Journal of Intelligent Systems and Internet of Things
Vol. 18, No. 01, PP. 361-373, 2026
DOI: https://doi.org/10.54216/JISIoT.180128

A Deep Convolutional Autoencoder with Metaheuristic Optimization based Feature Reduction Framework for Genetic Disorder Detection Model

S. Puvaneswari[1,*], G. Indırani[2]
[1]Research Scholar, Department of Computer Science and Engineering, FEAT, Annamalai University, Chidambaram, India
[2]Associate Professor, Department of CSE, Government College of Engineering, Sengipatti, Thanjavur, India
Emails: bhuvanakce2021@gmail.com; induk0992@gmail.com

Abstract

Genetic disorder is an outcome of transformation in deoxyribonucleic acid (DNA) system, which is progressed or natural from blood relation. Such transformations might lead to deadly illnesses like Alzheimer's, cancer, and much more. The disorder of single gene kind is affected by a change in a solitary gene in DNA. The chromosomal disorder kind is affected when a genetic material or a portion of chromosome is removed or substituted in the structure of DNA. Complex illnesses are caused by the alteration in over one gene exhibit in the DNA. In recent times, the usage of artificial intelligence (AI)-based deep learning (DL) systems has exposed excellent achievement in the prognosis and prediction of diverse illnesses. The latent of DL models are employed to forecast genetic disorder at an initial phase utilizing the genome data for appropriate treatment. This paper presents a Deep Feature Selection Framework for Genetic Disorder Detection Using Convolutional Autoencoder and Metaheuristic Optimization (DFSFGDD-CAEMO) model. The aim of DFSFGDD-CAEMO model is to develop an accurate DNA-based genetic disorder classification model using advanced techniques for early and reliable disease diagnosis. Initially, the min-max normalization method is employed in the data pre-processing stage for converting an input data into a beneficial format. Besides, the Aquila optimizer (AO) method has been deployed for the selection of feature process in order to select the most significant features from a dataset. For the classification procedure, the proposed DFSFGDD-CAEMO technique designs Convolutional Autoencoder (CAE) method. At last, the hyperactive parameter tuning process is performed through enhanced pelican optimization algorithm (EPOA) for improving the classification performance of CAE model. The experimental evaluation of the DFSFGDD-CAEMO technique occurs using benchmark dataset. The experimentation results indicated out the enhanced performance of the DFSFGDD-CAEMO system when equated to existing approaches.

Received: March 12, 2025 Revised: June 04, 2025 Accepted: July 16, 2025

Keywords: Genetic Disorder Detection; Convolutional Autoencoder; Enhanced Pelican Optimization Algorithm; Feature Selection; Deoxyribonucleic Acid

1. Introduction

The genetic disorder was triggered by an alteration in the genome or by a mutation in the gene structure. Since the genome holds the data, the variation in the genome leads to a mutation in the function or structure of an organ [1]. The genes are composed of deoxyribonucleic acid (DNA), and any variation in DNA arrangement leads to a genetic disorder. The genome information includes significant data and medical indicators, which are utilized to examine the genetic conditions that lead to illnesses. A specific division of genomics, bioinformatics, concentrates on the analysis of genomes [2]. There are numerous genetic illnesses: single-gene inheritance disorders, chromosomal disorders, complex disorders, and multi-factorial genetic inheritance disorders, and these are

analysed depending on the DNA structure [3]. The single-gene condition is triggered by an alteration in a one gene in the DNA. The chromosomal condition is affected when a DNA or a segment of chromosomes was removed or exchanged in the structure of DNA. Complex conditions are triggered by the alteration in several genes present in the DNA [4].

Genetic disorders can also be multi-factorial; for example, complex conditions that arise from the combined effects of genetic defects, environmental factors, and lifestyle, with genes contributing only partly to the phenotypes connected to these disorders [5]. A single-gene condition results from a mutation in only one gene. As this can occur in any genetic factor, single-gene conditions can affect overall functioning and are extremely different. Although they differ clinically, all single-gene conditions share the same organic basis, can be passed to the next generation, and require similar essential genetic and counselling support [6]. Early and precise diagnoses of genetic disorders remain a continuing issue in the medical field. However, substantial advancements have been made in identifying particular conditions, but the classification and estimation of disease across the range of genetic inheritance types have remained indeterminate [7]. The capability for systematically discerning genetic abnormalities in the earliest stages of life has deep clinical consequences. Initial identification assists quick intervention and enhances prognosis and life quality for affected persons [8]. Therefore, there is a crucial necessity for advanced yet accessible methods to define the wide varieties of genetic disorders and identify specific subtypes. Machine learning (ML) and deep learning (DL) are subdivisions of artificial intelligence (AI), widely employed in numerous genomics research [9]. DL is a development of ML, currently receiving positive attention for classifying and predicting genome problems. At present, DL has attained remarkable attention in progressing genetic research because of its ability to discern multi-dimensional interactions without assumptions [10].

2. Related Works

Pimpalkar et al. [11] introduced a new method, the effective Deep Convolutional Neural Networks (DCNN) method, to identify and classify DNA structures in genomics studies. Using the CNN's hierarchical learning abilities, this method automatically extracts complex features from raw DNA sequences, extracting global and local patterns vital for genomic understanding. A new k-mer embedding model converts raw DNA sequences into a numerical representation appropriate for CNN processing, maintaining sequence order and elevating data with contextual information to improve performance. Gala et al. [12] provided a novel machine-learned classification method for the numerous kinds of soft tissue that relied on genomics data, which solves a significant gap in sarcoma diagnostics. The earlier research has examined several traits of sarcoma; however, this research is innovative, which aims to predict sarcoma kinds utilizing genetic data. Random Forest (RF) was utilized as the meta-estimator, and a stacking ensemble approach encompassing RF, Light GBM, and Extreme Gradient Boosting was employed for this investigation. Das et al. [13] employed an efficient framework depending on the DL algorithm. At first, the information is gathered from 5 gene cancer datasets, which are afterward enhanced to increase the data size. Min-Max Normalization and then an Enhanced Chimp Optimizer (ECO) technique are used for selecting the most important genes although removing unwanted or redundant ones.

Zhang et al. [14] designed a new technique, named DRBPPred-GAT, to predict DBP and RBP on a graph multi-head attention network. This technique comprises 3 main steps: At first, Protein information is completely extracted by fusing 8 kinds of features. Following that, an Autoencoder (AE) is leveraged for removing unrelated features. At last, a GCNN alongside multi-head attention was employed for predicting DBP and RBP. In [15], the Enhanced Whale Optimizer (EWO) methodology and the Improved CNN (ICNN) model have been presented to predict ASD efficiently. This presented study includes pre-processing, feature selection (FS), and classification. EWO is deployed for identifying the most pertinent autistic dataset features for the FS process. Utilizing the EWO algorithm's objective function, choosing the best fitness features is one way to maximize classification accuracy.

In [16], a new DNA encoding system has been suggested, solutions were aimed at problems stated, and DNA enhancers were estimated alongside Bi-LSTM. This work comprises 4 diverse phases for 2 scenarios. Initially, DNA enhancer data was attained. Then, DNA sequences were transformed to numerical representation using the suggested encoded system and several DNA encoded methods, namely EIIP, atomic and integer number. Following that, the Bi-LSTM approach has been developed, and the data was classified. Thakur et al. [17] proposed a hybrid technique that depends upon CNN and Recurrent Neural Network (RNN) for predicting various kinds of cancer, including Lung, Kidney, Breast, Uterine, colon, and Prostate cancer from gene expression information. The bottleneck features are captured through a sandwich-stacked approach, depending on the VGG16 and VGG19 pre-trained methods. After that, the presented hybrid classifier derived from RNN-CNN is leveraged for classifying the data into different classes.

3. Proposed Methods

This article develops a DFSFGDD-CAEMO model. The main of DFSFGDD-CAEMO model is to develop an accurate DNA-based genetic disorder classification model using advanced techniques for early and reliable disease diagnosis. To achieve that, the proposed DFSFGDD-CAEMO model involves various stages such as data pre-

processing, feature selection, classification, and parameter tuning. Fig. 1 depicts the complete working flow process of DFSFGDD-CAEMO model.

A. Data Pre-processing

Initially, the min-max normalization method is employed for transforming an input data into a beneficial format. To normalize data, min-max normalization scaling input values among 0 and 1 that aids to better performance of model and converge more consistently. This model linearly alters features from one range of values to another [18]. The variables are frequently modified to fall from 0 to 1, or -1 and 1. The linear transformation has been typically employed to achieve rescaling.

$$z = \frac{w - \min(w)}{\max(w) - \min(w)} \tag{1}$$

Here min and max refers to least and maximal values in W and w represents collection of observed values of w. Specifically, the range of w equals $\max(w) - \min(w)$. The advantage of this normalization model exists in its capability to maintain data relations precisely.

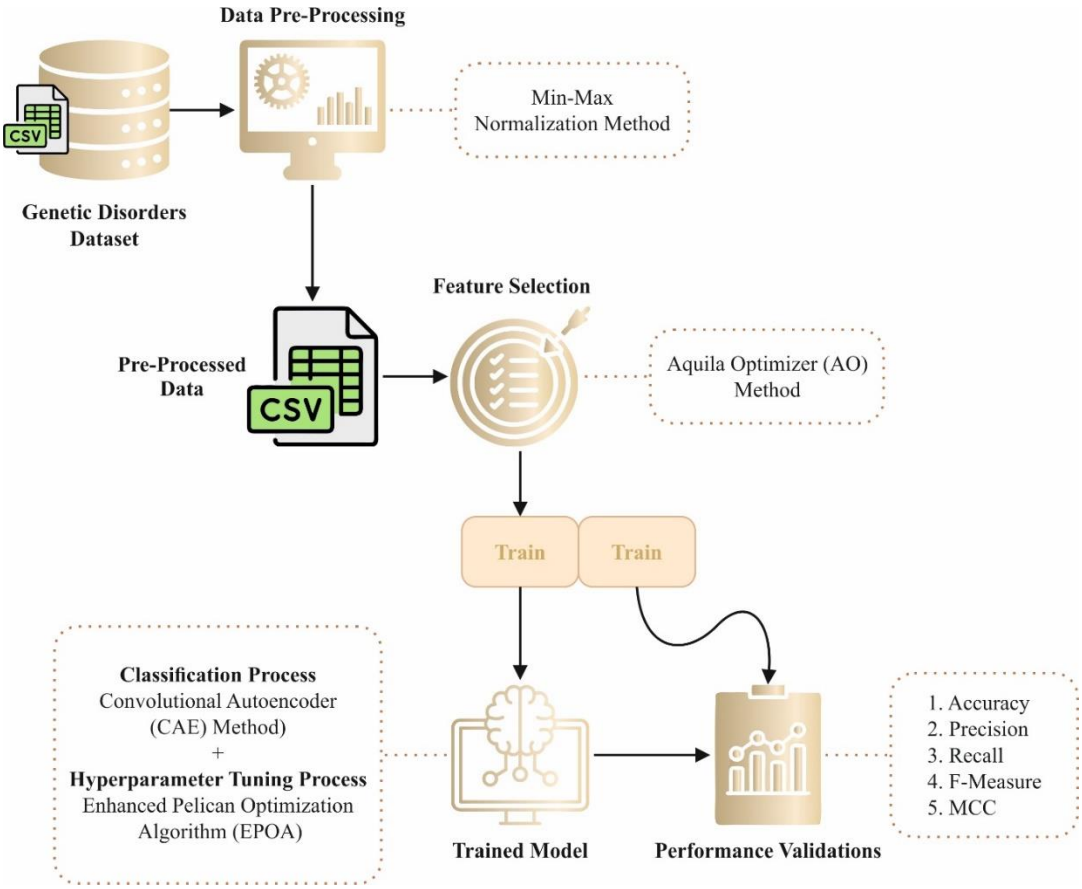

Figure 1. Overall Working Flow Process of DFSFGDD-CAEMO model

B. AO-based Feature Selection Process

Besides, the AO method has been deployed for the selection of feature process in order to select the most relevant features from a dataset. The AO is a population-based meta-heuristic model. The conventional AO follow 5 main stages [19]:

Stage 1: Initialization

In this step, a kind of proposed random solutions is given, and the AO parameters are initialized.

Stage 2: Expanded Exploration

This phase follows Aquila's major hunting model that consist of soaring to a greater height before implementing a vertical plunge. The model searches the area from the higher place, signified by mathematically in Eq. (2), while $X_1(t+1)$ is the following solution, and $X_{best}(t)$refers tooptimal solution thus far, utilizing Eq. (2):

$$x_1(t+1) = X_{best}(t) \times \left(1 - \frac{t}{T}\right) + (X_M(t) - X_{best}(t) \times rand) \tag{2}$$

Stage 3: Narrowed exploration

Here, Aquila involves in a shaped flight, glide above a shorter distance while encircling its prey. The behaviour is mathematically taken in Eq. (3):

$$X_2(t+1) = X_{best}(t) \times Levy(D) + (X_R(t) + (y - x) \times rand) \tag{3}$$

Stage 4: Expanded exploitation

Aquila implements a lower-level flight described by a slow growth in the attack angle, gradually approaches its prey that is exemplified by Eq. (4). Upper Bound (ub), Lower Bound (lb), and stable exploitation parameters α and δ are occupied.

$$X_3(t+1) = \big(X_{best}(t) \times X_M(t)\big) \times \alpha - rand + \big((ub - lb) \times rand + lb\big) \times \delta \tag{4}$$

Stage 5: Narrowed exploitation

The last stage is stimulated by Aquila's walk-and-grab attack, identified as Narrowed Exploitation, and defined in Eq. (5):

$$X_4(t+1) = QF \times X_{best}(t) - (G_1 \times X(t) \times rand) - (G_2 \times Levy(D) + rand \times G_1) \tag{5}$$

The fitness function (FF) reflects the classifier accuracy and the no. of chosen feature. It increases the classifier accuracy and decreases the set extents of preferred feature. Therefore, the below given FF is utilised to assess a discrete solution, as presented in Eq. (6).

$$Fitness = \alpha * \ ErrorRate + (1 - \alpha) * \frac{\#SF}{\#All\_F} \tag{6}$$

Where $ErrorRate$ represents the classifier rate of error by employing the preferred feature. $ErrorRate$ is intended as the proportion of incorrect classified to the amount of classification prepared amid 0 and 1, $\#SF$ represents the no. of preferred feature and $\#All\_F$ means a complete no. of attributes in an unique dataset. α is employed to control the significance of classifier quality and the length of sub-set.

C. CAE-based Classification Model

For the classification process, the proposed DFSFGDD-CAEMO technique designs CAE method.AE is a beneficial artificial neural network (ANN) device for diverse learning method of non-linear aspects and an unsupervised model, and its framework comprises decoder and encoder [20]. The encoder f alters high-dimension input data to low-dimension encode representation. The input data $x$is encode to h by encoder f:

$$h = f(x) = \sigma(Wx + b) \tag{7}$$

Now Wdenotes d×p weighted matrix, b specifies biased vector andσ signifies an activation function. The decoder g recreates h into x':

$$x' = g(h) = \ \sigma'(W'h + b') \tag{8}$$

Here W'depicts a p×d weighted matrix, b' indicates biased vector and σ' specifies an activation function. AE employs back-propagation model in training stage to decrease error of reconstruction J:

$$\underset{W,W',b.b'}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} J\big(x_{(i)}, x'_{(i)}\big) \tag{9}$$

Now $x_{(i)}$ and $x'_{(i)}$ refers to i^{th}training and decoded samples equivalent to $x_{(i)}$, correspondingly.

In this paper, distinct AE made by dense and convolution layer, which combines spatial data is employed. It is utilized a convolution process on input and transfers the result of succeeding layer. A convolution operation incorporated the feature within its receptive area and generates a particular value as output. Within convolutional layer, a kernel or filter moving through input data whereas managing element-wise multiplication and integrating the outcomes into a specific value of output. The kernel performed a uniform operation for every location as it moving through data. The decoder and encoder are accessible at top and bottom, correspondingly. While several

convolutional network frameworks are reported, a simple framework is intended to decrease the intricacy of method and time for computation calculation. Scaled Exponential Linear Unit (SELU), a non-linear function is employed as the active function in every layer except the last output layer that employed Sigmoid as the active function for reconstructing spectra. Usually 3x3 kernels are employed as the kernel dimension of convolution layers excepting the 1st and last 9th layers of encoder. Within these dual layers, point-wise convolution (1x1 kernel) is employed. The 1x1kernel function as dense layer for each individual aspects and computed the weighted amount for every point. Fig. 2 represents the structure of CAE model.

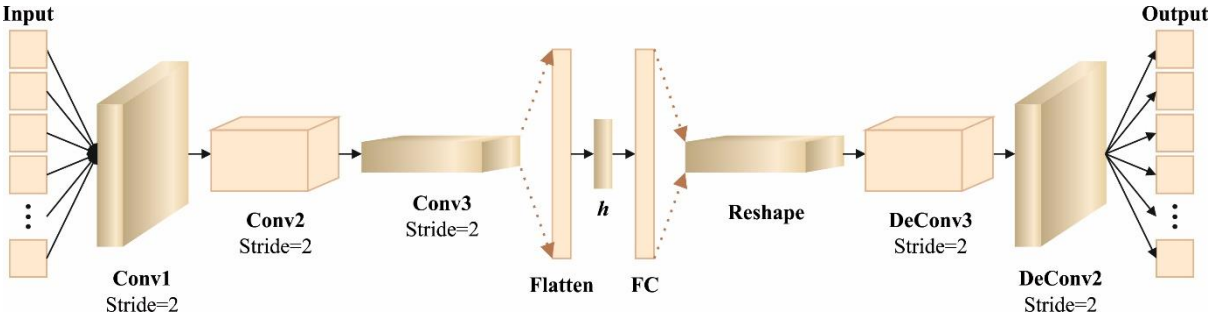

Figure 2. Structure of CAE model

D. EPOA-based Parameter Tuning Process

At last, the hyperactive parameter tuning process is performed through EPOA for improving the classification performance of CAE model. POA is an innovative swarm-based metaheuristic method replicates hunting process of pelicans [21]. It acquires searching and co-operative foraging mechanisms of pelican while search for food, particularly with group-based searching, exploitation and exploration models. While POA have rapid efficacy and convergence, it suffers from poor diversity and premature convergence, leads to sub-optimal solutions. To overwhelm these restrictions, the EPOA can integrate upgraded movement and dynamic mutation approaches to enhance search range and evade stagnation. POA replicating hunting process of pelicans that comprises dual basic approaches: approach for prey and soaring through water

Moving toward prey

During this phase, pelicans locate for prey and moving towards it. The movement approach relies on the fitness value in population. If the selected pelican have a greater fitness score, the existing pelican moving towards it; conversely, it moves away, improving exploration.

$$X_{i,d}^{t+1} = \begin{cases} X_{i,d}^{t} + rand.\left(X_{r,d}^{t} - I.X_{i,d}^{t}\right), \\ if F(X_r^t) < F(X_i^t) \\ X_{i,d}^{t} + rand.\left(X_{i,d}^{t} - I.X_{r,d}^{t}\right), \\ otherwise \end{cases} \tag{10}$$

Here $X_{r,d}^{t}$ indicates an arbitrarily chosen position of pelican has, $X_{i,d}^{t}$ denotes location of i^{th} pelican at iteration $t, F(X)$refers to fitness function assessing the excellence of solution and I signifies an arbitrary integer (1 or 2).

Winging on the water

Once move towards the prey, pelicans spreading their wings and disturbing the surface of water, force the prey to move upwards. This stage equivalent to the procedure of exploitation in optimization structure, refine solutions to modify their locations inside a localized neighbourhood.

$$X_{i,d}^{t+1} = X_{i,d}^{t} + R \cdot \left(1 - \frac{t}{T_{\max}}\right) \cdot (2 \cdot rand - 1) \cdot X_{i,d}^{t} \tag{11}$$

Now $T_{\max}$ depicts maximal iteration counts, R represents the neighbourhood radius and $rand$ refers to an arbitrary number among zero and one. R reduces linearly through time, safeguarding that hunting procedure transitions from global to local exploration, enhancing the set of solution.

Greedy selection mechanism

Every candidate solution is produced in the exploitation, exploration stages are assessed, and the finest solutions are kept to employ a greedy choice mechanism:

$$X_i^{t+1} = \begin{cases} X_i^{t+1}, & if F(X_i^{t+1}) < F(X_i^t) \\ X_{i,d}^t, & otherwise \end{cases} \tag{12}$$

These safeguards that every iteration preserves an enhancing or atleast stable set of solutions.

While POA is straightforward and efficient, it has restrictions regarding to restricted awareness of optimum solution and shrinking neighbourhood radius decreases diversity. Although reducing R enhances local search. The movement has been computed to employ Eq. (13).

$$X_{i,d}^{t+1} = X_{i,d}^t + \overrightarrow{r_3} \cdot \left(X_{best}^t - I \cdot X_{i,d}^t\right) \tag{13}$$

If a pelican lacking self-knowledge, it succeeds an arbitrary member of flock to guide utilizing Eq. (14).

$$X_{i,d}^{t+1} = X_{i,d}^t + \overrightarrow{r_4} \cdot \left(X_j^t - I \cdot X_{i,d}^t\right) \tag{14}$$

Here X_j^t indicates an arbitrarily chosen position of pelican.

While self-knowledge and member-based knowledge are inadequate, the pelican succeeds the finest leader and an arbitrary member:

$$X_{i,d}^{t+1} = X_{i,d}^t + \overrightarrow{r_5} \cdot \left(X_j^t - I \cdot X_i^t\right) + \overrightarrow{r_6} \cdot (2 \cdot \overrightarrow{r_7} \cdot X_{best}^t - I \cdot X_i^t) \tag{15}$$

$$X_{i,d}^t = \begin{cases} \max\left(X_{i,d}^t, LB_i^t\right) \\ \min\left(X_{i,d}^t, UB_i^t\right) \end{cases} \tag{16}$$

Now UB_i^t and LB_i^t refers to upper and lower bounds.

Furthermore, dynamic hunting learning (DHL) mutation improves diversity as below:

To compute adaptive searching neighbourhood

$$D_{\max}^t = |X_{best}^t - X_i^t| \tag{17}$$

To select neighbouring solutions

$$NS_i^t = \left\{X_N^t \middle| |X_i^t - X_N^t| \le D_{\max}^t, N = 1, \dots, N_{pop}\right\} \tag{18}$$

To implement adaptive mutation

$$X_{i,d}^{t+1} = X_{i,d}^t + R \cdot \left(1 - \frac{t}{T_{\max}}\right) \cdot (2 \cdot rand - 1) \cdot \left(X_{i,d}^t - X_{r,N,d}^t\right) \tag{19}$$

Compared with traditional POA that relies on direct positional upgrades depends upon the optimal solution, EPOA combines a several progressed approaches for improving exploration and stability of convergence.

The EPOA originates a FF to achieve enhanced performance of classifier. It expresses an optimistic numeral. to signify an enhanced performance of candidate solutions. The classifier rate of error reduction is measured as FF, as set in Eq. (20).

$$fitness(x_i) = ClassifierErrorRate(x_i)$$

$$= \frac{no.of\ misclassified\ samples}{Total\ no.of\ samples} * 100 \tag{20}$$

4. Performance Analysis

The experimental analysis of DFSFGDD-CAEMO model is examined under Genetic disorders dataset [22]. This dataset contains 18962 no. of instances under three class labels as portrayed in Table 1. The no. of features is 43, but only 29 features are chosen.

Table 1: Details of dataset

Class Labels	Description	No. of Instances
MIGID	"Mitochondrial genetic inheritance disorders"	9686
SGID	"Single-gene inheritance diseases"	7291
MUGID	"Multifactorial genetic inheritance disorders"	1985
Total No. of Instances		18962

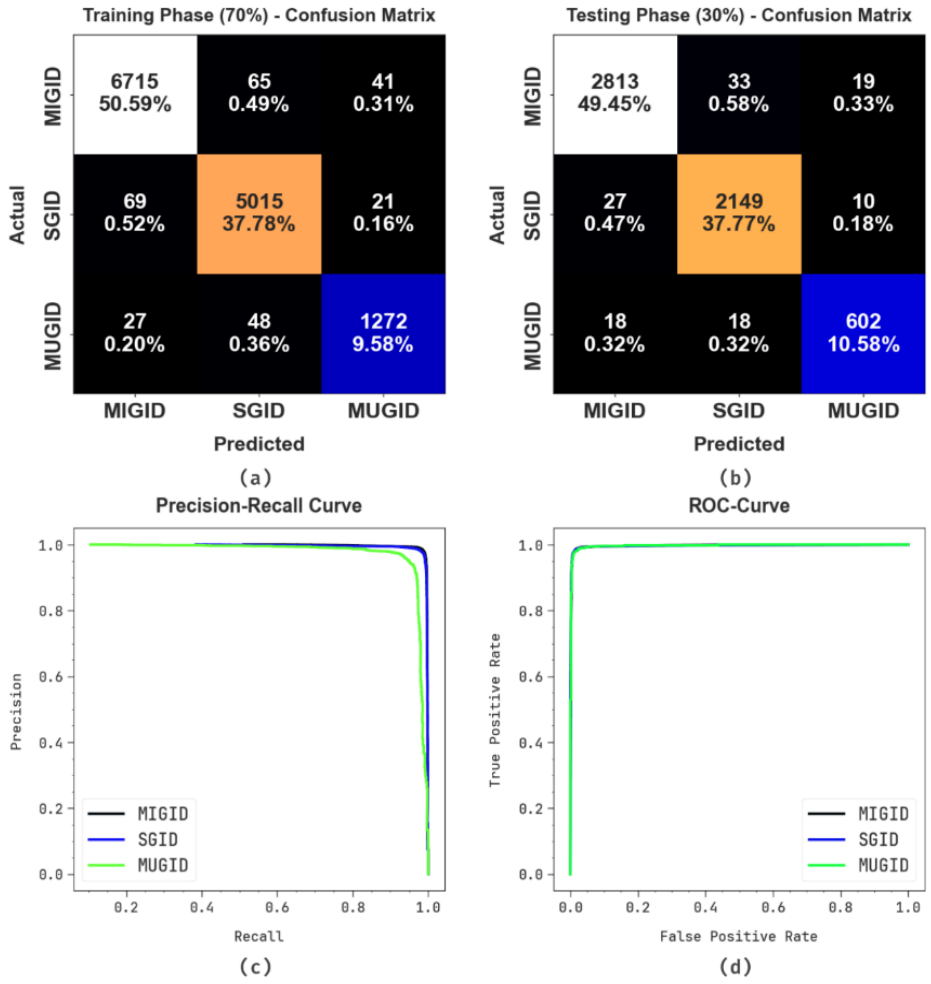

Figure 3. Classifier outcome of DFSFGDD-CAEMO model (a-c) confusion matrices and (b-d) PR and ROC curves

Fig. 3 depicts the classifier outcomes of DFSFGDD-CAEMO technique. Figs. 3a-3b demonstrate the confusion matrices with precise detection and classification of each class on 70:30. Fig. 3c shows the PR examination, denoting maximal performance in all classes. Finally, Fig. 3d demonstrates the ROC inspection, signifying efficacious outcomes with greater ROC values for separate class labels.

Table 2 and Fig. 4 present the genetic disorders detection of DFSFGDD-CAEMO system at 70:30. Based on 70% TRPHE, the proposed DFSFGDD-CAEMO model attains average $accu_y$ of 98.64%, $prec_n$ of 97.25%, $reca_l$ of 97.04%, $F_{Measure}$ of 97.14%, and MCC of 96.01%. Similarly, on 30% TSPHE, the proposed DFSFGDD-CAEMO model obtains average $accu_y$ of 98.54%, $prec_n$ of 97.17%, $reca_l$ of 96.95%, $F_{Measure}$ of 97.06%, and MCC of 95.85%.

Table 2: Genetic disorders detection of DFSFGDD-CAEMO model under 70:30

Class Labels	$Accu_y$	$Prec_n$	$Reca_l$	$F_{Measure}$	MCC
TRPHE (70%)					
MIGID	98.48	98.59	98.45	98.52	96.95
SGID	98.47	97.80	98.24	98.02	96.77
MUGID	98.97	95.35	94.43	94.89	94.32
Average	98.64	97.25	97.04	97.14	96.01
TSPHE (30%)					
MIGID	98.29	98.43	98.18	98.31	96.59
SGID	98.45	97.68	98.31	97.99	96.74
MUGID	98.86	95.40	94.36	94.88	94.24
Average	98.54	97.17	96.95	97.06	95.85

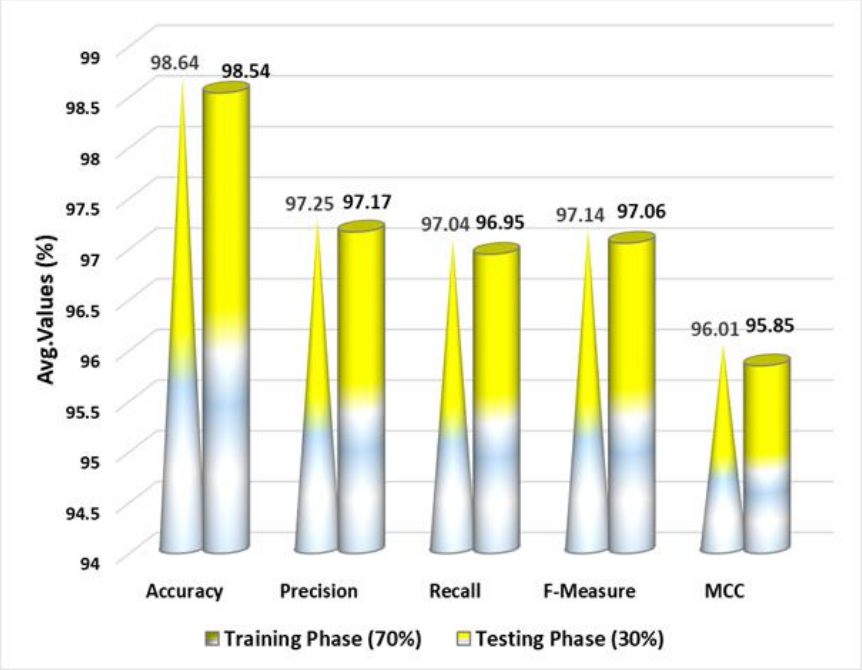

Figure 4. Average values of DFSFGDD-CAEMO model under 70:30

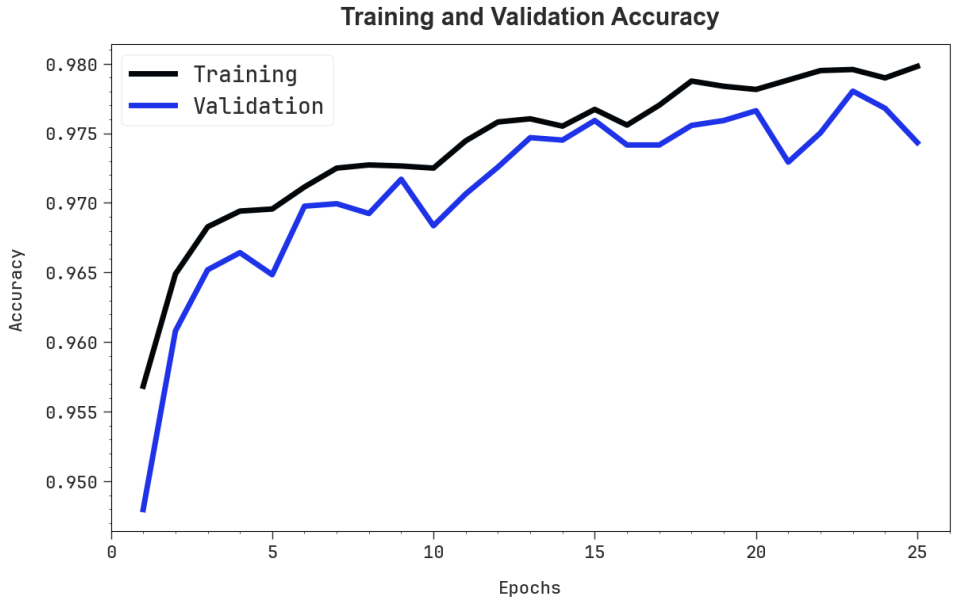

Figure 5. $Accu_y$ Curve of DFSFGDD-CAEMO model

Fig. 5 exemplifies the training (TRAIN) $accu_y$ and validation (VALID) $accu_y$ of a DFSFGDD-CAEMO method over 25 epochs. At first, both TRAIN and VALID $accu_y$ rise quickly, representing efficient pattern learning from the data. Around the epoch, the VALID $accu_y$slightly exceeds the training accuracy, implying good generalization without over-fitting. As training advances, it reflects higher performance and a lower performance gap between TRAIN and VALID. The close alignment of both curves in training suggests that the method is well regularized and generalized. This shows the method's stronger capability in learning and retaining valuable features across both seen and unseen data.

Fig. 6 demonstrates the TRAIN and VALID losses of DFSFGDD-CAEMO approach under 25 epochs. Initially, both TRAIN and VALID losses are higher, denoting that the approach starts with a partial understanding of the data. As training evolves, both losses persistently decline, displaying that the approach is efficiently learning and enhancing its parameters. The close alignment between the TRAIN and VALID loss curves during training implies that the model has not over-fitted and maintains good generalization to unseen data. This persistent and steady decrease in loss shows a reliable, well trained, and stable deep learning model.

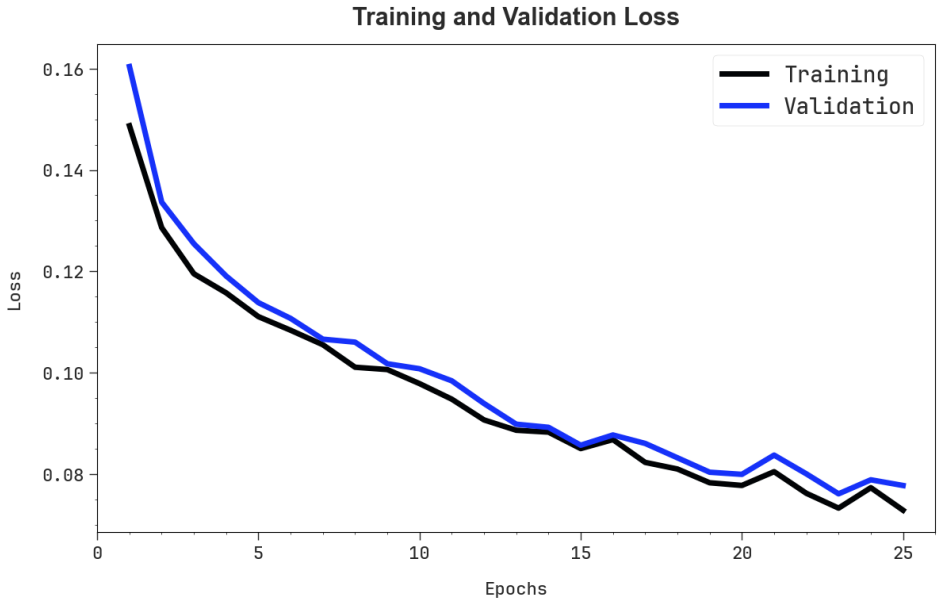

Figure 6. Loss curve of DFSFGDD-CAEMO method

Table 3 and Fig. 7 depict the comparative analysis of DFSFGDD-CAEMO system with present methods under various metrics [23-26]. The table values emphasized that the proposed DFSFGDD-CAEMO model got the highest $accu_y, prec_n$, $reca_l$, and $F_{Measure}$ of 98.64%, 97.25%, 97.04%, and 97.14%, respectively. While the current CNN-MGP method, SVM model, RNN technique, LSTM-RNN system, CNN-GRU methodology, Decision Tree algorithm, KNN model, HDLMOA-DGD method, ANN technique, AdaBoost algorithm, Ensemble Learning-GWO system, and EfficientNet model got worse performance.

Table 3: Comparative study of DFSFGDD-CAEMO model with existing systems

Methods	$Accu_y$	$Prec_n$	$Reca_l$	$F_{Measure}$
CNN-MGP	91.10	85.09	89.09	87.09
SVM Classifier	93.09	90.10	91.40	92.09
RNN Method	88.09	89.08	81.10	68.11
LSTM-RNN	95.20	88.07	85.59	87.09
CNN-GRU	97.32	96.18	95.53	95.07
Decision Tree	93.09	92.28	93.62	89.88
KNN Algorithm	93.48	92.57	90.72	89.10
HDLMOA-DGD	98.35	96.74	96.73	96.74
ANN	91.15	85.13	89.13	87.14
AdaBoost	93.12	90.14	91.43	92.11
Ensemble Learning-GWO	88.12	89.14	81.12	68.13
EfficientNet	95.24	88.12	85.63	87.15
DFSFGDD-CAEMO	98.64	97.25	97.04	97.14

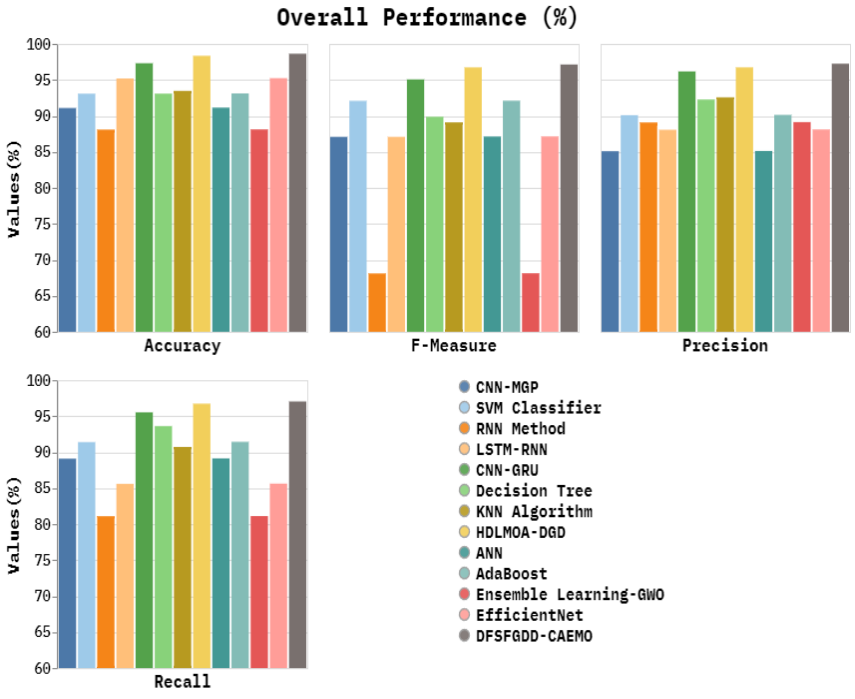

Figure 7. Comparative analysis of DFSFGDD-CAEMO model with existing techniques

In Table 4 and Fig. 8, the computational time (CT) outcome of DFSFGDD-CAEMO system with the current models is proven. The proposed DFSFGDD-CAEMO model offers less CT of 4.05sec while the CNN-MGP, SVM, RNN, LSTM-RNN, CNN-GRU, Decision Tree, KNN, HDLMOA-DGD, ANN, AdaBoost, Ensemble Learning-GWO, and EfficientNet methodologies got superior CT of 14.10sec, 8.47sec, 24.04sec, 11.89sec, 11.96sec, 14.43sec, 19.26sec, 6.38sec, 12.99sec, 13.09sec, 10.65sec, and 11.76sec, respectively.

Table 4: CT outcome of DFSFGDD-CAEMO model with existing methods

Methods	Computational Time (sec)
CNN-MGP	14.10
SVM Classifier	8.47
RNN Method	24.04
LSTM-RNN	11.89
CNN-GRU	11.96
Decision Tree	14.43
KNN Algorithm	19.26
HDLMOA-DGD	6.38
ANN	12.99
AdaBoost	13.09
Ensemble Learning-GWO	10.65
EfficientNet	11.76
DFSFGDD-CAEMO	4.05

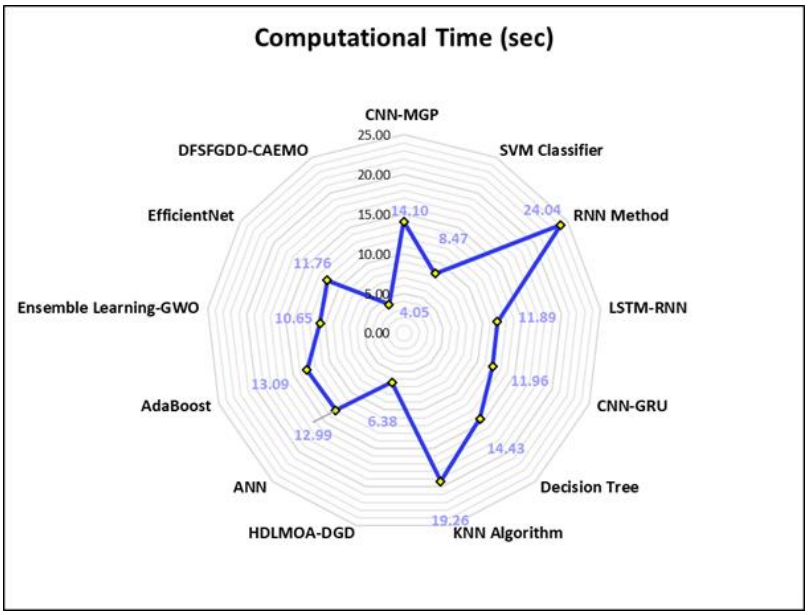

Figure 8. CT outcome of DFSFGDD-CAEMO model with recent methods

5. Conclusion

This paper presents a DFSFGDD-CAEMO model. The main of DFSFGDD-CAEMO model is to develop an accurate DNA-based genetic disorder classification model using advanced techniques for early and reliable disease diagnosis. Initially, the min-max normalization method is employed in the data pre-processing stage for transforming an input data into a beneficial format. Besides, the AO method has been deployed for the selection of feature process in order to select the most relevant features from a dataset. For the classification process, the proposed DFSFGDD-CAEMO technique designs CAE method. At last, the hyperparameter tuning process is performed through EPOA for improving the classification performance of CAE model. The experimental evaluation of the DFSFGDD-CAEMO technique occurs using benchmark dataset. The experimentation results indicated out the enhanced performance of the DFSFGDD-CAEMO system compared to existing approaches.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] Nasir, M.U., Gollapalli, M., Zubair, M., Saleem, M.A., Mehmood, S., Khan, M.A. and Mosavi, A., 2022. Advance genome disorder prediction model empowered with deep learning. *IEEE Access*, 10, pp.70317-70328.

[2] Zhou, J., Chen, Q., Braun, P.R., Perzel Mandell, K.A., Jaffe, A.E., Tan, H.Y., Hyde, T.M., Kleinman, J.E., Potash, J.B., Shinozaki, G. and Weinberger, D.R., 2022. Deep learning predicts DNA methylation regulatory variants in the human brain and elucidates the genetics of psychiatric disorders. *Proceedings of the National Academy of Sciences*, 119(34), p.e2206069119.

[3] Alharbi, W.S. and Rashid, M., 2022. A review of deep learning applications in human genomics using next-generation sequencing data. *Human Genomics*, 16(1), p.26.

[4] Park, C., Ha, J. and Park, S., 2020. Prediction of Alzheimer's disease based on deep neural network by integrating gene expression and DNA methylation dataset. *Expert Systems with Applications*, 140, p.112873.

[5] Eraslan, G., Avsec, Ž., Gagneur, J. and Theis, F.J., 2019. Deep learning: new computational modelling techniques for genomics. *Nature Reviews Genetics*, 20(7), pp.389-403.

[6] Luo, P., Li, Y., Tian, L.P. and Wu, F.X., 2019. Enhancing the prediction of disease–gene associations with multimodal deep learning. *Bioinformatics*, 35(19), pp.3735-3742.

[7] Koumakis, L., 2020. Deep learning models in genomics; are we there yet? *Computational and Structural Biotechnology Journal*, 18, pp.1466-1473.

[8] Ghazal, T.M., Al Hamadi, H., Umar Nasir, M., Gollapalli, M., Zubair, M., Adnan Khan, M. and Yeob Yeun, C., 2022. Supervised machine learning empowered multifactorial genetic inheritance disorder prediction. *Computational Intelligence and Neuroscience*, 2022(1), p.1051388.

[9] Novakovsky, G., Dexter, N., Libbrecht, M.W., Wasserman, W.W. and Mostafavi, S., 2023. Obtaining genetics insights from deep learning via explainable artificial intelligence. *Nature Reviews Genetics*, 24(2), pp.125-137.

[10] Zhang, Y., & Wang, Y. (2023). "Deep learning-based genomic data analysis: A survey." *IEEE Transactions on Computational Biology and Bioinformatics*, 20(2), pp.1234-1245.

[11] Pimpalkar, A., Gandhewar, N., Shelke, N., Patil, S. and Chhabria, S., 2025. An Efficient Deep Convolutional Neural Networks Model for Genomic Sequence Classification. *Genomics at the Nexus of AI, Computer Vision, and Machine Learning*, pp.345-375.

[12] Gala, P., Pandloskar, Y., Godbole, S., Hakim, F., Kanani, P. and Kurup, L., 2025. Classification of Sarcoma Based on Genomic Data Using Machine Learning Models. *Procedia Computer Science*, 252, pp.317-330.

[13] Das, A., Neelima, N., Deepa, K. and Özer, T., 2024. Gene selection based cancer classification with adaptive optimization using deep learning architecture. *IEEE Access*.

[14] Zhang, X., Wang, Y., Wei, Q., He, S., Salhi, A. and Yu, B., 2024. DRBPPred-GAT: Accurate prediction of DNA-binding proteins and RNA-binding proteins based on graph multi-head attention network. *Knowledge-Based Systems*, 285, p.111354.

[15] Meenaakshisundhari, R.P., Murali, L., Rajesh Sharma, R. and Nivethitha, A., 2023. Autism Spectrum Disorder Classification Using Enhanced Whale Optimization and Improved Convolutional Neural Network Algorithm.

[16] Alakuş, T.B., 2023. A novel repetition frequency-based DNA encoding scheme to predict human and mouse DNA enhancers with deep learning. *Biomimetics*, 8(2), p.218.

[17] Thakur, T., Batra, I., Malik, A., Ghimire, D., Kim, S.H. and Hosen, A.S., 2023. RNN-CNN based cancer prediction model for gene expression. *IEEE Access*, 11, pp.131024-131044.

[18] Chang, Y., the Role of Mathematical Algorithms in Advancing Computer Artificial Intelligence. Available at SSRN 5277622.

[19] Nassef, A.M., Ghadbane, H.E., Sayed, E.T. and Rezk, H., Reducing hydrogen consumption in hybrid electric vehicles using Aquila optimization algorithm.

[20] Gao, Y., & Wang, H. (2021). "Deep learning for bioinformatics: Overview and future directions." *Briefings in Bioinformatics*, 22(4), pp. 1160-1170.

[21] Chen, Y., Guo, Y., Gao, Y. and Liu, B., 2025. A novel lightweight deep learning framework using enhanced pelican optimization for efficient cyberattack detection in the Internet of Things environments. *Journal of Engineering and Applied Science*, 72(1), pp.1-26.

[22] Alshammari, T., 2024. Applying machine-learning algorithms for the classification of sleep disorders. *IEEE Access*.

[23] Shahin, O.R., Alanazi, F.M., ElDeeb, M.K., Kuriri, F.A., Kwa, F.A., Alenazy, F.O. and Kamal, T.M., 2025. Construction of Saudi computational gene models: Applications in healthcare of prevalent genetic disorders. *Alexandria Engineering Journal*, 117, pp.13-26.

[24] Alzahrani, A.A. and Alharithi, F.S., 2024. Machine learning approaches for advanced detection of rare genetic disorders in whole-genome sequencing. *Alexandria Engineering Journal*, 106, pp.582-593.

[25] Balasundaram, A., Shaik, A., Alroy, B.R., Singh, A. and Shivaprakash, S.J., 2024. Genetic algorithm optimized stacking approach to skin disease detection. *IEEE Access*.

[26] Kumar, A., & Gupta, R. (2022). "Machine learning for genomic data analysis: Recent trends and future directions." *IEEE Access*, 10, pp.12345-12358.