Received: August 25, 2022. Revised: September 24, 2022.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

A Self-Stabilizing and Parameter-Optimized Deep Learning Model for Predicting Mental Health Disorder

Saranya S[1,2]* **Kavitha N[1]**

[1]Department of Computer Science, Nehru Arts and Science College, Coimbatore, India
[2]Department of Computer Science, Dr. N.G.P. Arts and Science College, Coimbatore, India
* Corresponding author's Email: ssaranyamphil@gmail.com

Abstract: A mental health disorder (MHD) is a behavioural or mental pattern that causes significant distress or impairment of personal functioning. Daily distress faced in social, work, or family activities are associated with MHD which severely affects the person's life style leading to increase of death cases. So, it is very important to detect the MHD at earlier stage itself. Many machine learning (ML) techniques have been suggested to detect MHD based on patient data. But, the ML methods cannot process the large samples efficiently. So, the deep learning (DL) algorithms have been developed for the detection of MHD efficiently. But still, fixed network parameters used in DL methods cannot always perform well for large scale datasets. Also, irrelevant feature in the datasets degrades the performances of DL models. Moreover, stabilization will be another issue of the DL methods. To resolve the above mentioned problems, a chicken swarm intelligence improved self stabilized deep neural network (CSI-ISDNN) is proposed in this article to detect and diagnose the MHD more efficiently. The proposed method employs CSI on training dataset where CS will select the best optimal features and parameters for ISDNN. The fitness value of the CS is then calculated and analysed to find optimised prediction results of ISDNN for MHD with the goal of reducing prediction error. The stabilization of ISDNN is standardized by introducing some additional parameters along with original network parameters. The trained CSI-ISDNN model is used to predict test datasets. Finally, the experimental results exhibit that the CSI-ISDNN model achieves an accuracy of 97.4% on the OSMI mental health database. The accuracy of CSI-ISDNN is superior by 10.18%, 6.21%, 2.63%, and 0.0093% compared to classic model like IGCBA-BPNN, CNN-RNN, CNN-Bi-LSTM, and CSI-MLP respectively.

Keywords: Mental health disorder, Chicken swarm intelligence, Deep neural network, Network parameters, Objective function.

1. Introduction

A person's MHD is the result of variance consequence in brain chemistry. According to a WHO report, MHD reportedly cost the world nearly one trillion dollars in lost productivity. Depression affects the majority of people for a variety of reasons [1, 2]. In India, one in every seven people suffered from a MHD of varying severity, with depression being the most common. Depression was responsible for 33.8% of all MHD [3]. MHD is expanding at a widespread quantity so it is important to upgrade mental fitness services, eliminating the social stigma, in-sighting mental consciousness, and contributing an admittance to disorder treatments. It is very much necessary to detect the MHD at earlier stage and diagnose them with proper treatments [4]. People struggling with MHD can benefit from early discovery, precise diagnosis, and successful treatment. On the other hand, traditional procedures are often labor-intensive and time-consuming. Previously, statistical models were proposed for MHD prediction.

The prominent statistical strategies for forecasting MHD are the autoregressive integrated moving average model (ARIMA) and exponential smoothing (ES), the negative binomial model (NBM), and fractional polynomials [5-7]. On the other hand, these existing methods cannot identify the fundamental association among the distinctive factors and the prediction outcomes. As a result, statistical models' accuracy in predicting MHD is considerably low.

In recent years, there has been a significant growth in the ML application in medical care services especially in MHD prediction. ML is extremely efficient and competent of detecting MHD from complicated datasets allowing data analyst to discover MHD in their early stages. The ML categorization method k-NN [8] efficiently identifies the categories of MHD. However, this strategy does not perform well with larger datasets. A ML strategy based on logistic regression was described [9] to predict mental health problems in mois-adolescence. However, the inability of ML to handle large amounts of data may result in increased costs.

With the CSI model, an enhanced multilayer perceptron is constructed for the early diagnosis of MHD condition [10]. This system includes major attributes evaluated using the lasso feature selection method [11] and its parameters are efficiently calibrated by the CSI approach. This model further addresses the issue of MLP overfitting by employing lasso regression, which finds and excludes redundant and characteristics. The multilayer perceptron's performance is fully reliant on the characteristics that are extremely influential for categorization. Chicken swam intelligence fine-tunes its performance by optimizing certain parameter values. However, this strategy takes a long time to train the data and produces lesser results when compared to other methods. As a result, effective prediction/classification model construction is critical for more accurate detection and treatment of MHD.

Currently, the DNN model is introduced to detect and diagnose the MHD. DNN models developed with the right architectural decisions are critical for achieving the desired classifier performance. To build an effective DNN model, it is critical to consider parameters such as the number of hidden layers, the number of nodes in each layer, and training features such as learning rate and regularization methods. A DNN is best suitable model for the MHD detection. But still, the DNN model has default values that do not always perform well on large-scale datasets. Also, increasing the number of features reduces the model's training performance.

Hence in this article, CSI-ISDNN algorithm is developed for the early detection of MHD. Initially, population of CS is initialised. The position of initialized CS randomly selects features from training dataset and parameters from specified ranges. Then, ISDNN model is trained for each CS with the selected features and parameters. The training accuracy of ISDNN is the fitness value of the CSs. The highest fitness CS is considered as best CS. Next, the position of other CSs is updated towards best CS. The CS then select other set of features and parameters with the updated position. Again, fitness value CSs are found from ISDNN. The updating is continued until all CS obtained same fitness values. At the end of the iteration, ISDNN training model obtained for best accuracy (fitness) with optimal selected feature and parameter is selected as optimal model. The stabilization of ISDNN is standardized by introducing some additional parameters like global and local learning rate along with original network parameters. Finally, the trained optimal CSI-ISDNN method is utilized to classify the test datasets in the testing stage. The different metrics like accuracy, precision, recall and RMSE are used to show the efficiency of CSI-ISDNN recognition on larger datasets.

The optimally selected features and parameter improve the accuracy of the prediction. The stabilization of the network reduces the error rate during training and testing process. The process of optimization and stabilization of classier improve MHD prediction more efficiently.

This manuscript's remaining parts are organised as follows: section 2 evaluates the literature on MHD prediction. The CSI-ISDNN is described in section 3, and its effectiveness is demonstrated in section 4. The study is summarized in section 5, which also includes recommendations for future research.

2. Literature review

ML approaches [12] were utilized to predict MHD problems in adolescence. The dataset of MHD in adolescents was pre-processed and categorised using ML approaches in this model. This method analyses multiple ML algorithms and standard logistic regression using integrated questionnaire and register data to discover which algorithm offers the best results. On the other hand, the classification error was high in this model.

A context DNN framework [13] was presented for detecting depression, a type of MHD. In conjunction with DNN and context information, this model identifies the depression risk context. DNN model inputs included context data relating to predicting depression factors. The DNN connection was subjected to risk prediction regression analysis to estimate the probability of depression. However, it performed worse than other existing systems.

A DL model [14] was created for detecting MHD through social media user content. During this procedure, several posts from Reddit's mental health communities were collected. This algorithm effectively identifies whether a user's post relates to a given MHD by absorbing posting data offered by users. However, model requires additional features because basic DL architecture was difficult to detect the MHD.

A class-contrastive human-interpretable ML approach [15] was presented to predict mortality in severe MHD. In this process, the ML was used to predict the mortality and class-contrastive reasoning to explain the model's performance. However, the major drawback would be the heat-maps reproducibility which results to provide imbalances in the training set.

A comprehensive prediction system called improved global chaos bat back propagation neural network (IGCBA-BPNN) was created [16] to determine and classify the most relevant elements influencing medical professionals' mental health. IGCBA-BPNN not only enhances MHD prediction accuracy, but it also chooses the smallest feature variables. However, IGCBA simply lowers the input dimension of BPNN. The design of the BPNN was not enhanced, and the parameters in the BP system were not optimized properly.

A MHD prediction model [17] was introduced with the ability of deep information expression using convolution neural networks. In this model, psychological health status of students was collected and classified using CNN structure. The model can learn on its own, allowing it to diagnose psychological disorders in students and supports for college psychological counselling and the psychological health team. However, this method does not provide accurate results on larger dataset.

An ensemble DL model 18] was presented for classifying different sorts of MHD. A multiple sentiment features were identified in this model to predict the user's MHD by using the ensemble DL model. The Reddit social networking platform was employed for the data collection and the collected data were classified using CNN and recurrent neural network (RNN) to identify the various MHD. However, this approach had a considerable computational cost.

A hybrid DL approach [19] was developed for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM (Bi-LSTM). The developed model comprises of different modules like collecting the data produced by the online users. Then, the collected dataset was pre-processing to clean data. The feature extraction procedure was modified to produce machine-readable data. Finally, depression categorization was conducted for depressive tweets versus non-depressive tweets. But still, this model has very low convergence rate.

A dual testing technique [20] was developed to achieve a seafarers' mental health status. In this model, the Seafarers' mental health Test scale (SMHT) and fuzzy factor analysis was used to test the data. An intelligent framework was developed for the classification process. Finally, the hybrid scoring mechanism was constructed to achieve the current mental health of seafarers. On the smaller datasets, lower performance was observed.

3. Proposed methodology

In this section, proposed CSI-ISDNN model for detecting the MHD is described briefly. Fig. 1depictsthe overall flow of the presented CSA-ISDNN based MHD recognition system. First, the CSI is used to selects the best optimal features and parameters. Then, the fitness value of the features and parameters is individually calculated. Then, the dataset is divided into training and testing set for training and validating the ISDNN model to provide efficient MHD performance.

3.1 Chicken swarm intelligence (CSI)

CSI is a sophisticated analytical method that has been designed to analyse a range of characteristics associated with chickens, roosters, and hens when they are seeking for food. The CS can be divided into several groups, each of which consists of a rooster and a number of hens and chicks. Distinct chickens follow different rules of motion. There is competition among diverse chickens in a given hierarchical sequence [21, 22]. The CSO implements swarm optimization based on the innate behaviour of chickens in the swarm. On considering the hierarchical structure of CS, the structure with highest fitness value will be considered by roosters, and those with the lowest fitness values i.e., at the bottom of the structure will be considered as chicks. Similarly, formation at middle are considered as hens. The swarm is randomly differentiated into groups, each of which contains a rooster, a group of hens and chicks [23]. The rooster with the highest fitness value has the ability to look for food in more places and on a greater scale. The CSI model is categorized into different steps which are listed below.

3.1.1. Population generation

Initialize the set of CS population. The position

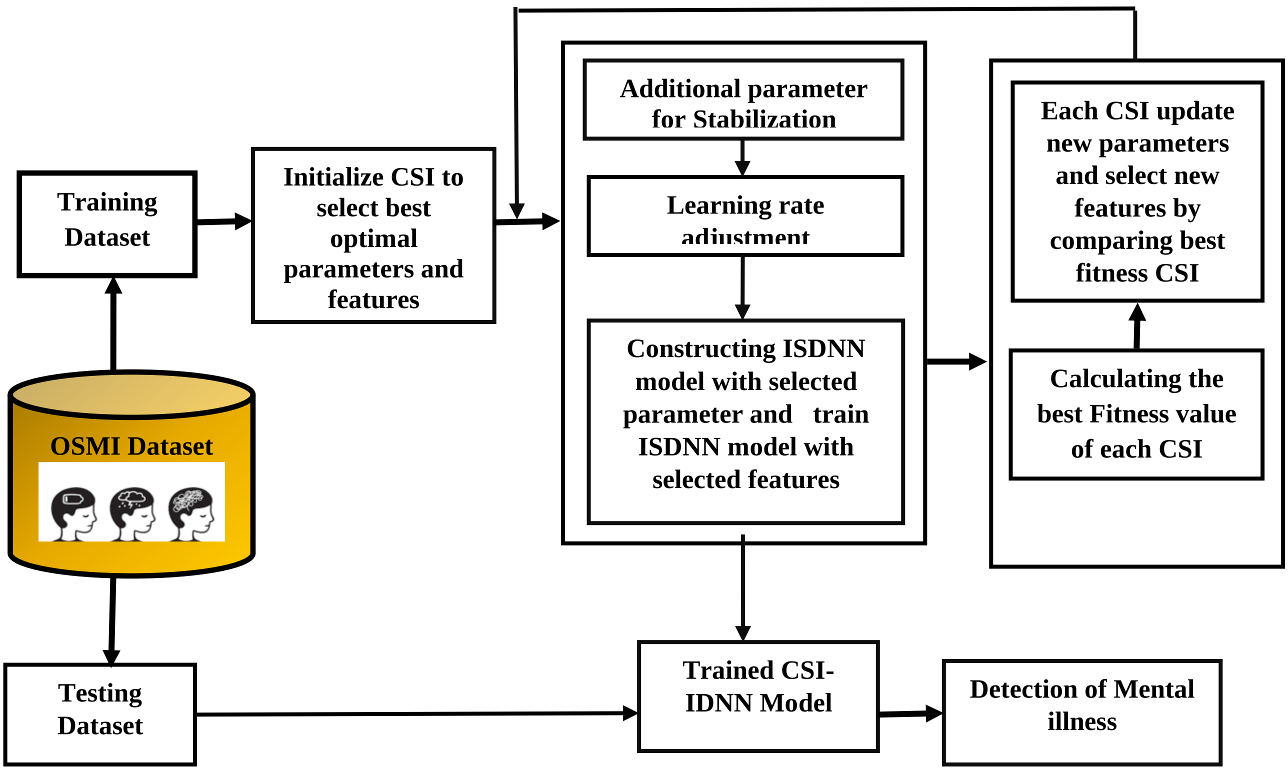

Figure. 1 overall flow of the presented CSA-ISDNN

of initialized CS randomly selects features from training dataset and parameters from specified range values of parameters. ISDNN is constructed based on the selected parameters of CSs. Then, the selected features by the CS are used to construct ISDNN with the goal of less prediction error. The collected features form the training datasets by CS population are represented as $f = (F_1, F_2, F_3, .., F_N)$ andparameters selected by CS is represented $p = P(P_1, P_2, P_3, .., P_N)$ respectively.N denotes the number of CS initialized in the population.

3.1.2. Fitness calculation

The training accuracy of ISDNN for features and parameters selected by each CS is considered as the fitness value. Based on these fitness values, the CS is categorized into different types like Rooster, Chicks and Hens. The chickens with high fitness values are referred to as roosters, while those with low fitness values are referred to as chicks and remaining rages of values are referred to as hens.

3.1.3. Updating the CSI for new parameter and feature selection

The highest fitness CS is considered as best CS. All other CSs other than the highest fitness value will be moving towards or updating towards best CS. The CS then select other set of features and parameters with the updated position iteratively until all CSs obtain same accuracy.

The different position update can be resulted for each chicken i.e., roosters, chicks and hens. Roosters are high fitness CSs can seek new positions i.e., search for optimal parameters and characteristics in a broader range of places than those with lower fitness values as shown in Eq.(1) and Eq. (2).

$$CS_{xp,yp}^{u+1} = CS_{xp,yp}^{u} \times \left(1Rdn(0, \sigma^2)\right) \quad (1)$$

$$CS_{xf,yf}^{u+1} = CS_{xf,yf}^{u} \times (1 + Rdn(0, \sigma^2)) \quad (2)$$

In Eq. (1), Eq. (2) $CS_{xp,yp}^{u+1}$, $CS_{xp,yp}^{u}$ is the roosters position for selecting parameters p, $CS_{xf,yf}^{u+1}$, $CS_{xf,yf}^{u}$ depicts the roosters position for selecting feature f with indexes x and y in $u + 1$ and u instances respectively. $Rdn(0, \sigma^2)$ is the Gauss distribution whose mean value is 0 whose standard deviation will be σ^2. The variable σ^2 is expressed for both parameters and features are given in Eq. (3), Eq. (4)

Received: August 25, 2022. Revised: September 24, 2022.

$$\sigma^2 = \begin{cases} 1, & fit_{xp,yp} \leq fit_{zp} \\ exp \left(\frac{fit_{p,z}-fit_{xp,yp}}{|fit_{xp,yp}|+\varepsilon}\right), & else \end{cases} z\epsilon[1,n], \quad z \neq x,y \tag{3}$$

$$\sigma^2 = \begin{cases} 1, & fit_{xf,yf} \leq fit_{zf} \\ exp \left(\frac{fit_{zf}-fit_{xf,yf}}{|fit_{xf,yf}|+\varepsilon}\right), & else \end{cases} z\epsilon[1,n], \quad z \neq x \tag{4}$$

where $x,y,z \in [1, Rsize]$ and $x \neq z$, $y \neq z$ $Rsize$ will be described as the number of rooster swarms. Where ε will be constant and considered as a small number. Z arbitrarily selected roosters rank form the from the roosters group, $fit_{xp,yp}$ and $fit_{xf,yf}$ will return the fitness value of the associated rooster CS will the fitness value of the corresponding rooster $CS^u_{xp,yp}$ and $CS^u_{xf,yf}$ respectively.

The position of hens is updated towards best CSs for selecting parameter and features for obtaining best fitness than its previous fitness is represented in Eq. (5) and Eq. (6) Its random search will be carried out by hen population is decided by Roosters positions and its previous position [25].

$$\begin{aligned} CS^{u+1}_{xp,yp} = \ & CS^u_{xp,yp} \\ & + T1 \times Rd \\ & \times \left(CS^u_{R_1(xp,yp)} - c^u_{xp,yp}\right) \\ & + T2 \times rd \\ & \times \left(CS^u_{R_2(xp-yp)} - c^t_{xp,yp}\right) \end{aligned} \tag{5}$$

$$\begin{aligned} CS^{u+1}_{xf,yf} = \ & CS^u_{xf,yf} \\ & + T1 \times Rd \\ & \times \left(CS^u_{R_1(xf,yf)} - c^u_{xf,yf}\right) \\ & + T2 \times rd \\ & \times \left(CS^u_{R_2(xf,yf)} - c^t_{xf,yf}\right) \end{aligned} \tag{6}$$

Where, Rd is a persistent value ranging form [0,1]. $CS^u_{R_1(xp,yp)}$ and $CS^u_{R_1(xf,yf)}$ is the roosters position followed up by hens. $CS^u_{R_2(xp,yp)}$ and $CS^u_{R_2(xf,yf)}$ is the any individuals position except the available hen [24]. $T1$ and $T2$ are the learning factors for the parameters and features derived in Eq. (7-10).

$$T1(P) = \exp\left(\frac{fit_{xp,yp}-fit_{R_1(xp,yp)}}{abs(fit_{xp.yp})+\varepsilon}\right) \tag{7}$$

$$T2\ (P) = \exp\left(fit_{R_2(xp,yp)} - fit_{xp,yp}\right) R_1 \in [1,2,\dots n], R_2 \in [1,2,\dots n], R_1 \neq R_2 \neq x,y \tag{8}$$

$$T1\ (F) = \exp\ ((fit_{xf,yf} - fit_{R_1(xf,yf)}) / (abs(fit_{xf,yf}) \tag{9}$$

$$T2\ (F) = \exp\left(fit_{R_2(xf,yf)} - fit_{xf,yf}\right) R_1 \in [1,2,\dots n], R_2 \in [1,2,\dots n], \ R_1 \neq R_2 \neq x,y \tag{10}$$

In the above equation, fit_{R_1} is the fitness of the rooster, which is followed by hens. The fitness of any individual i^th hen is represented by fit_{R_2}. R_1 is the index of the rooster preceded by the hens, R_2 is the score of any individual excluding the i^th hen, and both R_1 and R_2 differ from the present index. The mobility of chicks for food will be centered on their mother hen.

The position of chick swarm are updated for parameter and features selection is shown in Eq. (11), Eq. (12) as

$$CS^{u+1}_{xp,yp} = CS^u_{xp,yp} + CF \times \left(CS^u_{zp} - CS^u_{xp,yp}\right) z \in [1,2,\dots n] \tag{11}$$

$$CS^{u+1}_{xf,yf} = CS^u_{xf,yf} + CF \times \left(CS^u_{zf} - CS^u_{xf,yf}\right) z \in [1,2,\dots n] \tag{12}$$

Where CF is the chick coefficient ranging from0 to 2. CS^u_{zp} and CS^u_{zf} represent the hens position, i.e., the acquired parameter and features preceded by the chicks. Its arbitrary search will be performed out by chicks whose population is determined by the hens positioning and its prior position.

3.1.4. Termination process

The position updating is continued until all CS obtained same fitness values or else the process is repeated until the result of expected fitness value is determined. At the end of the iteration, ISDNN training model obtained for best accuracy (fitness) with optimal selected feature and parameter is selected as optimal model.

3.2 Architecture of CSI-ISDNN

The Fig. 2 provides the structure of CSI-ISDNN model.

3.3 Improving ISDNN by self-stabilization

With their original parameters and features, Self-Stabilized DNN training models are efficiently used to train and test the MHD dataset. Consequently, the structures are trained using mini batch Stochastic gradient descent. For better optimal result with SGD, the first and last learning levels must be tuned as well as an optimum learning rate schedule should be

Received: August 25, 2022. Revised: September 24, 2022.

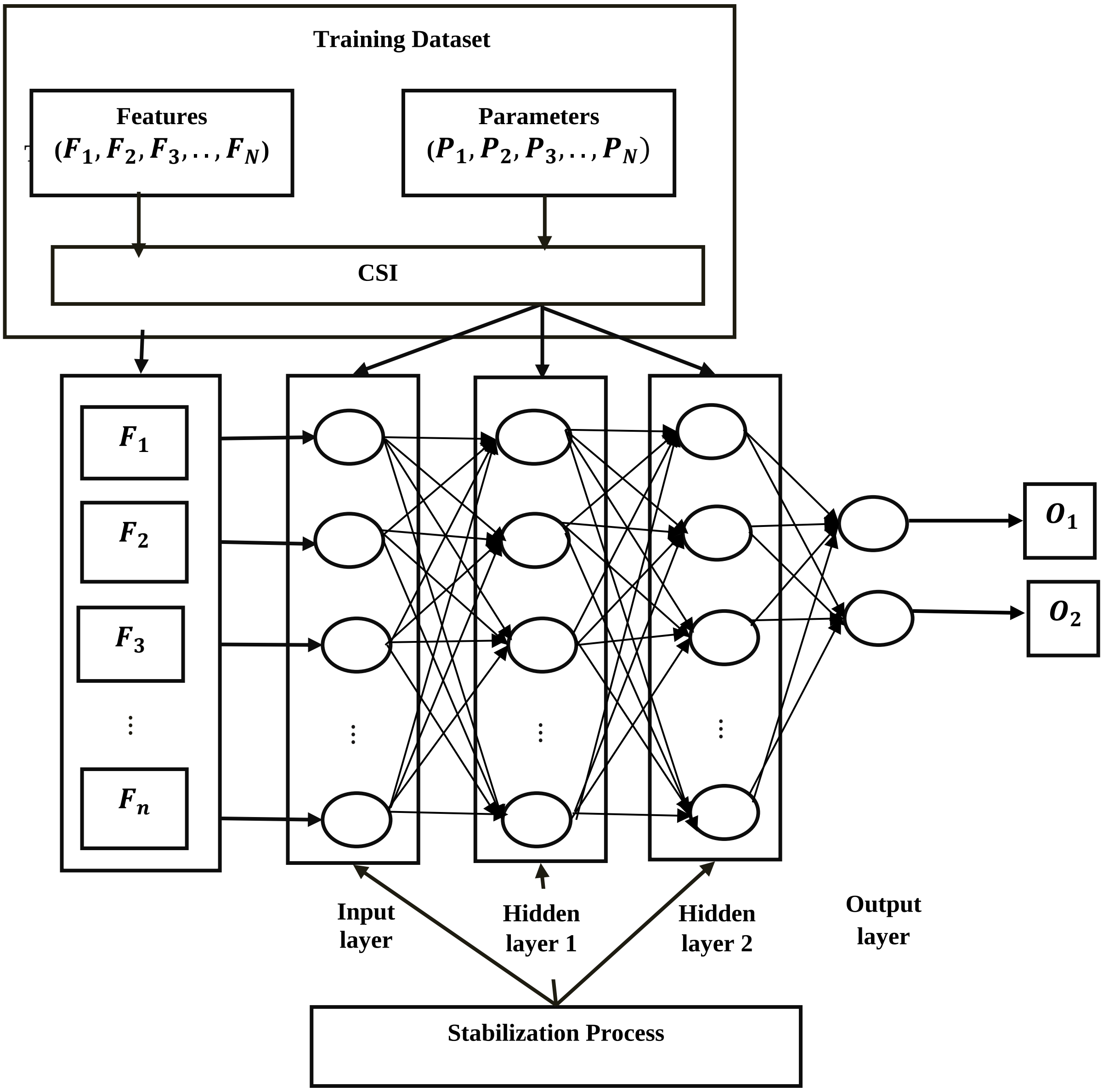

Figure. 2 Stricture of CSI-ISDNN model

designed effectively. Finding the initial value will be the most difficult as lower values might result in delayed learning and high values can create inconsistency and variance in training. To circumvent this, the self-stabilized DNN is enhanced with multiple learning algorithms such as Global learning rate scheduling methods and local per-dimension learning rate scheduling methods [25]. The SDNN is substantially enhanced by employing these methods, and the parameters and features are not shrunk making the entire model less susceptible to prediction mistakes in MHD identification

In DNN model an extra scalar parameters and features is applied to each layer. It is designed to balance the SGD training and it is trained along with the original network parameters. These additional variables are constructed as a per-layer stabilizer with data determining whether to reduce or enhance their value in order to advance towards reducing the target function $\mathcal{F}$. In its most basic form, the variables of a DNN layer with a scalar γgiven in Eq. (13),

$$D = \phi\,(\gamma \times Pi + c) \tag{13}$$

$$\frac{\partial \mathcal{F}}{\partial i} = \gamma\, P^{S} \frac{\partial \mathcal{F}}{\partial D} \tag{14}$$

In Eq. (15), the derivative with reference to variable is computed as

$$\frac{\partial \mathcal{F}}{\partial \gamma} = \frac{\partial \mathcal{F}}{\partial D} \times \frac{\partial D}{\partial \gamma} = \frac{\partial F^S}{\partial D} Pi \qquad (15)$$

So the gradient for parameter γcan be expressed in Eq. (16),

$$\frac{\partial \mathcal{F}}{\partial \gamma} = \frac{1}{\gamma}\left(\frac{\partial \mathcal{F}}{\partial \gamma}\right)^S i = \frac{1}{\gamma} < \left(\frac{\partial \mathcal{F}}{\partial \gamma}\right), i > \qquad (16)$$

So, γ is computed in Eq. (17),

$$\gamma_{u+1} = \gamma_u - \frac{\eta}{\gamma} < \frac{\partial \mathcal{F}}{\partial i}, i > \qquad (17)$$

The variation in γ corresponds to how the input layer i relates to the target function derivative with reference to that of input variable. If the target function is raised by levelling $i$up, γ will also be raised. If the objective function is enhanced by using a lower i, then γ will be reduced. If the input's comparative orientation and gradient are both random then it will stabilize. This is likely to occur near the point of convergence. The stride size for parameterv_{xy} identirelyregulatedby thecurrentvalue of $\gamma$is stated in Eq. (18) as

$$m_{xy} = \frac{\partial \mathcal{F}}{\partial v_{xy}} = \gamma \frac{\partial \mathcal{F}}{\partial j_x} i_y \qquad (18)$$

In practice, $exp(\gamma)$ is utilized instead of γ in this research work. Also, the $\gamma = 0$ is initialized. As a result, the effective stabiliser must be positive and deterioration more slowly as it equalizes to 0 which is depicted in Eq. (19),

$$D = \phi\,(\exp(\gamma) \times Pi + c) \qquad (19)$$

This network is less susceptible to learning rate decision and parameter adjustments. Based on the selection of optimal parameters and features form CSI, an ISDNN is designed to assist the network in scaling the parameters at the start of training to be closer to the local optima of the objective function. The dataset is split into training and testing phase after GAN execution is fully completed. The actual network settings are utilized to train the ISDNN model on the training set which then extracts the most effective parameters and features and detects the particular illness. Further, the trained model is tested using the test set to analyse the efficiency of CSI-ISDNN detection.

3.4 Algorithm for the CSI-ISDNN model to predict the MHD

The process of CSI-ISDNN is depicted in algorithm 1.

Input: OSMI dataset(ds), a is the input parameter, b class parameter of the pertinent data.
Output: Determining in presence or absence of MHD
Initialize a set of parameters and features for ISDNN
For all (a, b)) $\in ds$ do

a. Fed the network with the input a
b. Calculate the network output a
c. Calculate the discrepancy between the executed and resulted output
d. Stop for

Determine the RMSE for the f
End for
Compute the RMSE of the $f(x)$
Adding of total training pattern weights
Upgrade the weights of CSI
Give the appropriate parameters for the n-person population of chickens.
Calculate the fitness value for all chickens $(n)$and set $t = 0$.

While $t < max - gen$

e. If (t % G $==$ 0) then
   Sort the chicks' fitness values in order.
   Develop a hierarchal structure in the swarm
   Divide the swarm into different section
   Establish the relationships among the hens and rooster chicks in a group.
f. End if

For $j = 1$ to n

g. IF j $==$ rooster then

Upgrade the position of the rooster for p and f using the Eq. (1) and Eq. (2)

h. Else if$j$ $==$ hen then
i. Upgrade the location of the superior hen p and $f$using the Eq. (5) and Eq. (6)
j. Else if j == chick then improved position of chick using the Eq. (11) and Eq. (12)
k. End if

Add all weights over all training patterns
Perform one update step of the minimization approach
Compute the new search solution
IF the new solution is better than the previous, update it
Map the selected parameters by using ISDNN to reduce the prediction errors
Train the ISDNN classifier;

End for
Obtain the trained model and validate it
Test the model to detect the MHD
End while
End

By this process, it is concluded that the CSI-ISDNN algorithm delivers optimum predictions of mental disease issues at the earlier stage in order to improve the quality of life for patients and lower prediction error.

Table 1. Dataset description

Attributes	**Type**
Age	Numeric
Year	Nominal
Age-Group	Categorical
Gender	Categorical
Sought Treatment	Numeric
Describe Past Experience	Categorical
Prefer Anonymity	Categorical
Rate Reaction to Problems	Categorical
Negative Consequences	Categorical
Location	Categorical
Access to information	Numeric
Insurance	Categorical
Diagnosis	Categorical
Discuss Mental Health Problems	Categorical
Responsible Employer	Categorical
Family History	Categorical
Company Size	Categorical
Disorder Notes	Categorical
Disorder	Numeric
Primarily a Tech Employer	Numeric

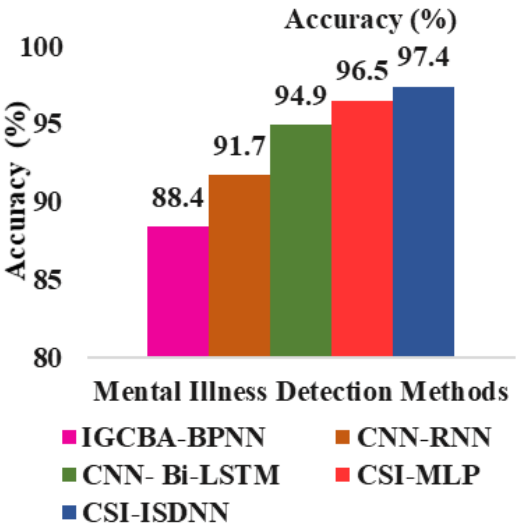

Figure. 3 Performance Evaluation of accuracy for MHD detection

4. Dataset description

The open sourcing mental illness (OSMI) Mental Health in Tech Survey 2019 provided the dataset for this prediction, which contains 4218 cases and 20 attributes [26]. The dataset includes information about working people, intends to raise MHD awareness at an early stage, which will be advantageous to both employees and employers. Table 1 lists the attributes included in the OSMI dataset

5. Result and discussion

This section discusses about performance analysis of MHD detection using proposed CSI-ISDNN and it is implemented using python code. The dataset description is given in section 4. The developed model performance is assessed with various existing detection techniques like IGCBA-BPNN [16], CNN-RNN [18] CNN-Bi-LSTM [19] CSI-MLP [10]. For determine the performance efficiency, the proposed and existing methods are evaluated with different metrics which are listed below.

5.1 Accuracy

It is the proportion of exact detection of MHD over the total samples or labels tested. It is computed as

$$Accuracy = \frac{True\ Positive\ (TP) + True\ Negative\ (TN)}{TP + TN + False\ Positive\ (FP) + False\ Negative\ (FN)} \quad (20)$$

TP measures an outcome where the CSI-ISDNN exactly classifies the MHD as MHD. FP measures an outcome where the CSI-ISDNN inexactly classifies the MHD labels as non- MHD.FN measures an outcome where the CSI-ISDNN inexactly classifies the non- MHD labels as MHD. TN measures an outcome where the CSI-ISDNN exactly classifies the non- MHD labels as non- MHD.

Fig. 3 compares the accuracy of various MHD detection models used to forecast the existence of MHD. The X-axis represents the detecting techniques, and the Y-axis represents the accuracy range. Compared to IGCBA-BPNN, CNN-RNN, CNN-Bi-LSTM, and CSI-MLP, the accuracy of CSI-ISDNN is superior by 10.18%, 6.21%, 2.63%, and 0.0093%, respectively. The cause is that existing models do not adequately address the issue of class imbalance. The parameters and characteristics are heavily predetermined in the dataset while employing

CSI-ISDNN. Thus, the proposed model achieves the maximum accuracy when compared to other models.

5.2 Precision

It is computed according to the amount of correctly detected MHD at TP and FP.

$$Precision = \frac{TP}{TP+FP} \tag{21}$$

Fig. 4 shows the accuracy rate attained by several prediction models for identifying MHD. The evaluation shows that the suggested CSI-ISDNN outperforms the IGCBA-BPNN, CNN-RNN, CNN-Bi-LSTM, and CSI-MLP by 6.86%, 5.03%, 2.93%, and 1.02%, respectively. This is due to the integration of ISDNN and CSI's with lesser hierarchical order which optimizes the MHD prediction. It is clear from this analysis that the proposed CSI-ISDNN offers higher precision than other current approaches for MHD prediction.

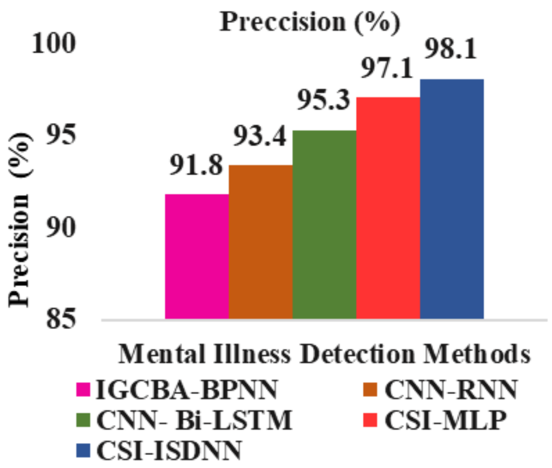

Figure. 4 Comparison of precision values for MHD detection

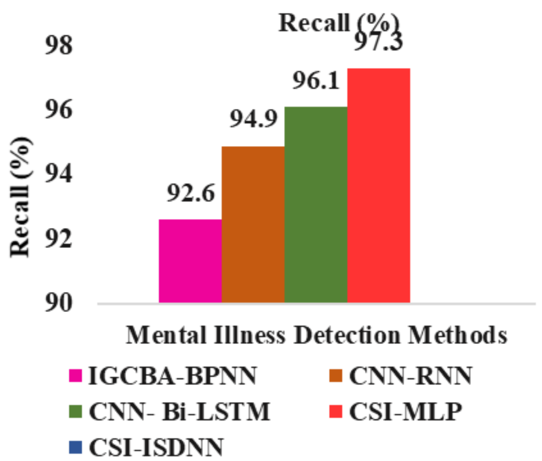

Figure. 5 Comparison of recall values for MHD detection

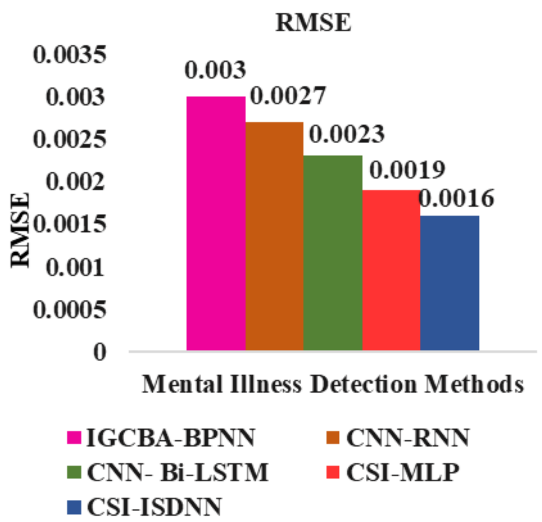

Figure. 6 Evaluation of RMSE for MHD detection

5.3 Recall

It is calculated according to detection of the MHD at TP and FN rates.

$$Recall = \frac{TP}{TP+FN} \tag{22}$$

Fig. 5 shows a comparison of recall metrics for several prediction models used to identify MHD. According to the analysis above, CSI-ISDNN has a larger specificity than IGCBA-BPNN, CNN-RNN, CNN-Bi-LSTM, and CSI-MLP by 6.04%, 3.47%, 2.18%, and 0.0092%, respectively. Applying ISDNN, which manages the over fitting issue by reducing penalty parameters, reduces the attributes in the dataset. This proposed work adapts CSI based on its fitness value during the training phase as opposed to employing random assignment of parameter values. As a result, the recall's results are better than those of other current techniques.

5.4 Root mean square error

RMSE is the defined as the accurate identification of the prediction error throughout the research work. RMSE is frequently employed and is regarded as a superior all-purpose error metric for numerical forecasts. It is calculated as

$$RMSE = \sqrt{\sum_{j=1}^{N}(CS_{xp,yp}^{u} - CS_{xf,yf}^{u})^2/N} \tag{23}$$

In above Eq. (23), $CS_{p,x}^{u}$ and $CS_{f,y}^{u}$ depicts the location of x^{th} and y^{th} dimension of parameter p and feature f in u. N is the total of CSs populations.

Received: October 18, 2021. Revised: December 20, 2021.

In above Eq. (23), $CS_{p,x}^{u}$ and $CS_{f,y}^{u}$ depicts the location of x^{th} and y^{th} dimension of parameter p and feature f in u. N is the total of CSs populations.

The comparison based on RMSE derived by classification models used to diagnose MHD is explored in Fig. 6. The error rate of CSI-ISDNN is 0.46%, 0.40%, 0.30%, and 0.15% lesser than that of the existing methods like IGCBA-BPNN, CNN-RNN, CNN-Bi-LSTM and CSI-MLP. The proposed model efficiently addresses the both over fitting and class imbalance by concentrating on highly contributing qualities and the ideal distribution of parameter values such as weight and bias in DNN models. Low percentages of classes with MHD are a problem for conventional models. Thus, they produce the highest error rate when compared to CSI-ISDNN.

6. Conclusion

This research utilized a newly established technique called CSI-ISDNN to perform early identification of MHD. This model employs CSI on training dataset where CS will select the best optimal features and parameters for ISDNN. The fitness value of the CS is then calculated and analyzed to find optimized prediction results of ISDNN for MHD with the goal of reducing prediction error. The stabilization of ISDNN is standardized by introducing some additional parameters along with original network parameters. The trained CSI-ISDNN model is used to predict test datasets. In addition, the suggested CSI-ISDNN generated a significant contribution with the greatest accuracy of 97.4% for MHD detection in OSMI dataset 2019, compared to other three traditional classifications methods with reduced detection error. In future, online training or incremental training method will be developed in transfer learning to reduce training time effectively.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

Conceptualization, methodology, software, validation, Saranya; formal analysis, investigation, Kavitha; resources, data curation, writing—original draft preparation, Saranya; writing—review and editing, Kavitha; visualization; supervision, Kavitha

References

[1] U. S. Reddy, A. V. Thota, and A. Dharun, "ML techniques for stress prediction in working employees", In: *Proc. of 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pp. 1-4, 2018.

[2] S. E. Hanisch, C. D. Twomey, A. C. Szeto, U. W. Birner, D. Nowak and C. Sabariego, "The effectiveness of interventions targeting the stigma of mental illness at the workplace: a systematic review", *BMC Psychiatry*, Vol. 16, No. 1, pp. 1-11. 2016.

[3] T. Theorell, A. Hammarström, G. Aronsson, L. T. Bendz, T. Grape, C. Hogstedt, and C. Hall, "A systematic review including meta-analysis of work environment and depressive symptoms", *BMC Public Health*, Vol. 15, No. 1, pp. 1-14, 2015.

[4] P. C. Gronholm, N. Chowdhary, C. Barbui, J. D. Munshi, K. Kolappa, G. Thornicroft, ... and T. Dua, "Prevention and management of physical health conditions in adults with severe mental disorders", *WHO Recommendations. International Journal of Mental Health Systems*, Vol. 15, No. 1, pp. 1-10, 2021.

[5] M. H. Browning, L. R. Larson, I. Sharaievska, A. Rigolon, O. McAnirlin, L. Mullenbach, ... and H. O. Alvarez, "Psychological impacts from COVID-19 among university students: Risk factors across seven states in the United States", *PloS One*, Vol. 16, No. 1, p. e0245327, 2021.

[6] X. Yang, M. Jin, and L. Zheng, "The prediction model of college students life stressor on mental health", *China J Health Psychol*, Vol. 26, pp. 775-8, 2018.

[7] J. Margraf, X. C. Zhang, K. L. Lavallee, and S. Schneider, "Longitudinal prediction of positive and negative mental health in Germany, Russia, and China", *PLoS One*, Vol. 15, No. 6, p. e0234997, 2020.

[8] P. Morillo, H. Ortega, D. Chauca, J. Proaño, D. V. Huanga and M. Cazares, "Psycho web: a machine learning platform for the diagnosis and classification of mental disorders", In: *Proc. of International Conference on Applied Human Factors and Ergonomics*, pp. 399-410, 2019.

[9] R. Garriga, J. Mas, S. Abraha, J. Nolan, O. Harrison, G. Tadros and A. Matic, "Machine learning model to predict mental health crises from electronic health records", *Nature Medicine*, Vol. 28, No. 26, pp. 1240-1248, 2022.

[10] S. Saranya and N. Kavitha, "Robust Feature Selection With Chicken Swarm Intelligence Improved Multilayer Perceptron For Early Detection Of Mental Illness Disorder", *Journal of Theoretical And Applied Information*

Received: October 18, 2021. Revised: December 20, 2021.

Technology, Vol. 100, No. 10, pp. 3337-33345, 2022.

[11] Y. Jiang, Y. He and H. Zhang, "Variable selection with prior information for generalized linear models via the prior LASSO method", *Journal of the American Statistical Association*, Vol. 111, No. 513, pp. 355-376, 2016.

[12] E. Tate, R. C. McCabe, H. Larsson, S. Lundström, P. Lichtenstein, and R. K. Halkola, "Predicting mental health problems in adolescence using machine learning techniques", *PloS One*, Vol. 15 No. 4, p. e0230389, 2020.

[13] J. W. Baek and K. Chung, "Context deep neural network model for predicting depression risk using multiple regression", *IEEE Access*, Vol. 8, pp. 18171-18181, 2020.

[14] J. Kim, J. Lee, E. Park, and J. Han, "A DL model for detecting mental illness from user content on social media", *Scientific Reports*", Vol. 10, No. 1, pp. 1-6, 2020.

[15] S. Banerjee, P. Lio, P. B. Jones, and R. N. Cardinal, "A class-contrastive human-interpretable machine learning approach to predict mortality in severe mental illness", *Npj Schizophrenia*, Vol. 7, No. 1, pp. 1-13, 2021.

[16] X. Wang, H. Li, C. Sun, X. Zhang, T. Wang, C. Dong, and D. Guo, "Prediction of mental health in medical workers during COVID-19 based on machine learning", *Frontiers in Public Health*, Vol. 9, p. 697850, 2021.

[17] P. Huang, "A Mental Disorder Prediction Model with the Ability of Deep Information Expression Using Convolution Neural Networks Technology", *Scientific Programming*, Vol. 2022, pp. 1-8, 2022.

[18] S. Nasrullah and A. Jalali, "Detection of Types of Mental Illness through the Social Network Using Ensembled DL Model", *Computational Intelligence and Neuroscience*, p. 2022, 2022.

[19] H. Kour and M. K.Gupta, "An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM", *Multimedia Tools and Applications*, pp. 1-37, 2022.

[20] Z. Zhen, R. Wang, and W. Zhu, "A deep learning based method for intelligent detection of seafarers mental health condition", *Scientific Reports*, Vol. 12, No. 1, pp. 1-11, 2022.

[21] X. Meng, Y. Liu, X. Gao and H. Zhang, "A new bio-inspired algorithm: chicken swarm optimization", In: *Proc. of International Conference in Swarm Intelligence*, pp. 86-94, 2014.

[22] C. Qu, S. A. Zhao, Y. F, and W. He, "Chicken swarm optimization based on elite opposition-based learning", *Mathematical Problems in Engineering*, p. 2017, 2017.

[23] F. Chebihi, M. E. Riffi, A. Agharghor, S. C. B. Semlali, and A. Haily, "Improved Chicken Swarm Optimization Algorithm to Solve the Travelling Salesman Problem", *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 12, No. 3, pp. 1054-1062, 2018.

[24] X. Huang, C. Ye, and J. Zheng, "Chicken swarm optimization algorithm of hybrid evolutionary searching strategy", *Computer Engineering and Applications*, Vol. 54, No. 7, pp. 176-181, 2018.

[25] P. Ghahremani and J. Droppo, "Self-stabilized deep neural network", In: *Proc. of 2016 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 5450-5454. 2016.

[26] https://www.kaggle.com/osmihelp/osmi-mental-health-in-tech-survey-2019.