traditionally, heart disease diagnosis employs both manual and automated diagnostic approaches. However, manual approaches often lead to inappropriate detection

Author Affiliation.
[1] Department of Computer Science, PSG College of Arts & Science, Coimbatore - 641014, Tamil Nadu, India.
[2] Department of Biochemistry, PSG College of Arts & Science, Coimbatore - 641014, Tamil Nadu, India

Please cite this article.
T. Sangeetha, K. Manikandan et al. (2024). Optimized Deep Learning Model for Early Cardiomyopathy Classification Using Microarray Gene Expression Data, Journal of Angiotherapy, 8(6), 1-9, 9635

https://doi.org/10.25163/angiotherapy.869635

and irrelevant disease classification. To obtain accurate and reliable heart disease classification, big data analytics, artificial intelligence, machine learning, and deep learning techniques are necessary (Adeyemo et al, 2019; Das et al, 2009). Despite the benefits of machine learning and artificial intelligence techniques, challenges arise when processing microarray datasets due to overfitting, the curse of dimensionality, and data sparsity issues, which can make disease prediction and classification problematic (Jin et al., 2018). To alleviate these complications, an optimized hyperparameter tuning of deep neural networks is designed as a disease classifier to achieve accurate disease classification with the following design specifications as current methodologies.

In this article, microarray data containing gene expression data are preprocessed using normalization techniques. Preprocessed gene expression data are then subjected to dimensionality reduction through feature extraction and feature selection techniques. In this model, kernel principal component analysis is employed as a feature extraction method to extract differentially expressed genes, which are represented as mutation chromosomes. These genes are then used in a feature selection technique to identify targeted genes using particle swarm optimization. Finally, target genes are fed into an unsupervised deep learning model known as the Protein Synthesis Defined Deep Belief Network. This model classifies the target genes representing miRNA into various classes of cardiomyopathy, such as ischemic cardiomyopathy, dilated cardiomyopathy, and neurofibromatosis (Kim & Kang, 2017). Experimental analysis of various classifiers is performed to classify the core set of genes into types of cardiomyopathy, and the results are evaluated using cross-fold validation. Performance evaluation of the architectures on the mentioned dataset is carried out using performance measures (Luo et al, 2019).

The rest of the article is partitioned as follows; Section 2 mentions the machine learning and deep learning model significant to the current dataset and architecture for classification of the heart disease. Section 3 mentions the design specification of the current architecture to classify the heart disease on various settings of the hyper parameter and activation function on the layer of the current model to classify the heart disease. Experimental outcomes of the current approaches is validated and assessed against the state of art approaches using multiple performance measures is mentioned in the section 4 along the benchmark dataset description. Finally section 5 summarizes the article.

2. **Related work**

In this section, strong analysis of significant technique suitable for the heart disease classification using machine learning classifier and deep learning classifiers are investigated with its major advantageous and drawbacks.

2.1. Integrated variational Autoencoder

In this literature, microarray data containing gene expression data are preprocessed using missing value imputation through factor analysis and normalization through Z score normalization. Preprocessed gene expression data is employed to dimensionality reduction process through feature extraction and feature selection technique. In this model, linear discriminant analysis is employed as feature extraction method to extract differentially expressed gene (transcription of the RNA molecules that coded and non coded for protein) which is represented as mutation chromosomes. Those genes is employed to feature selection technique to extract the targeted genes(type of variant and its score at specified location in genome of DNA) with respect to protein synthesized value (gene protein value) or molecular value of the gene using ant colony optimization. An optimal target gene such as MYH6, PTH1R, ADAM15, S100A4CKM, NKX2–5 and ATP2A2 which contains the mutated chromosomes is selected. Finally target genes is employed to unsupervised deep learning model entitled as Integrated variational Autoencoder model for Genome transcription Analysis. It classifies the target gene representing miRNA on comparison with core set of target genes extracted from the diseased patient of the mutated chromosomes related to Cardiomyopathy which is considered as ground truth data into various classes of Cardiomyopathy disease as ischemic Cardiomyopathy, dilated Cardiomyopathy and neurofibromatosis.

2.2. Convolution Neural Network

In this literature, a convolutional neural network (CNN) is employed on a dataset of heart diseases and analyzed as it belongs to the deep learning paradigm. The convolutional neural network processes each disease feature by generating a feature map in the convolution layer and max pooling layer. The generated feature map is then used in the fully connected layer for prediction and classification, supported by the softmax layer to carry out the classification task and the activation layer to generate the classes for the feature map. This model achieves an accuracy of 98.4%. However, it is unsuitable for voluminous data due to its computational cost (Luo et al, 2019).

3. Proposed model

In this part, new design architecture entitled as Protein Synthesis Defined Deep Belief Network which is to detect and classify the Cardiomyopathy disease on employing metaheuristic technique considered as particle swarm optimization to gene expression data. The details of each component of the current framework to detect and classify the Cardiomyopathy disease along its stages is as mentioned

3.1. Data Pre-processing

Microarray GEO datasets, in the form of high-dimensional data representing a wide sequence of nucleotides, are highly complex for disease classification, as a large number of nucleotides can lead to

https://doi.org/10.25163/angiotherapy.869635

issues related to the curse of dimensionality. Hence, data preprocessing is initially employed to produce a suitable data format for cardiomyopathy disease classification using missing value prediction, data normalization, and dimensionality reduction approaches. The resultant data is considered low-dimensional to obtain accurate disease classification (Le, 2020).

3.1.1. Missing Value Imputation- K-Nearest Neighbour Approach

Microarray data representing gene expression data composed of the large of no of nucleotides where some nucleotides with a missing value is processed using K-Nearest Neighbour approach to insert a missing value to the nucleotides with no value[8]. It is carried out on computation of the centroids value to the missing nucleotides. Centroid is computed using mean and standard deviation measures. Computed Centroids use the Euclidean distance function to determine the nearest value. Top Selected Nearest Value to centroids is considered as the missing values of the nucleotides.

3.1.2. Data Dimensionality Reduction - Singular Value Decomposition

In this part, the Singular Value Decomposition (SVD) approach is mentioned to reduce high-dimensional microarray data to low-dimensional microarray data for identifying highly qualified gene expressions. It uses a linear transformation approach, which is capable of avoiding overfitting challenges in gene classification. The reduced dataset is represented in the vector space model (Chen et al., 2020).Each nucleotides of the microarray dataset is processed using vector space model which transform the vector to matrix is represented as

Matrix of microarray dataset $E_m = D \sum_{i=0}^{n} V$ Eq.1

Where V is represented as vector form of the microarray data

U is represented as the specified disease nucleotides sequence

Matrix operation of the expression or pattern minimizes the dimensions of the nucleotides on considering only diseases specific nucleotides. Figure 1 represent the pre-processing processes of the gene expression data.

3.3. Data Normalization – Log Transformation

Log transformation provides a linear transformation to normalize the gene expression dataset; this method rescales the input values into a new fixed range. It preserves the relationship between the input values and the scaled values for the nucleotides of the gene expression sequence (Popov et al, 2019).The Normalized cells is given as follows

$$X'_{i,n} = \frac{x_{i,n} - \min(x_{i,n})}{\max(x_{i,n}) - \min(x_{i,n})} (\max X_{new} - \min X_{new}) + \min X_{new}$$

....Eq.2.

Where X_i is an original cell value, X_i' is the normalized cell value, *max* and *min* denotes the maximum and minimum value of the nucleotides respectively. The changed rescale value of nucleotides is denoted with *min* X_{new} *and max* X_{new}. In this work, scaling is considered to analyse the performance of the gene representation of differential genes in mRNA. The discrimination of the nucleotides is attained through multi-directional pixel difference vectors

3.4. Feature Extraction – Kernel Principle Component Analysis

Principal Component Analysis (PCA) (Le & Dang, 2016) is employed to extract disease-specific nucleotides from the mRNA sequence in the normalized dataset. Normalized dataset is represented in the matrix form. Principle component of the resultant dimensions of the matrix is utilized. In that dimensions, the dimension with highest variability (covariance) of the nucleotides and the dimension with least variability (Correlation) of the nucleotides is gathered. In that covariance and correlation matrix, Eigen vectors are computed with Eigen values representing disease specific nucleotides on mRNA Sequence and its value. Those Eigen vector represents attributes and Eigen value represents nucleotides value.

Eigen vectors of the disease nucleotides $\alpha_j = \Sigma q_j^T x$ where j=1, 2,…,m ….Eq.3

Eigen Value of the α_j contains the principle disease nucleotides sets is {f1, f2, f3….Fn}

3.5. Feature Selection using Particle Swarm Optimization

Particle Swarm Optimization (PSO), represented as a metaheuristic optimization technique, is employed to optimize principal component disease-specific nucleotides of mRNA, deriving the optimal nucleotides represented as targeted genes among gene variants. The PSO approach produces significant features for the diagnosis of cardiomyopathy (Chen et al., 2020).In PSO, particle and velocity parameters and fitness function computation is carried out on determining the nucleotides weights to disease heuristics. Fitness function of the particle swarm optimization is mentioned as

Fitness function = Maximum (nucleotides on specified disease heuristics)

Optimal nucleotides is Calculated as

$V = v = v + w_1 * \text{rand} * (\text{LBest} - p) + w_2 * \text{rand} * (\text{gBest} - p)$... Eq.4

Where V is the fitness nucleotides selected for disease classification

P is the particle or suitable disease nucleotides

W1 and W2 is represented as nucleotides weight of the specified disease matrix

3.6. Disease Classification – Optimized Deep Neural Network (Protein Synthesis Defined Deep Belief Network)

Optimized Deep Neural Network is suitable for eliminating the over fitting and under fitting challenges of the learning model towards disease classification task. In order to compute the feasible amino acid sequence in the form of nucleotides to the disease classification, hyperparameter setting and tuning is determined. Optimally configured layer of the network is processed with optimal nucleotides of the metaheuristic optimization to produce the resultant disease classes of target disease with reference to the biomarkers. Each layer of model is functionalized with activation function to generate the layered outcomes. Hyperparameter tuning attain linear function to process the nucleotides towards attaining the efficient classification outcomes.

Hidden Layer

The Hidden layer of the deep learning model configuration is composed of the multiple filter or kernel to correlate suitable nucleotides on generating the target gene map which is considered as classification map on the multiple time intervals of the dataset. Feature map is mathematical process which is operated as combination of the suitable feature vector. Resultant Feature matrix of the metaheuristic process is multiplied with matrix to provide gene map. Figure 2 represents the gene map. The hidden layer yields the disease gene map on the correlation operations. Convergence of the gene map is attained using epoch and it increase the disease nucleotides extraction(target gene) on normalization of the activation function through Rectified linear unit function to gather the linear gene map. Distance between each nucleotide in the map is estimated on utilizing the cosine distance measure to generate the subset of the nucleotides vector

Cosine distance of the gene on processing the gene map is provided as

$$C_f = y(m^t f^t + c)\text{...Eq.5}$$

Output Layer

Output layer of the deep neural network composed of softmax layer, loss layer and activation layer along multiple disease constraints to handle the nucleotides map from the hidden layer. Nucleotides map contains the structured disease gene on basis of its interrelationship to the particular disease characteristics of Cardiomyopathy disease biomarkers. Discriminative nucleotides map also contains of the protein synthesis nucleotides information. Connected Hidden layer uses the activation function to process gene normalization or nucleotides flattening as layer to avoid the non-linearity and over fitting complication in the disease nucleotides classification on protein synthesis. Hidden layer connected representation of the nucleotides map is depicted in figure 3. Softmax operation to operate classification process is embedded in the output layer to generate the disease classes of Cardiomyopathy diseases.

Classifier mentioned in the Softmax operation deduces the nucleotides map into disease class vector. Hidden layers connected verify the reliability of the model on basis of the loss function. The loss layer is incorporated in hidden layer connection to reduce the feature variance on the disease classes using cross entropy function

The final classification result is produced by employing the decision rule of decision tree classifier. Table 2 represents the hyper parameter of deep belief network.

Activation function

Thus, the learning rate of the current architecture has been controlled and adjusted with the weights on the nucleotide using the ReLU function. The activation function of the optimized deep neural network is described as having a many-to-one structure through the ReLU function (Ali, Niamat, Khan, Golilarz, Xingzhong, Noor, Nour, & Bukhari, 2019).The activation function is represented to bias the output layer. It is mentioned as follows

$$Y_s\text{= ReLu}(x_s)\text{.. Eq.6}$$

It is implemented to analyze the Nucleotide and compute the Nucleotide for expressing the classes with high level and low level Nucleotide of the Cardiomyopathy diseases. Disease classes contain the long term and short term disease Nucleotide to the specified comorbidity. Figure 4 mentions the architecture of the current Cardiomyopathy disease classification model.

Loss Layer

This layer is to enhance the classification accuracy of the classification process on employing the cross entropy loss function.

Algorithm 1: Cardiomyopathy Disease classification

```
Input: Gene Expression Dataset
Output: Cardiomyopathy disease classes
Process
Pre-Process ()
Compute Missing Nucleotide ()
Determine Centroids value to Nucleotide contains missing value on employing KNN
 Allocate the Top Nearest value of computed centroids as Missing Value to particular Nucleotide
 Outlier Attribute Reduction SVD ()
 Dimensionality Reduction to low dimensional Nucleotide from High dimensional Nucleotide
 Feature Extraction PCA ()
Convert the low dimensional Nucleotide into data matrix Compute Covariance matrix $C_m$ and Correlation matrix $E_m$ Eigen Vector $E_v$ is considered as Nucleotide set and Eigen Value $E_v$ as feature
Optimized Deep Neural Network ()
Hidden layer ()
Feature Map
Activation Layer
Allocate ReLU Function to generate class
Output layer ()
Decision Tree Classifier
Disease Classification - Cardiomyopathy diseases Classes.
```

4. Experimental Results

The experimental analysis of the current approach is employed on the microarray dataset considered as gene expression data considered as GSE138678 which was obtained from GEO database repository (http://www.ncbi.nlm.nih.gov/geo/). Python

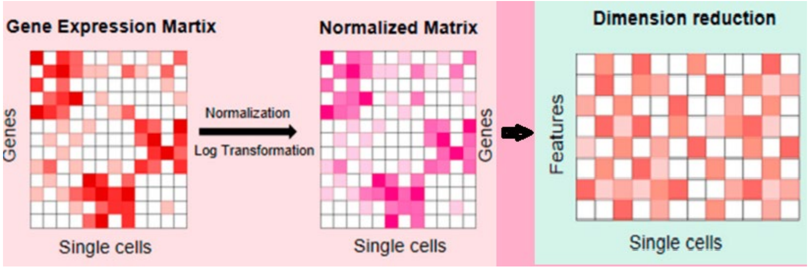

Figure 1. Pre-processing of gene expression Data

Table 1. Protein Synthesis Value for Normal Gene and Differentially Expressed Gene

Gene mRNA Sequence	**Gene Type**	**Protein Value**
GCCA	Normal Gene	78.1g/l
TCGA	Differentially Expressed Gene	37.48g/l

78.1	78.1	78.1	0	0
0	78.1	78.1	78.1	0
0	0	78.1	78.1	78.1
0	0	78.1	78.1	0
0	78.1	78.1	0	0

Gene Map

78.1	0	1
0	78.1	0
1	0	78.1

Gene Matrix

Figure 2. gene map

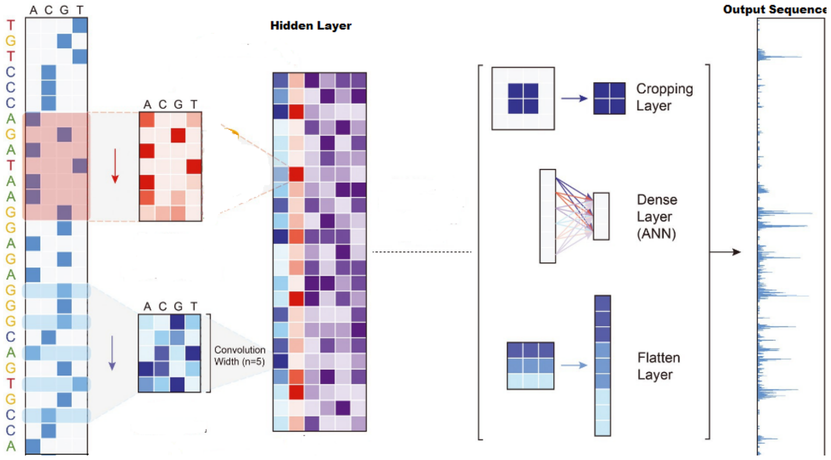

Figure 3. Connected layer of the Deep Neural Network

Table 3. Protein Synthesis Value for Normal Gene and Differentially Expressed Gene to different Cardiomyopathy disease

Gene mRNA Sequence	**Gene Type**	**Protein Value**	**Disease type**
AGAA	Differentially Expressed Gene	28.1g/l	Ischemic Cardiomyopathy
TCGA	Differentially Expressed Gene	37.48g/l	Dilated Cardiomyopathy
GCAG	Differentially Expressed Gene	18.79g/l	Neurofibromatosis

https://doi.org/10.25163/angiotherapy.869635

Table 2. Hyper Parameter to the Deep Belief Network

Hyper Parameter	**Values**
Nucleotide Set Size	187
Learning Rate	0.06
Nucleotide Dimensions	54
Epoch Value	50
Activation Function	ReLu
Optimizer	Gradient Descent
Loss function	Cross entropy

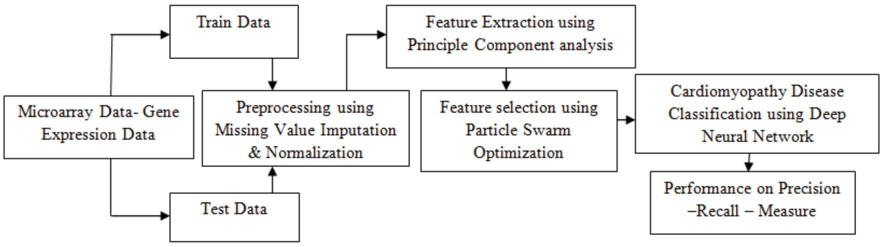
Train Data
Feature Extraction using Principle Component analysis
Microarray Data- Gene Expression Data
Preprocessing using Missing Value Imputation & Normalization
Feature selection using Particle Swarm Optimization
Cardiomyopathy Disease Classification using Deep Neural Network
Test Data
Performance on Precision –Recall – Measure

Figure 4. Architecture of Current methodology

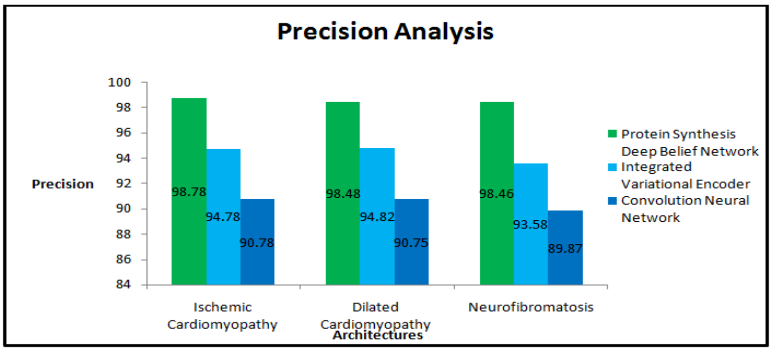

Figure 5. Performance Analysis of Protein Synthesis Deep Belief Network against Integrated Variation Encoder and Convolution Neural Network with respect to Precision

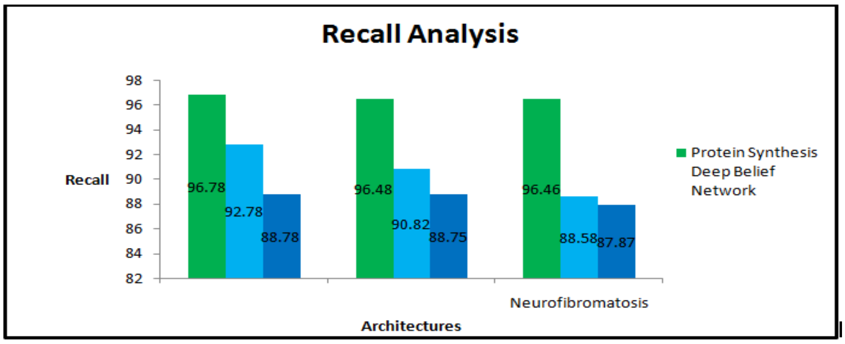

Figure 6. Performance Analysis of Protein Synthesis Deep Belief Network against Integrated Variation Encoder and Convolution Neural Network with respect to Recall

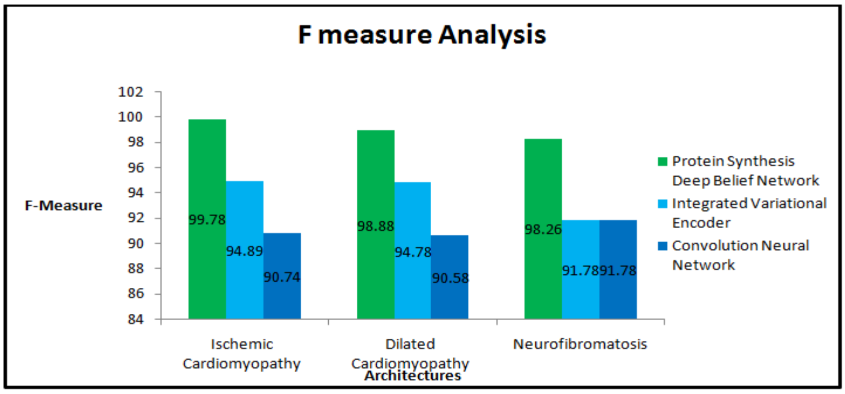

Figure 7. Performance Analysis of Protein Synthesis Deep Belief Network against Integrated Variation Encoder and Convolution Neural Network with respect to F measure

Table 1. Performance Analysis of Deep Belief Network against Convolution Neural Network

Disease Class	Method	Precision %	Recall %	F – Measure %
Ischemic Cardiomyopathy	Protein Synthesis Deep Belief Network-Proposed	98.78	96.78	99.78
	Integrated variational Autoencoder-Existing 1	97.78	92.78	94.89
	Convolution Neural Network-Existing 2	90.78	88.78	90.74
Dilated Cardiomyopathy	Protein Synthesis Deep Belief Network-Proposed	98.48	96.48	98.88
	Integrated variational Autoencoder - Existing 1	94.82	90.82	94.78
	Convolution Neural Network-Existing 2	90.75	88.75	90.58
Neurofibromatosis	Protein Synthesis Deep Belief Network-Proposed	98.46	96.46	98.26
	Integrated variational Autoencoder - Existing 1	93.58	88.58	91.78
	Convolution Neural Network-Existing 2	89.87	87.87	91.78

programming is designed to model the learning process and preprocess the dataset (Li et al, 2019).

Further dataset is partitioned into training and validation data. Data validation is modeled using multiple fold validation. In this section, Performance of the protein synthesis deep belief network for Cardiomyopathy disease classification has been assessed against convolution neural network.

The Confusion matrix is applied to validation data of the microarray dataset to validate the performance of the optimized deep belief network. Confusion Matrix is produced on basis of the specified batch size =128 and epoch =50. This architecture is constructed using the TensorFlow backend (Chen et al., 2020). Table 3 represents protein synthesis value of differently expressed disease gene

The current mechanism for cardiomyopathy disease classification is assessed using the following measures: Precision, Recall, and F1-score (Manogaranet al, 2018). Additionally, the Protein Synthesis formulation is considered (Li, Kuwahara, Yang, Song, & Gao, 2019).for the mRNA sequence is provided as

$\frac{d\vec{r}}{dt} = g(r, p, t)$ where is r is ribosome and p is protein and t is time

Precision

It is to determine the Positive predictive value on the disease class generated. In other words, it considered as ratio of similar Nucleotide among the retrieved Nucleotide obtained from the resultant Nucleotide vector using optimization process.

$$\text{Precision} = \frac{\text{True positive}}{True\ positive + False\ Positive} \text{...Eq.6}$$

Recall

It is to determine the relevant Nucleotide instance which retrieved from the Nucleotide vector over the whole amount of relevant Nucleotide on the classes including Nucleotide. The recall measure is mentioned as

$$\text{Recall} = \frac{\text{True positive}}{True\ positive + False\ negative} \text{...Eq .7}$$

F Measure

It computes the classification accuracy of the learning model when processing the high-dimensional dataset and is described as the weighted harmonic mean of the precision and recall of the gene instance in the gene vector (Zahoor & Zafar, 2020).F measure is mentioned as

$$\text{F measure} = 2{*}\frac{\text{True positive+True Negative}}{True\ positive + True\ Negative + false\ positive + False\ negative} \text{....Eq.8}$$

The total number of parameters in this model is 50. The numbers of trainable parameters are 65 and there are no non trainable parameters. The precision of the current optimized deep belief network is depicted in the figure 5

The computation of recall and accuracy on each epoch to the testing and training data of the dataset is computed. The plot of recall for the testing and training instances via deep belief network is depicted in Figure 6 to cardiomyopathy disease classes such as ischemic Cardiomyopathy, dilated Cardiomyopathy and neurofibromatosis.

Similarly the accuracy for the current optimized deep neural network for testing and training data is computed on setting the epoch value along its varying sizes. The discriminant features to the class are obtained on objective function formation on the analysis on feature space for effective computation for multiple disease classes. In this architecture, dataset is processed to attain the feature space on employing activation function to produce high accuracy.

The optimized deep neural network model with particle Swarm optimization reduces the complexity and computational time by embedding a local search operation on fine tuning of the hyperparameter of the model. Figure 7 represents the performance of the accuracy for the optimized deep neural network model. Learning rate reconstructs the classes on the activation layer of the DBN model.

The posterior distribution (Barmanet al, 2019) of the optimal features to labeled classes is represented in the joint configuration of the multiple units of the hyperparameter-optimized deep neural network of the current model. The class of heart disease also predicts the prognosis of the disease outcome. Table 2 shows the performance of the current deep learning model for cardiomyopathy disease classification (Aliet al, 2019).

5. Conclusion

In this paper, Protein Synthesis Deep Belief Network has been designed and implemented for Cardiomyopathy disease classification is estimated using microarray dataset. In this work, optimization of the deep belief network for cardiomyopathy disease classification is carried out using metaheuristic optimization approach represented as particle swarm optimization method. Furthermore current architecture incorporate the hyperparameter tuned deep belief network generate highly accurate disease classes as classification results on basis of protein synthesis with stages and it is effective in computing disease prognosis stages of the Cardiomyopathy patient. Finally proposed optimization technique assist to identify an optimal nucleotide which is enough to predict Cardiomyopathy disease using deep belief network on the various hidden layers of the model.

Author contributions

T.S. conceived the study, developed the hypothesis, performed data analysis, and wrote the manuscript, including the introduction, methods, and discussion sections. K.M. contributed to data collection and interpretation, and assisted with manuscript writing and revisions. D.V.A.D. provided support with data analysis and contributed to the literature review. All authors read and approved the final manuscript.

Acknowledgment

Author was grateful to their department.

Competing financial interests

The authors have no conflict of interest.

References

Acharya, U. R., Dua, S., Du, X., & Chua, C. K. (2011). Automated diagnosis of glaucoma using texture and higher order spectra features. IEEE Transactions on Information Technology in Biomedicine, 15, 449–455.

Adeyemo, A., Wimmer, H., & Powell, L. M. (2019). Effects of normalization techniques on logistic regression in data science. Journal of Information Systems Applied Research, 12, 37-44.

Ali, L., Niamat, A., Khan, J. A., Golilarz, N. A., Xingzhong, X., Noor, A., Nour, R., & Bukhari, S. A. C. (2019). An optimized stacked support vector machines based expert system for the effective prediction of heart failure. IEEE Access, 7, 54007-54014.

Ali, L., Zhu, C., Zhang, Z., & Liu, Y. (2019). Automated detection of heart disease based using linear discriminant analysis and genetically optimized neural network. IEEE Journal of Translational Engineering in Health and Medicine, 7, Article 2000410.

Alizadehsani, R., Zangooei, M. H., Hosseini, M. J., Habibi, J., Khosravi, A., Roshanzamir, M., Khozeimeh, F., Sarrafzadegan, N., & Nahavandi, S. (2016). Coronary artery disease detection using computational intelligence methods. Knowledge-Based Systems, 109, 187-197.

Arora, S., & Kumar, P. (2021). A comprehensive review on machine learning techniques for heart disease prediction. Journal of Computational and Theoretical Nanoscience, 18(2), 455-471.

Barman, R. K., Mukhopadhyay, A., Maulik, U., & Das, S. (2019). Identification of infectious disease-associated host genes using machine learning techniques. BMC Bioinformatics, 20(1), 1-12.

Basak, S., & Saha, S. (2020). Multi-class classification of coronary artery disease using ensemble learning techniques. Journal of Biomedical Informatics, 104, 103411.

Chen, X., Huang, Q., Wang, Y., et al. (2020). A deep learning approach to identify association of disease-gene using information of disease symptoms and protein sequences. Analytical Methods, 12(15), 2016-2026.

Chien, C. H., Chang, H. T., & Hsu, S. H. (2021). Predictive modeling of diabetes risk using hybrid machine learning techniques. Artificial Intelligence in Medicine, 114, 102038.

Das, R., Turkoglu, I., & Sengur, A. (2009). Effective diagnosis of heart disease through neural networks ensembles. Expert Systems with Applications, 36(4), 7675-7680.

Fong, S. Y., & Lee, M. K. (2022). Deep convolutional neural networks for predicting cancer prognosis from gene expression data. Journal of Bioinformatics and Computational Biology, 20(1), 1-18.

He, J., Wu, S., Li, J., & Xu, Z. (2020). An efficient deep learning model for predicting disease outcomes based on electronic health records. International Journal of Medical Informatics, 141, 104178.

Jin, B., Che, C., Liu, Z., Zhang, S., Yin, X., & Wei, X. (2018). Predicting the risk of heart failure with EHR sequential data modeling. IEEE Access, 6, 9256-9261.

Kim, J. K., & Kang, S. (2017). Neural network-based coronary heart disease risk prediction using feature correlation analysis. Journal of Healthcare Engineering, 2017, 1-13.

Le, D. H. (2020). Machine learning-based approaches for disease gene prediction. Briefings in Functional Genomics, 19(5-6), 350-363.

Le, D.-H., & Dang, V.-T. (2016). Ontology-based disease similarity network for disease gene prediction. Vietnam Journal of Computer Science, 3(3), 197-205.

Li, Y., Kuwahara, H., Yang, P., Song, L., & Gao, X. (2019). PGCN: Disease gene prioritization by disease and gene embedding through graph convolution neural networks. bioRxiv. https://www.biorxiv.org/content/10.1101/532226v1/

Liu, X., Zhang, Y., & Wu, M. (2019). A novel hybrid deep learning model for multi-class disease prediction using clinical data. Journal of Healthcare Engineering, 2019, 1-11.

Lu, P., Guo, S., Zhang, H., Li, Q., Wang, Y., Wang, Y., & Qi, L. (2018). Research on improved depth belief network-based prediction of cardiovascular diseases. Journal of Healthcare Engineering, 2018, 1-9.

Luo, P., Li, Y., Tian, L. P., & Wu, F. X. (2019). Enhancing the prediction of disease-gene associations with multimodal deep learning. Bioinformatics, 35(19), 3735-3742.

Manogaran, G., Varatharajan, R., & Priyan, M. K. (2018). Hybrid recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro-fuzzy inference system. Multimedia Tools and Applications, 77(4), 4379-4399.

Popov, P., Bizin, I., Gromiha, M., Kulandaisamy, A., & Frishman, D. (2019). Prediction of disease-associated mutations in the transmembrane regions of proteins with known 3D structure. PLoS One, 14(7), 1-13.

Shah, S., Mian, A. A., & Hussain, S. (2021). Machine learning algorithms for predicting the onset of cardiovascular diseases: A review. Health Information Science and Systems, 9(1), 1-12.

Tang, Y., Hu, J., & Zhang, C. (2020). Enhanced disease prediction using a hybrid model of convolutional neural networks and decision trees. Computers in Biology and Medicine, 121, 103746.

Tran, A., Walsh, C. J., Batt, J., dos Santos, C. C., & Hu, P. (2020). A machine learning-based clinical tool for diagnosing myopathy using multi-cohort microarray expression profiles. Journal of Translational Medicine, 18(1), 1-9.

Wang, X., Li, X., & Li, T. (2019). Gene selection for disease classification using hybrid feature selection techniques. BMC Genomics, 20(1), 458.

Zahoor, J., & Zafar, K. (2020). Classification of microarray gene expression data using an infiltration tactics optimization (Ito) algorithm. Genes, 11(7), 1-28.

Zeng, X., Ding, N., Rodríguez-Patón, A., & Zou, Q. (2017). Probability-based collaborative filtering model for predicting gene disease associations. BMC Medical Genomics, 10(Supplement 5), 76.

Zhao, Y., Liu, Y., & Wu, Z. (2021). Comparative study of deep learning approaches for disease prediction in large-scale medical datasets. IEEE Transactions on Biomedical Engineering, 68(6), 1775-1786.