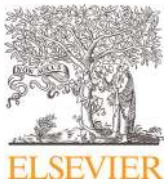
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Computers and Electrical Engineering 87 (2020) 106765



Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Deep learning disease prediction model for use with intelligent robots

Srinivas Koppu[a], Praveen Kumar Reddy Maddikunta[a], Gautam Srivastava[b,c,*]

[a] School of Information Technology and Engineering, VIT - Vellore, Tamilnadu, India
[b] Department of Mathematics and Computer Science, Brandon University, 270 18th Street, Brandon, R7A 6A9 Canada
[c] Research Center for Interneural Computing, China Medical University, Taichung 40402, Taiwan, Republic of China

ARTICLE INFO

Article history:
Received 30 October 2019
Revised 29 May 2020
Accepted 13 July 2020

Keywords:
COVID-19
Deep learning
Intelligent robotics
Data cleaning
Disease prediction
Dragonfly optimization
Feature extraction
Fitness basis

ABSTRACT

Deep learning applications with robotics contribute to massive challenges that are not addressed in machine learning. The present world is currently suffering from the COVID-19 pandemic, and millions of lives are getting affected every day with extremely high death counts. Early detection of the disease would provide an opportunity for proactive treatment to save lives, which is the primary research objective of this study. The proposed prediction model caters to this objective following a stepwise approach through cleaning, feature extraction, and classification. The cleaning process constitutes the cleaning of missing values ,which is proceeded by outlier detection using the interpolation of splines and entropy-correlation. The cleaned data is then subjected to a feature extraction process using Principle Component Analysis. A Fitness Oriented Dragon Fly algorithm is introduced to select optimal features, and the resultant feature vector is fed into the Deep Belief Network. The overall accuracy of the proposed scheme experimentally evaluated with the traditional state of the art models. The results highlighted the superiority of the proposed model wherein it was observed to be 6.96% better than Firefly, 6.7% better than Particle Swarm Optimization, 6.96% better than Gray Wolf Optimization ad 7.22% better than Dragonfly Algorithm.

1. Introduction

A robot is an active agent that interacts with the real world and often works under benevolent and uncontrolled circumstances. While robots may have predefined or pre-trained capabilities, they should be able to adapt or expand to carry out new and useful tasks. It is still challenging to deploy robots with pre-built skills and the ability to acquire new skills. Therefore, applying deep learning to robotics stimulates research questions [1]. Deep Reinforcement Learning (DRL) has been successfully applied in various workplaces to learn complex arrangements through vast volume perceptions, e.g., pictures. Robots can, therefore, be used to carry out day-to-day activities such as washing and folding clothes, cooking, and cleaning. Applying Deep Learning (DL) approaches to real humanoid robots, however, remains a significant challenge, as traditional DRL techniques involve a large number of learning samples [2]. DL and Artificial Intelligence (AI) have received a lot of attention over the last few years [3,4]. These technologies are revolutionizing various applications such as healthcare, computer vision, pattern recognition, robots, self-driving cars, and automatic machine translation, etc.

* Corresponding author.
E-mail addresses: srinukoppu@vit.ac.in (S. Koppu), praveenkumarreddy@vit.ac.in (P.K.R. Maddikunta), srivastavag@brandonu.ca (G. Srivastava).

https://doi.org/10.1016/j.compeleceng.2020.106765
0045-7906/

Healthcare is one of the most widely used applications of these technologies. DL stack up a large volume of patient data, including patient, medical, and insurance records, into neural networks to develop better results [5,6]. Disease prediction systems have been playing a significant role in the life of people, and it has been considered an important topic by many academics [7]. In healthcare management, data mining holds a significant role in predicting diseases.

Disease Prediction Systems (DPSs), using a variety of data mining techniques, recently has attracted considerable attention. One of the most common data mining methods, the Single-Layer Perceptron (SLP) classifier, was widely used to predict various diseases. AI and Machine Learning (ML) approaches have been integrated to resolve medical care issues in many real-world situations. In recent times, Neural Network (NN) ensembles were effectively exploited in several applications and may assist in medical analysis. NN ensembles could considerably enhance the simplification capabilities of learning systems via training a predetermined amount of NN and then merging their outcomes.

Many researchers have implemented ML algorithms around the globe to help aid the fight against COVID-19, Heart disease, and Breast Cancer. Although prediction results achieved are promising, there is further scope in the improvement of the results by improved feature engineering. These factors are the main motivation for the work in front of you.

Several pre-processing techniques, such as Spline Interpolation (SI), Principal Component Analysis (PCA), have been studied to improve the prediction results of AI-based methodologies.SI is used for filling missing values in the dataset, whereas PCA is utilized for selecting the optimal features from the dataset considered. To further improve the prediction accuracy of Deep Belief Network Models, hyper-parameter tuning can be performed by the Fitness Oriented Dragonfly Optimization algorithm (F-DA).

The steps involved in the proposed model are given below:

1. The disease prediction model involves three modules: (a) data cleaning, (b) feature extraction, and (c) classification.
2. Firstly, the two heart disease datasets, “Cleveland and Statlog” and a breast cancer dataset known as Wisconsin Breast Cancer (WBC) were attained from the UCI data repository. These datasets are fed to the data cleaning process in preprocessing. Filling up of missing values and outlier detection are the two phases in this step. Spline Interpolation (SI) is exploited for filling of missing value, and entropy, correlation is used for outlier detection.
3. The resultant data from the data cleaning process is fed into the feature extraction process, where Principle Component Analysis (PCA) is employed for dimensionality reduction. Further, the extracted features are multiplied with a weight, and then subjected to the classification process.
4. The resultant feature vector is then fed to the Deep Belief Network (DBN) framework. Accordingly, the multiplied weight is optimally tuned by Fitness Oriented Dragonfly Optimization, such that the error between the actual and predicted output is minimized.
5. The proposed approach is then evaluated over other traditional models, namely, Particle Swarm Optimization (PSO), FireFly (FF), Grey Wolf Optimization (GWO), and Dragonfly Algorithm (DA).

The rest of the paper is organized as follows. Section 2 describes the literature works related proposed system. Section 3 introduces the disease prediction model with various adapted stages. Section 4 explains the weight optimization by the F-DA algorithm. Then disucss our main results in Section 5. Finally, Section 6 concludes the paper with future work directions and closing remarks.

2. Literature reviews

2.1. Related works

Loey *et al.* [8] have proposed a novel predication deep learning-based Generative Adversarial Network (GAN) model for COVID-19 X-ray datasets. The main idea of the proposed model is to take input as X-ray images. Virus detection is performed by the GAN Deep Learning Classification Model. The developed system reduces complexity, memory consumption, and time. Zhou *et al.* [9] have developed an automatic COVID-19 CT segmentation model using U-net spatial and channel attention architecture. U-net contains two models, such as encoder and decoder, for the segmentation of the medical image of COVID-19. Contextual relationships are identified in ROI by spatial and channel attention for better representation in segmentation. However, the proposed system performed experimentally on limited datasets. Ghoshal *et al.* [10] used a Bayesian Convolutional Neural Network (BCNN) deep learning dataset to detect Coronavirus Uncertainty and Interpretability (COVID-19) X-ray dataset to improve diagnostic performance. Estimation of uncertainty in deep learning results in more reliable disease predictions, which could lead practitioners to false assumptions.

Nilashi *et al.* [11] presented a novel system for predicting diseases through clustering methods. Here, CART (Classification and Regression Trees) was used to develop a set of fuzzy-based rules. In addition, the results have shown that the technique adopted significantly improves the accuracy of disease predictions. Chuan *et al.* [12] have adopted Privacy-Preserving Disease Prediction (PPDP). Here, the patient’s historical medical data was outsourced and encrypted, which could be further used to train predictive approaches safely. The threat of disease to emerging medical statistics could be assessed on the basis of prediction methods. Chen *et al.* [13] introduced a computational methodology of an ensemble approach to detect miRNA-disease. The investigations were conducted on three major tumors. Besides, Kidney Cancer, Prostate Cancer, and Lymphoma of the top fifty predicted MiRNAs were established by current experimental results, which illustrated the consistent predictability of HAMDA.

Parisot *et al.* [14] suggested a methodological assessment of the generic structure, including non-imagery and imaging data referred to as Graphic Convolutional Networks (GCNs) and which could be used for brain research in dense populations. In addition, the broad estimation identifies the consequences of each element of this model on disease prediction and further assesses it on a number of baselines. Weng *et al.* [15] designed a framework for examining the performance of different classifiers together with the individual classifiers concerned in the Ensemble Classifier (EC). Statistical tests were carried out to assess the performance differentiation between the types of classifiers. Kumar *et al.* [16] have formulated a new methodological regime for diabetes. In addition, a novel classification methodology based on the fuzzy rule for the treatment of disease has been established. The experiments were carried out using the original health records collected from different hospitals. Luo *et al.* [17] presented a new technique for the MiRNA-Association of Diseases Dependant on the Graphic Regularization Framework (MDAGRF). The performance of this technique was not susceptible to a range of constraints. By distinguishing from other traditional models, MDAGRF was able to achieve improved predictive outcomes for certain diseases. Sengupta and Asit [18] modeled a scheme based on the concepts of PSO and Association Rule Mining (ARM). The incremental classifier was appropriate to be relevant for predicting disease, as the characteristics of the diseases vary in terms of time due to changes in climate, geographical, and biological aspects.

2.2. Reviews

At first, the Fuzzy Logic System (FLS) was introduced in [11], which offers a better diagnosis of diseases, and it offers better noise elimination. However, it needs more contemplation on other datasets. Privacy-Preserving Disease Prediction (PPDP) was exploited in [12] that offers improved privacy protection with low computational complexity, but it requires more effective PPDP models. Also, Hybrid Approach for MiRNA-Disease Association prediction (HAMDA) approach was deployed in [13] that provides better performance and improved prediction ability. Anyhow, it needs to be confirmed by the biological observations in the future. Likewise, the GCN scheme was exploited in [14], which offers increased accuracy, and it also predicts several labels. However, there were occurrences of imbalance proportions of enormous class. Also, EC was employed in [15], which offers increased accuracy, and it is highly cost-effective; however, the exterior validity of EC has to be measured in advance. A fuzzy rule was exploited in [16] that offers enhanced specificity, and it offers better security levels, anyhow, it requires considerations on cryptographic approaches. MDAGRF was implemented in [17], which provides a better prediction of diseases, and it also offers consistent data, but this approach needs consideration on noise effects. Also, At last, the PSO algorithm was suggested in [18] that provides more accurate outcomes and improved efficiency. However, it requires contemplation on incremental data with diverse feature sets. The discussed limitations must be considered to enhance the performance of disease prediction models successfully in the presented work.

3. Disease prediction model: Various adopted stages

3.1. Proposed architecture

The diagrammatic representation of the proposed architecture for disease prediction is given by Fig. 1.
The different steps of the proposed model are explained below:

1. In the first step, the collected data on heart disease and breast cancer are applied to the data cleaning process. Two steps are involved in the cleaning process, the first one is missing value filling, and the subsequent one is outlier detection. Here, in the proposed work, Spline Interpolation (SI) is used for missing value filling, whereas entropy and the correlation-based procedure is followed for detecting the outlier.
2. In the second step, the resultant data from data cleaning are given to the feature extraction process, where Principal Component Analysis (PCA) is used.
3. In the third step, features extracted from PCA is transformed into another form of feature vector by multiplying with a weight function.
4. In the fourth step, the resultant feature vector is conveyed to the Deep Belief Network (DBN) framework. As the main contribution, the multiplied weight is tuned optimally by modified DA known as F-DA, such that the error among the actual and predicted output is minimized during the classification process. The classification output provides the labels that differentiate whether the patient is affected or not.

3.2. Data cleaning

Data cleaning refers to “the process of detecting and correcting (or removing) defective or inaccurate information from a recordset, table or database and helps recognize incomplete, incorrect, inaccurate or irrelevant data parts and then replace, change or remove dirty or coarse data” [19].

3.2.1. Missing value filling

Spline Transformation (ST) fills missing value. Splines are polynomials, smoothly linked together. The meeting points of polynomials are called “knots.” Considering a spline of n degree, coefficients of $n+1$ are needed to explain each piece. There is an extra smoothing factor that stimulates the spline ’s stability to order $n-1$ at the “knots”.

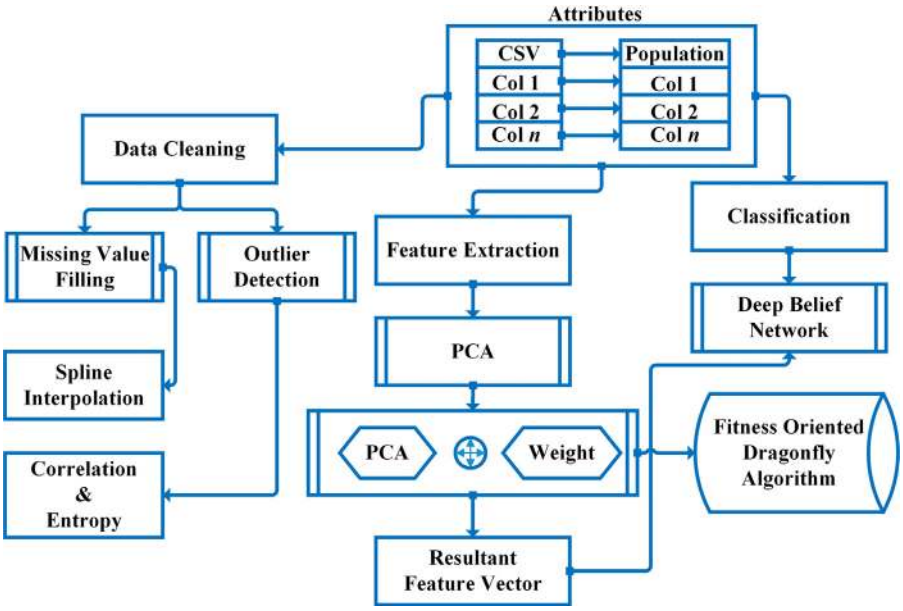


Fig. 1. Proposed Disease Prediction Model.

3.2.2. Outlier detection

Outlier detection may be determined by a procedure that exploits entropy and correlation.

1. Find the correlation: Correlation is one of the statistical metrics that indicates the extent to which two or more parameters oscillate together. The formulation for correlation is given by Eq. 1, where a_o indicates the attributes and $o = \{1, 2, 3, \ldots N\}$, in which N denotes the number of attributes. The common correlation formulation is given by Eq. 2, where x and y indicate the variables and k refers to the number of point pairs.

$$C_o = Corr(a_0, label) \tag{1}$$

$$Corr = \frac{k(\sum xy) - (\sum x)(\sum y)}{\sqrt{[k\sum x^2][k\sum y^2 - (\sum y^2)]}} \tag{2}$$

2. Find the entropy: Entropy is "a measure of the unpredictability of the state, or equivalent, of its average information content," as given by Eq. 3, and Eq. 4 gives the formulation of entropy. In Eq. 4, p denotes the probability of occurrence, and b indicates the base.

$$E_n = Ent(C_n) \tag{3}$$

$$Ent = -\sum_{i=1}^{\nu} p_i log_b(p_i)) \tag{4}$$

3. Find elements: If the user-defined threshold is more significant, replace it with zero or else replace it with one.
4. Drop the number of an outlier (records) has maximum zeros. Here, the number of outliers is user-defined.

Thus the outlier detected attribute resulting from the above procedures is indicated by D_H. D_D indicates the resultant data cleaning attribute attained from missing value filling and outlier detection.

3.3. Feature extraction and feature tranformation

The attained D_D from data cleaning is given to PCA for feature extraction.PCA [20] is a conventional method that reduces the enormous dimensional features of the data. The extracted features from PCA fe_b are not directly given to the classfier, rather, it is multiplied with a weight function w_b and forms a new feature vector as given by Eq. 5.

$$F_b = fe_1w_1 + fe_2w_2 + fe_3w_3 + \ldots . fe_Nw_f \tag{5}$$

Thus the final feature vector indicated by F_b is further given to DBN for classification, which can predict the concerned disease.

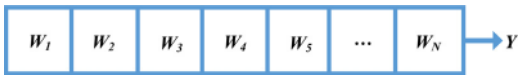


Fig. 2. Weight Encoding.

3.4. Classification using DBN

DBN is an intelligent classification framework that employs independent layers, visible and hidden neurons that develop the output layer [21]. The link between hidden and visible neurons is exclusive and symmetric.

4. Weight optimization by fitness oriented dragonfly optimization

4.1. System classification and fitness function

The weights that are multiplied with the features are given as solutions for encoding, as given by Fig. 2. Consequently, weights must be designed to minimize the difference between the expected results and the actual results, which is given by Eq. 6. This paper uses a new modified algorithm called F-DA for weight optimization.

$$Objective \quad Function, R = Min\left(e_l^{\hat{Z}}\right) \tag{6}$$

4.2. Dragonfly optimization

For weight optimization, we proposed a modified F-DA. Generally speaking, DA [22] approach emerges from adaptive and stable swarming processes. With the application of meta-heuristics, these two processes are closely correlated with the two main stages of optimization, namely (i) exploration phase and (ii) exploitation phase. These two phases are as follows: Demarcation formulation is assessed as given in Eq. 7, Where *YJ* determines a neighbor's J^{th} location, *Y* represents the current individual's location, and *N* represents the neighboring entities.

$$O_i = -\sum_{j=1}^{N_e} Y - Y_j \tag{7}$$

As shown in Eq. 8, where *Qj* is the velocity of an adjacent j^{th} individual. Furthermore, Eq. 9 gives the cohesion formulation, where *Yj* implies neighboring j^{th} position,*Ne* represents the neighborhood quantities and *Y* represents the current individual's area.

$$B_i = \frac{\sum_{j=1}^{N_e} Q_j}{N_e} \tag{8}$$

$$G_i = \frac{\sum_{j=1}^{N_e} Y_j}{N_e} - Y \tag{9}$$

Eq. (10) represents attractiveness to food, where the food source is Y^+, and the present location is *Y*.

$$F_i = Y^+ - Y \tag{10}$$

Eq. (11) describes diversion to an enemy, where Y^- represents the enemy's place, and *Y* represents the individual's place.

$$E_i = Y^- + Y \tag{11}$$

Two vectors are assessed to change dragonflies location in the exploration phase and their activities are described as : step ($\triangle Y$) and position (Y).

The step vector shows the dragonfly movement path, which is represented in Eq. (12). Here, *Oi* denotes the isolation of i^{th} entity, *p* indicates the separation value, *a* indicates the alignment value, *G* is the i^{th} cohesion, *c* is the cohesion value, *B* is the individual i^{th} alignment, F_i is the food resource, *f* is the food component, *e* is the enemy element, *w* is the inertia weight, E_i is the enemy's position of the i^{th} entity and *t* is the iteration counter.

$$\triangle Y(t+1) = (pO_i + aB_i + cG_i + fF_i + eE_i) + w\triangle Y(t) \tag{12}$$

Following the phase vector evaluation, Eq. (13), manipulates position vectors, where *t* is the latest iteration.

$$Y(t+1) = Y(t) + \triangle Y(t+1) \tag{13}$$

To improve artificial dragonfly's deterministic efficiency, flying around the exploration space using an arbitrary walk is essential in such situations, Eq. (14) modifies the position of the dragonfly, where *z* implies location vector and *t* is the current state.

$$Y(t+1) = Y(t) + Levy(z) * Y(t) \tag{14}$$

Eq. (15) evaluates Levy flight, where β is a prime factor and $r1$, $r2$ are arbitrary numbers of [0, 1]. δ is calculated using Eq. (16), in which $\tau(x) = (x-1)$.

$$Levy(x) = 0.01 * \frac{r_1 * \delta}{\mid r_2 \mid^{\frac{1}{\beta}}} \tag{15}$$

$$\delta = \left[\frac{\tau(1+\beta) * \sin(\frac{\pi\beta}{2})}{\tau\frac{(1+\beta)*}{2} * \beta * 2^{\frac{(\beta-1)}{2}}} \right]^{\frac{1}{\beta}} \tag{16}$$

4.3. Proposed F-DA algorithm

Although interesting factual information about the traditional DA algorithm, it has some drawbacks like diminished internal storage and weak convergence. In the conventional method, dragonfly's neighborhood is measured by comparing the current solution to its surrounding solutions. Accordingly, the nearby solutions with minimum distance will be selected as an optimal neighborhood. The proposed approach selects the neighborhood based on its fitness function *f*. Initially, the mean fitness, $f(i)$ is determined, and the fitness values that are less than the mean fitness selected as neighborhood. All the other fitness functions are neglected. Algorithm 1 describes the pseudo-code of the proposed F-DA model.

Algorithm 1: Proposed F-DA Algorithm.

```
Set up the dragonfly population as Y_i(i = 1, 2, ....n)
Set up the step vectors as △Y_i(i = 1, 2, ....n)
while condition is not satisfied do
    Calculate the fireflies' fitness value
    Upgrade food, enemy
    Upgrade w, e, a, f, c, and s
    Measure B, O, F, E and G using Eq.7 - 11
    Determine mean fitness, f_i
    if fitness, f < f_i then
        Select it as neighborhoods
    else
        Neglect it
    end
    if dragonfly concerns a neighbor's dragonfly, then
        Upgrade velocity vector with the aid of Eq. 12
        Upgrade position vector with the aid of Eq. 13
    else
        Upgrade position vector with the aid of Eq.14
    end
    Review and accept new positions, based upon variable thresholds
end
```

5. Results and discussion

5.1. Simulation mechanism

To carry out the experimentation process, we used a Windows 10 Operating System laptop with 8 GD RAM. The proposed disease prediction model implemented in MATLAB R2015a. The operation performed using three datasets attained from the UCI data repository. Cleveland's heart data set with 303 instances and 75 attributes. Statlog heart data set with 270 instances and 13 attributes. Wisconsin Breast Cancer (WBC) data set with 569 cases of cancer, and 32 attributes.

The implemented disease prediction model distinguished with other traditional methods together with FF [23], PSO [24], GWO [25] and DA [22]. Specific performance measurements like accuracy, precision, specificity, sensitivity, F1-Score, NPV, MCC, FNR, FPR, FDR and corresponding results were obtained.

5.2. Performance evaluation using cleveland dataset

Fig. 3 determines the performance evaluation of the proposed disease prediction model using the Cleveland dataset for heart disease. Fig. 3(a), shows the proposed F-DA model produces better accuracy of 3.7% at 80th iteration than PSO, 1.23% greater than GWO, and 2.47% greater than DA algorithms. Fig. 3(b), shows the proposed model for sensitivity at 80th

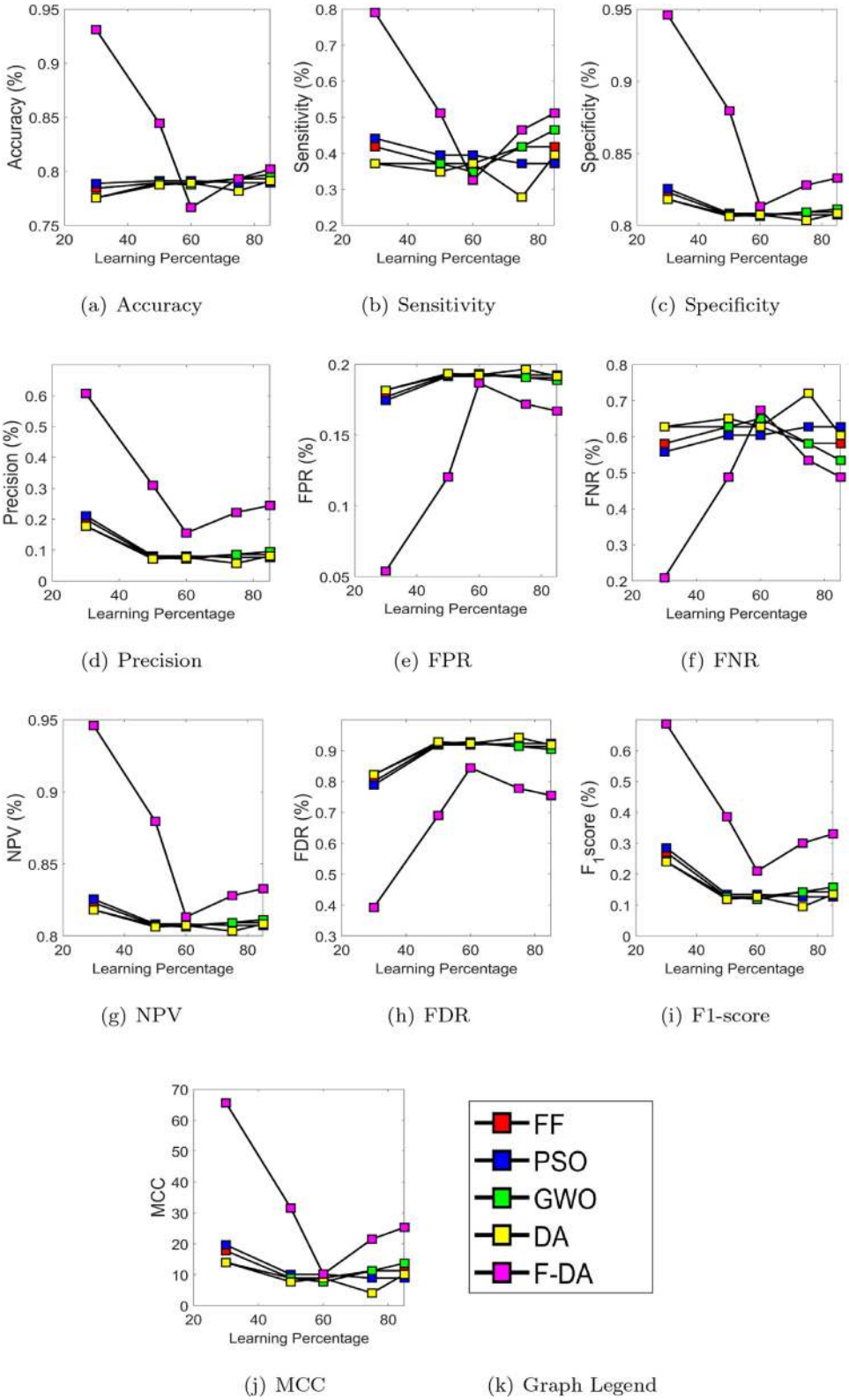


Fig. 3. Performance evaluation of proposed and conventional disease prediction models for different learning percentage using Cleveland Dataset (a) Accuracy (b) Sensitivity (c) Specificity (d) Precision (e) FPR (f) FNR (g) NPV (h) FDR (i) F1-score (j) MCC (k) Graph Legend.

Table 1
Performance measures using the proposed model and traditional models for Cleveland Dataset
.

Cleveland Dataset					
Measures	FF[23]	PSO[24]	GWO[25]	DA[22]	F-DA
Accuracy	0.78947	0.79139	0.78947	0.7875	0.8444
Sensitivity	0.37209	0.39535	0.37209	0.3488	0.5116
Specificity	0.80739	0.80838	0.80739	0.8063	0.8796
Precision	0.076555	0.08134	0.076555	0.0717	0.3098
FPR	0.19261	0.19162	0.19261	0.1936	0.1203
FNR	0.62791	0.60465	0.62791	0.6511	0.4883
NPV	0.80739	0.80838	0.80739	0.8063	0.8796
FDR	0.92344	0.91866	0.92344	0.9282	0.6901
F1_score	0.12698	0.13492	0.12698	0.1190	0.3859
MCC	0.089126	0.10117	0.089126	0.0770	0.3155

iteration is 18%, 22%, 10%, and 20% better than FF, PSO, GWO, and DA algorithms respectively. Also, Fig. 3(c), shows the precision at 30th iteration is 67.2% greater than FF, 62.3% greater than PSO, and 70.49% greater than DA algorithms. In Fig. 3(d), the F-DA model shows the results of FPR at 30th iteration is 68.15%, 67.74%, and 72.22% better than FF, PSO, and DA algorithms, respectively.

5.3. Performance evaluation using statlog dataset

Fig. 4 exhibits the performance analysis of the proposed F-DA model for predicting heart disease using statlog dataset. Fig. 4(a), shows the accuracy of the proposed model at 30th iteration is 3.19% greater than FF, 0.53% greater than PSO, 2.13% greater than GWO, and 3.72% greater than DA algorithms. In Fig. 4(b), the sensitivity of the F-DA model at 60th iteration is 2.17%, 4.35%, 4.89%, and 8.69% better than FF, PSO, GWO, and DA models, respectively. Fig. 4(f), presents the FNR of the proposed model at 30th iteration is 45.83% better than FF, 8.33% better than PSO, 26.67% better than GWO and 66.67% better than DA algorithms. Also, from Fig. 4(g), the F-DA model in terms of NPV at 30th iteration is 1.58% better than FF, 0.53% better than PSO, 1.05% better than GWO, and 2.11% better than DA algorithms. Thus, the enhancement of the proposed F-DA model is verified by the attained results.

5.4. Performance evaluation using wisconsin dataset

The WBC dataset deployed for predicting breast cancer performance analysis is shown in Fig. 5. Fig. 5(a), shows the accuracy of the adopted F-DA model for 30th iteration is 40.5% greater than GWO and 43.03% greater than DA algorithms. Consequently, the specificity of the presented model is shown in Fig. 5(c) at 45th iteration is 24.05% better than GWO and 26.58% better than DA algorithms. Also, from Fig. 5(h), at 30th iteration, the FDR of the implemented F-DA model is 51.09% greater than GWO and 52.63% greater than DA algorithms. Moreover, Fig. 5(i), exhibits the adopted scheme at 30th iteration in terms of F1-score is 91.67% better than GWO and 65% better than DA algorithms. Finally, Fig. 5(j), shows the proposed method at 30th iteration for MCC is 37.78% greater than GWO and 33.33% greater than DA algorithms. Therefore, the improvements in the F-DA technique for weight optimization was substantiated successfully.

5.5. Overall performance evaluation

Table 1 gives the overall performance evaluation of the disease prediction for Cleveland dataset. The proposed system obtained an accuracy of 6.96%, 6.7%, 6.96%, and 7.22% better than FF, PSO, GWO, and DA models, respectively. The sensitivity of the F-DA model is 37.5%, 29.41%, 37.5%, and 46.67% better than FF, PSO, GWO, and DA techniques, respectively. The specificity of the introduced model is 8.94% greater than FF, 8.81% greater than PSO, 8.94% greater than GWO, and 9.08% greater than DA algorithms, respectively. Similarly, Table 2, shows the overall performance evaluation using the statlog dataset, whose accuracy by F-DA is 0.22% greater than FF, 0.67% greater than PSO, 0.44% greater than GWO and 1.34% greater than DA algorithms, respectively. Consequently, the precision of the adopted model is 1.22% better than FF, 3.75% better than PSO, 2.47% better than GWO, and 7.79% better than DA algorithms, respectively. Furthermore, the FPR of the implemented F-DA scheme is 0.75% greater than FF, 2.2% greater than PSO, 1.48% greater than GWO, and 4.32% greater than DA algorithms, respectively. Moreover, Table 3, shows the disease prediction using the Wisconsin dataset, whose NPV values of the F-DA based scheme is 23.87%, 24.06%, 23.87%, 24.06% better than FF, PSO, GWO and DA algorithms, respectively. The FDR of the improved F-DA model is 77.56%, 77.67%, 77.56%, and 77.67% greater than FF, PSO, GWO and DA algorithms, respectively. Finally, the F1-score of the implemented model is 91.38% better than FF, 91.95% better than PSO, 91.38% better than GWO, and 91.95% better than DA algorithms. The time comparison of various models is shown in Fig. 6. Hence, from the attained outcomes, it has been clearly shown that the proposed F-DA method can offer better performance on hidden neuron optimization in DBN for disease prediction than previous state-of-the-art methodologies.

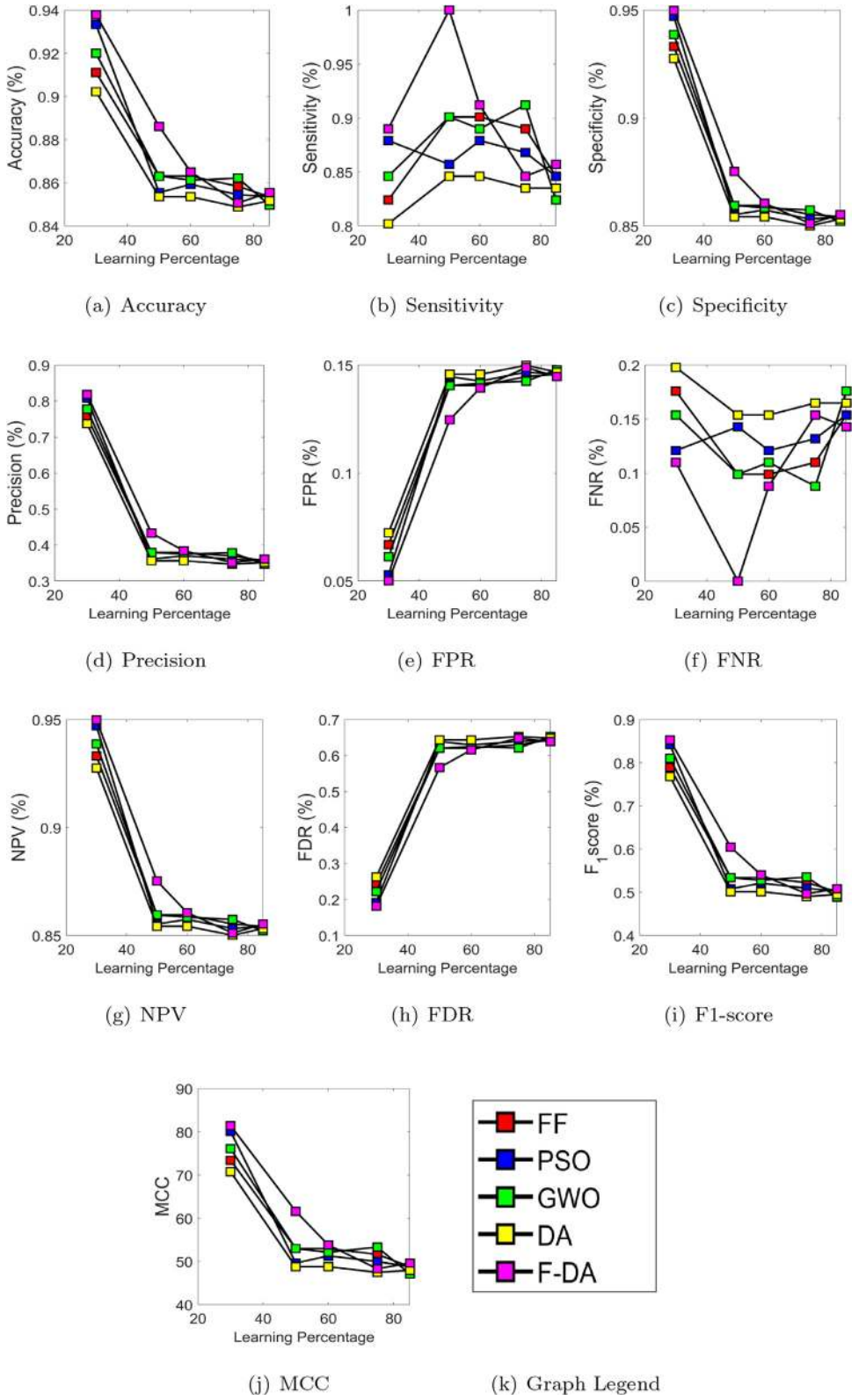


Fig. 4. Performance evaluation of proposed and conventional disease prediction models for different learning percentage using Statlog Dataset (a) Accuracy (b) Sensitivity (c) Specificity (d) Precision (e) FPR (f) FNR (g) NPV (h) FDR (i) F1-score (j) MCC (k) Graph Legend.

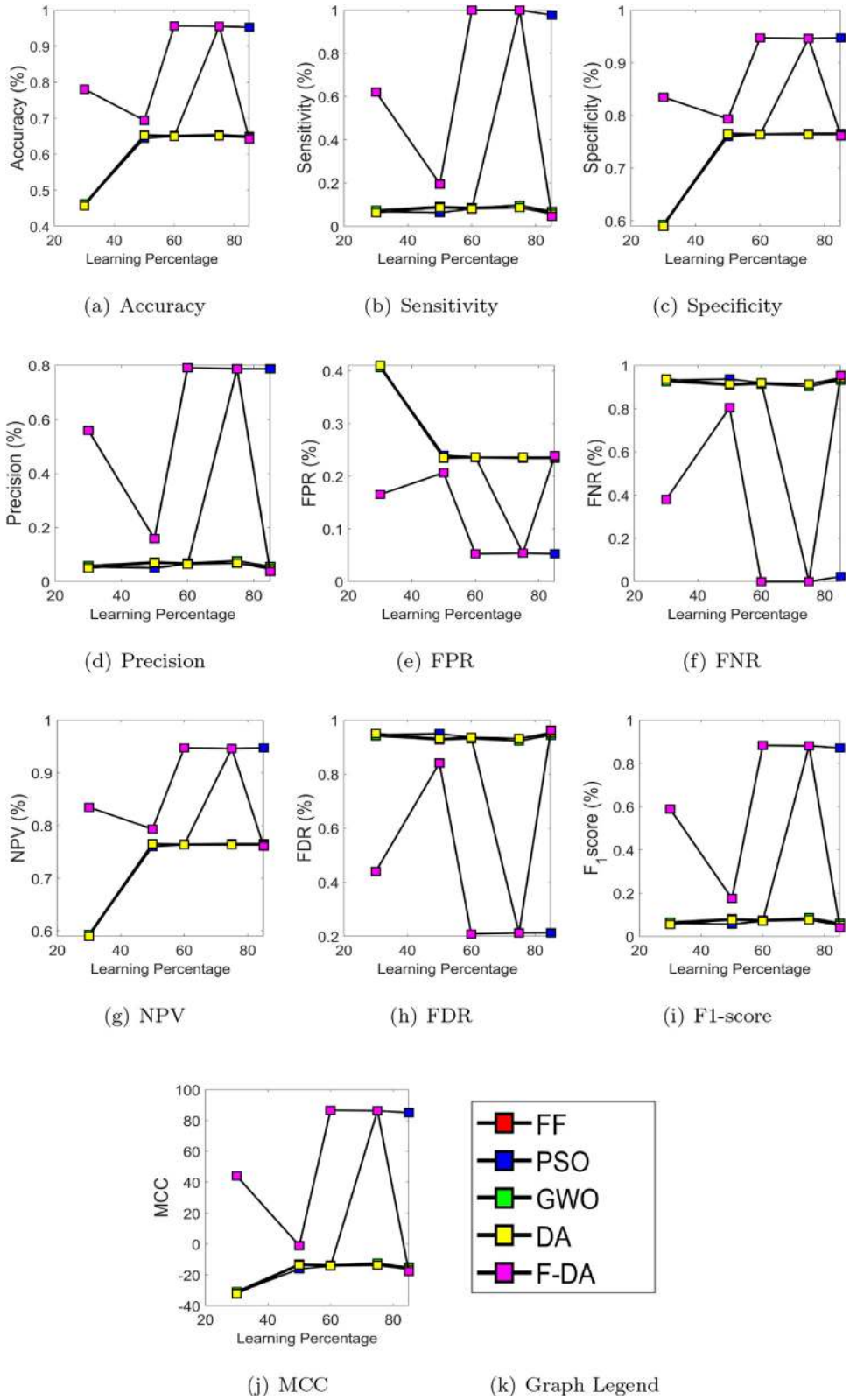


Fig. 5. Performance evaluation of proposed and conventional disease prediction models for different learning percentage using Wisconsin Dataset (a) Accuracy (b) Sensitivity (c) Specificity (d) Precision (e) FPR (f) FNR (g) NPV (h) FDR (i) F1-score (j) MCC (k) Graph Legend.

Table 2
Performance evaluation of the proposed model and traditional models using Statlog Dataset .

Statlog Dataset					
Measures	FF[23]	PSO[24]	GWO[25]	DA[22]	F-DA
Accuracy	0.86316	0.85933	0.86124	0.8535	0.8650
Sensitivity	0.9011	0.87912	0.89011	0.8461	0.9120
Specificity	0.85954	0.85744	0.85849	0.8543	0.8605
Precision	0.37963	0.37037	0.375	0.3564	0.3842
FPR	0.14046	0.14256	0.14151	0.1457	0.1394
FNR	0.098901	0.12088	0.10989	0.1538	0.0879
NPV	0.85954	0.85744	0.85849	0.8543	0.8605
FDR	0.62037	0.62963	0.625	0.6435	0.6157
F1_score	0.5342	0.52117	0.52769	0.5016	0.5407
MCC	0.52963	0.51286	0.52124	0.4877	0.5380

Table 3
Performance evaluation of the proposed model and traditional models using Wisconsin Dataset.

Wisconsin Dataset					
Measures	FF[23]	PSO[24]	GWO[25]	DA[22]	F-DA
Accuracy	0.6516	0.6497	0.6516	0.6497	0.9559
Sensitivity	0.0862	0.0804	0.0862	0.0804	1
Specificity	0.7646	0.7634	0.7646	0.7634	0.9471
Precision	0.0681	0.0636	0.0681	0.0636	0.7909
FPR	0.2353	0.2365	0.2353	0.2365	0.0528
FNR	0.9137	0.9195	0.9137	0.9195	0
NPV	0.7646	0.7634	0.7646	0.7634	0.9471
FDR	0.9318	0.9363	0.9318	0.9363	0.2090
F1_score	0.0761	0.0710	0.0761	0.0710	0.8832
MCC	-.1363	-.1426	-.1363	-.1426	0.8655

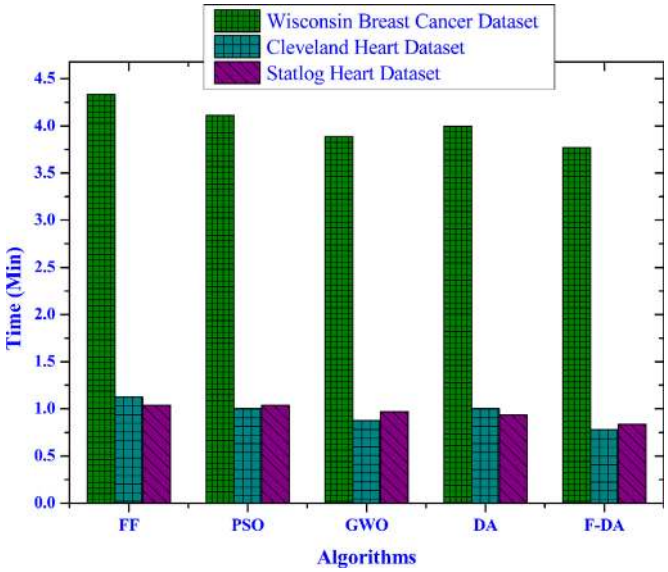


Fig. 6. The time comparison of various models.

6. Conclusion

The proposed disease prediction model is made up of three phases, namely (i) data cleaning (ii) feature extraction, and (iii) classification. The main contribution lies in its use of a modified Dragonfly Algorithm coined as F-DA to optimally tune multiplied weights and ensure that the error among the actual and predicted output is minimized. The model has also been experimentally analyzed against traditional state-of-the-art approaches. The results reveal the superiority of the proposed model yielding enhanced resultant accuracy of 6.96% higher than FF, 6.7% greater than PSO, 6.96% greater than GWO, and 7.22% greater than the DA algorithms, respectively. Moreover, the sensitivity of the F-DA model is observed to be 37.5%

better than FF, 29.41% better than PSO, 37.5% better than GWO, and 46.67% better than DA algorithms, respectively. The results thus justify the effectiveness of the proposed method in disease prediction. Although the results are promising, the datasets being used are relatively smaller in comparison to the big data culture predominant in the present day and age. Thus, the research's future path will be to test the proposed model on large-high-quality datasets to check the effectiveness and reliability of the proposed model.

Declaration of Competing Interest

- All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.
- This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.
- The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript
- The following authors have affiliations with organizations with direct or indirect financial interest in the subject matter discussed in the manuscript:

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was partially funded by the Natural Sciences Research Council of Canada (NSERC) Discovery Grant program (RGPIN-2020-05363) held by Dr. Gautam Srivastava.

References

[1] Lu H, Li Y, Chen M, Kim H, Serikawa S. Brain intelligence: go beyond artificial intelligence. Mobile Netw Appl 2018;23(2):368–75.
[2] Sakai Y, Lu H, Tan J-K, Kim H. Recognition of surrounding environment from electric wheelchair videos based on modified yolov2. Future Generat Comput Syst 2019;92:157–61.
[3] Zhang N, Ding S, Zhang J, Xue Y. An overview on restricted boltzmann machines. Neurocomputing 2018;275:1186–99.
[4] Zhang J, Ding S, Zhang N, Shi Z. Incremental extreme learning machine based on deep feature embedded. Int J Mach Learn Cybern 2016;7(1):111–20.
[5] Shi K, Wang J, Zhong S, Zhang X, Liu Y, Cheng J. New reliable nonuniform sampling control for uncertain chaotic neural networks under markov switching topologies. Appl Math Comput 2019;347:169–93.
[6] Xu X, Lu H, Song J, Yang Y, Shen HT, Li X. Ternary adversarial networks with self-supervision for zero-shot cross-modal retrieval. IEEE Trans Cybern 2019.
[7] Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. IEEE Access 2019;7:81542–54.
[8] Loey M, Smarandache F, M Khalifa NE. Within the lack of chest covid-19 x-ray dataset: anovel detection model based on gan and deep transfer learning. Symmetry (Basel) 2020;12(4):651.
[9] Zhou T, Canu S, Ruan S. An automatic covid-19 ct segmentation based on u-net with attention mechanism. arXiv preprint arXiv:200406673 2020.
[10] Ghoshal B, Tucker A. Estimating uncertainty and interpretability in deep learning for coronavirus (covid-19) detection. arXiv preprint arXiv:200310769 2020.
[11] Nilashi M, bin Ibrahim O, Ahmadi H, Shahmoradi L. An analytical method for diseases prediction using machine learning techniques. Comput Chem Eng 2017;106:212–23.
[12] Zhang C, Zhu L, Xu C, Lu R. Ppdp: an efficient and privacy-preserving disease prediction scheme in cloud-based e-healthcare system. Future Generat Comput Syst 2018;79:16–25.
[13] Chen X, Niu Y-W, Wang G-H, Yan G-Y. Hamda: hybrid approach for mirna-disease association prediction. J Biomed Inform 2017;76:50–8.
[14] Parisot S, Ktena SI, Ferrante E, Lee M, Guerrero R, Glocker B, et al. Disease prediction using graph convolutional networks: application to autism spectrum disorder and alzheimers disease. Med Image Anal 2018;48:117–30.
[15] Weng C-H, Huang TC-K, Han R-P. Disease prediction with different types of neural network classifiers. Telemat Informat 2016;33(2):277–92.
[16] Kumar PM, Lokesh S, Varatharajan R, Babu GC, Parthasarathy P. Cloud and iot based disease prediction and diagnosis system for healthcare using fuzzy neural classifier. Future Generat Comput Syst 2018;86:527–34.
[17] Luo J, Ding P, Liang C, Chen X. Semi-supervised prediction of human mirna-disease association based on graph regularization framework in heterogeneous networks. Neurocomputing 2018;294:29–38.
[18] Sengupta S, Das AK. Particle swarm optimization based incremental classifier design for rice disease prediction. Comput Electron Agric 2017;140:443–51.
[19] Reddy GT, Reddy MPK, Lakshmanna K, Kaluri R, Rajput DS, Srivastava G, et al. Analysis of dimensionality reduction techniques on big data. IEEE Access 2020;8:54776–88.
[20] Gadekallu TR, Khare N, Bhattacharya S, Singh S, Reddy Maddikunta PK, Ra I-H, et al. Early detection of diabetic retinopathy using pca-firefly based deep learning model. Electronics (Basel) 2020;9(2):274.
[21] Reddy GT, Reddy MPK, Lakshmanna K, Rajput DS, Kaluri R, Srivastava G. Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. Evol Intell 2019:1–12.
[22] Jafari M, Chaleshtari MHB. Using dragonfly algorithm for optimization of orthotropic infinite plates with a quasi-triangular cut-out. Eur J Mech A/Solids 2017;66:1–14.
[23] Wang H, Wang W, Zhou X, Sun H, Zhao J, Yu X, et al. Firefly algorithm with neighborhood attraction. Inf Sci (Ny) 2017;382:374–87.
[24] Zhang J, Xia P. An improved pso algorithm for parameter identification of nonlinear dynamic hysteretic models. J Sound Vib 2017;389:153–67.
[25] Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. Adv Eng Softw 2014;69:46–61.

Srinivas Koppu working as Associate Professor in SITE, VIT, India. He received the Phd, M. Tech and B. Tech (CSE) from, VIT, IIIT Allahabad, JNTU Hyderabad, respectively, India. He had published more than 25 articles in reputed international journals. He was a Visiting Professor with the Neusoft Institute, Guangdong, 2018, China. His Research interest also includes Robotics, IoT, Medical Image Processing.

Praveen Kumar Reddy Maddikunta received the B.Tech in CSE from JNT University, M.Tech and Phd in CSE from VIT, Vellore, India. He is currently working as an Assistant Professor in SITE, VIT. He was a Visiting Professor with the Guangdong University of Technology, China, in 2019. He produced more than 50 international/national publications. His area of interests, ML, IoT, Robotics.

Gautam Srivastava (Senior Member, IEEE) received the B.Sc. degree from Briar Cliff University, USA, in 2004, and the M.Sc. and Ph.D. degrees from the University of Victoria, Canada, in 2006 and 2011, respectively. He is currently working as Associate Professor, Brandon University, Canada. He has published more than 60 high-IF articles and area interest data mining and big data.