Bulletin of Electrical Engineering and Informatics
Vol. 13, No. 5, October 2024, pp. 3726~3736
ISSN: 2302-9285, DOI: 10.11591/eei.v13i5.7309

Prediction of mental illness using ensemble model and grid search hyperparameter optimization

Srinath Kudlapura Shivaiah[1], Kiran Krishnappa[1], Naveen Kumar Boraiah[2], Punjalkatte Deepa Shenoy[1], Venugopal Kuppanna Rajuk[3]
[1]Department of Computer Science and Engineering, University of Visvesvaraya College of Engineering, Bengaluru, India
[2]Department of Information Science and Engineering, Sahyadri College of Engineering and Management, Mangaluru, India
[3]Department of Computer science and Engineering, Bangalore University, Bengaluru, India

Article Info

Article history:

Received Aug 2, 2023
Revised Mar 13, 2024
Accepted Mar 28, 2024

Keywords:

Depression anxiety stress scale
Gradient descent neural network
Large scale group decision making
Long short-term memory
Machine learning algorithms
Mental illness classification

ABSTRACT

The early prediction of mental illnesses reduces the severity of the disease. The symptoms like poor concentration, unstable energy of the body, pressure, and loss of interest cause depression. A large-scale group decision making (LSGDM) method is proposed in this paper along with the ensemble classifier model by combining convolutional neural network (CNN), recurrent neural network (RNN), and long short-term memory (LSTM) for effective classification of depression, anxiety and stress (DAS) levels. The data is collected from the depression anxiety stress scale-42 (DASS-42) dataset for efficient classification and predictions of mental health problems. The min-max normalization is used to pre-process the data, and the feature selection is done for extracting informative features. The extracted features are provided as input to the ensemble classifier. The proposed LSGDM model maximizes the classification accuracy with the help of grid search CV hyperparameter tuning, and results in an accuracy of 98.88%, precision of 98.21%, recall of 99.62%, F1-Score of 98.90%, and MCC of 99.41%. The proposed LSGDM method gives superior results when compared to the existing machine learning (ML) based ensemble model, a principal component analysis along with modified fast correlation based filtering (PCA-mFCBF), and LSTM based RNN (LSTM- RNN).

Corresponding Author:

Srinath Kudlapura Shivaiah
Department of Computer Science and Engineering, University of Visvesvaraya College of Engineering
Bengaluru, India
Email: solansrinath@gmail.com

1. INTRODUCTION

The process of predicting the problem of mental illness is a complex task that includes the problems of depression, anxiety, and stress at various levels. Depression is a difficult problem that affects mental health randomly, alongside affecting acting, thinking, and feeling, and potentially leads to suicide attempts in some cases. The physical and emotional problems also arise because of mental disorders [1], accompanied by the symptoms like stress, nervous [2], uneasy feelings and anxiety [3]. The input data is collected from depression anxiety stress scale-42 (DASS-42) dataset, resulting in efficient classifications and predictions of mental health problems. Further, mental illness issues occur in employees due to work pressure, discrimination, inequality, and insecurity in the job. The prediction of mental problems is a complex task for psychologists during the time of pandemic [4]. The automatic prediction technique using deep learning algorithms is useful, but some complexity results in the prediction of mental problems on social media data [5] due to the extraction of subjective data from social media, having lesser efficiency which leads to negative prediction that is called fake depression [6]. The semi-supervised learning algorithms are used for training the data and the convolutional neural network (CNN), graph neural network (GNN), recurrent neural network (RNN) and long short-term memory (LSTM) algorithms are used for efficient classification with unlabeled and labelled data on social media platforms [7]. The machine learning models like eXtreme gradient boosting (XG-Boost) result in the best predictions of mental illness problems [8], [9]. The mental disorders in children and adults can identify based on the symptoms of anxiety, stress, and depression by using machine learning techniques without depending on the mood of the users [10]. Mental illness like depression mainly affects pregnant women and leads to severe problems hence reduction of risk factors with the optimal cost is done by using machine learning algorithms [11]. The ML approaches like random forest (RF), and k-nearest neighbours (KNN) can efficiently predict stress levels from the extracting features of the input DASS-42 dataset compared to the existing Machine learning aproaches [12]-[15]. The cognitive training technique is presented to evaluate high-stress prediction performance on machine learning algorithms [16]. The process of analysing human behaviour, and recognizing depression are predicted efficiently by using machine learning algorithms [17]. The machine learning algorithms result in efficient prediction of depression by analysing [18] various levels of severities like mild, severe, moderate, extremely severe, and normal [19] of input DASS-42 dataset [20]. Information about appropriate approaches to shorten mental health problems has enhanced as a outcome of increased research in a field of mental health. However, it is still unclear and unknown what exactly causes mental diseases. Therefore, the rise in mental health issues and the demand for high-quality medical care have prompted research into the utilization of DL in mental health problems. In order predict mental health issues early, this research presents a thorough assessment of deep learning algorithms.

Sumolang and Maharani [21] presented the Bi-LSTM method to identify the depression of Twitter users with the extraction technique of wor2vec. The dataset used has been obtained from DASS-42. The entire data collected from the 159 Twitter users were classified and pre-processed into normal and depressed categories based on the DASS-42 dataset. The Bi-LSTM model classifies the data mostly into positive labels than negative labels due to the presence of unbalanced data labels where the hyper parameters consist of epoch-8, 256 neurons, and 32-size in the hidden layer. The Bi-LSTM does not result in efficient predictions in case of fewer datasets and making the data for analysis was a complex task. Rahman and Halim [22] presented a spectral, temporal, and mel frequency cepstral coefficient (MFCC) to extract the features and the correlation between DAS. The efficiency feature extraction and understanding of the drawing signal of the MFCC were proposed. The classification was improved using the Bi-LSTM network and the emotions were predicted efficiently by the in-depth investigation. The drawing, and writing signal features were extracted efficiently and resulted in low and high-frequency signal extraction by using the MFCC method. The main limitation of MFCC does not give efficient results in the case of digital handwriting and drawing processes due to blur text. Sun *et al*. [23] proposed a machine learning based ensemble model known as long to short assessments that assess people's behavior, attitude, mental and physical states. A long to short model was evaluated on adults initially by using DASS-42 dataset. In order to forecast the ground truths, the feature selection technique such as minimum redundancy maximum relevance (MRMR) utilizing mutual information quotient (MIQ) was utilized to determine which DASS-42 questionnaire questions were most crucial for predicting anxiety. Finally, a small anxiety evaluations were implemented into a web tool for future assessment participants to use. The proposed approach has acquired better accuracy in predicting anxiety levels. However, the data imbalance and small size data were the limitations of this work which is difficult to train the ML model to obtain high robustness and accuracy.

Nolazco-Flores *et al*. [24] suggested a merging feature technique of temporal, statistical, kinematic, spectral, and cepstral domain features to identify the state of mood. These features were chosen by utilizing a principal component analysis (PCA) along with modified fast correlation based filtering (mFCBF) from EMOTHAW database for DASS levels. The automated ML model was used to train and test ten plain text and ensemble classifier. The proposed approach achieved highest accuracy in identifying two possible grades of mood severities and achieved better performance in detecting DAS levels. The results obtained during trinary detection were only impressive when compared to binary detection, which is a limitation of this work. Amanat *et al*. [25] proposed an LSTM based RNN model for predicting depression from text to protect individuals from mental disorders and suicidal thoughts. In order to detect sadness from text, semantics, and written content, RNN was trained on textual data. The presented approach represented a viability of RNN and LSTM by developing positive outcomes for the early detection of depression in the emotions of multiple social media subscribers. The proposed approach obtained higher accuracy when compared to SVM, CNN, and Decision tree. The behavior of mentally disturbed individuals was not identified with this approach, but the implementation of hybrid RNN was useful in overcoming this limitation. Acik *et al*. [26] presented a Mediterranean Diet Prevention (PREDIMED) dataset to detect anxiety of people following diet and having psychological problems presented by using the DASS-42 dataset. The mental illness and diet score problems resulted inversely, where the relation among mental health and diet users was analysed by utilizing the logistic regression models. The limitation of using PREDIMED presented fewer quality results in the prediction of mental disorder problems like anxiety and depression, and also failed in accessing the diet level randomly. The limitations found from the literature review are overfitting, data misclassification, data imbalance, and inaccurate results in prediction task. In this research, a deep learning algorithm is proposed for efficient classification of data with accurate results by reducing overfitting in prediction task. The primary contributions of this research are:

– Here, deep learning algorithms such as CNN, RNN and LSTM are used for predicting the emotions of mental disorders like anxiety, stress, and depression from the input DASS-42 dataset.
– Then, hyperparameter tuning using the grid search optimization technique is employed for identifying the best operating conditions of the classifier models.
– At last, an ensemble model including the LSGDM mechanism is applied for improving the performance in prediction, where the objective is more stable and dependent on the decision weights which mainly helps in the estimation of accurate values of DAS levels.

This research paper is further arranged as follows: the proposed method is presented in section 2, and the hyperparameter tuning optimization is explained in section 3, while the results and discussion of proposed algorithm are demonstrated in section 4. At last, the conclusion of this research is given in section 5.

2. PROPOSED METHOD

The proposed LSGDM is used in choosing the best classifier output for classifying DAS levels. Figure 1 represents the flow of the experiment. The ensemble model using the LSGDM gives higher performance compared to the existing loss optimizers like Gradient descent, Adam, and Adagrad. The initial step is data collection, while data pre-processing is the next step which helps in filtering the data and removing noisy words in the data.

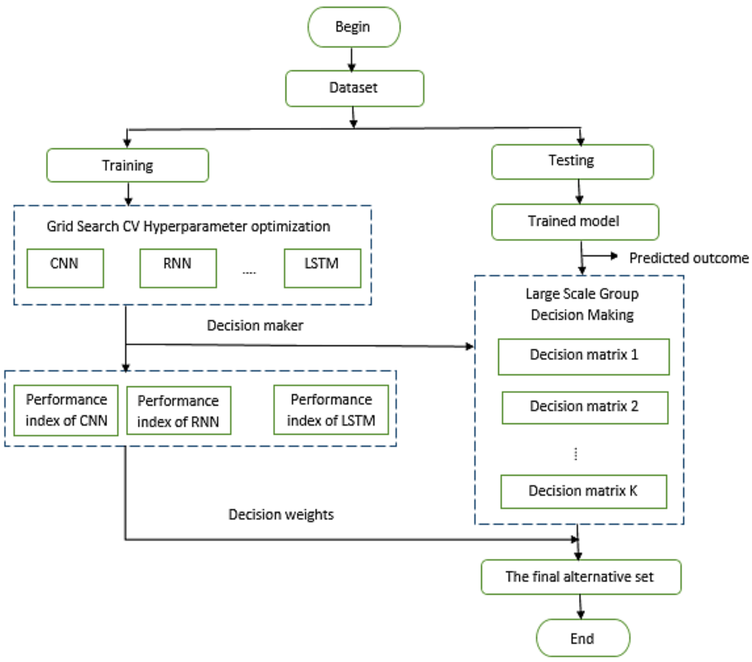

Figure 1. Flow chart of the proposed model

The extraction of features takes place to simplify the data, which makes it easy to understand the model. The extracted data is classified into two categories as testing and training. The ensemble model is trained by using the training data, and the testing ensemble method takes place for testing the ensemble model. Finally, the evaluation is carried out to enhance the performance of the ensemble model.

2.1. Data collection

The dataset is obtained from an online source which is available publicly based on DASS-42 [27] and consists of 39,776 data samples. The DASS is a 42-item self-report tool for assessing stress, anxiety, as well as depression. Each contain three 14-item subscales assessed on a Likert scale with four possible outcomes from 0 to 3. The total scores are identified by adding the items from each subscale, providing every scale a score range of 0-42. The scores for depression, anxiety, and stress subscales over 10, 8, and 15, accordingly. According to threshold scores for every stage, the severity of every situation is classified as none, mild, moderate, severe, or extremely severe, depending on the scores for each condition. Further, the subjects are divided into two classes for the purposes of the current study: none and mild (low anxiety), and rest (high anxiety) [24]. The total score for every state (e.g., anxiety) is computed by totaling scores of all 14 items. Each item involves a statement concerning the participant's experience, and the user can select how closely they match the statement in a previous week, with options "never", "sometimes", "often", and "almost always", which respects to scores 0, 1, 2, and 3, correspondingly. The negative emotions like Anxiety, Stress, and Depression are calculated using the measurement tool of DASS-42. The process of predicting the severity of depression takes place by including three variables like psychology, behaviour, and physical emotions where the severity score is between 1 and 4 (http://www2.psy.unsw.edu.au/groups/dass/Download%20files/Dass42.pdf). The collected dataset is then provided to the pre-processing step.

2.2. Data preprocessing

In this section, the collected data is utilized as input to the pre-processing step. A min-max normalizer [28]-[30] is used for pre-processing, which filters and removes unwanted and low-quality data from the input DASS-42 dataset, thereby maximizing the performance of the model. The detailed description about this technique with the mathematical expression is discussed below.

2.2.1. Min_Max normalization for rescaling

The normalization of the input feature is done by using the Min_Max scaler where the features are transferred to the range of (0,1). The variables that are measured at various scales do not result equal to the model fitting; hence, to normalize the fitness, the Min_Max scale is used. The formula for calculating Min_Max limits is represented in (1). The absolute minimum and maximum values of attribute data-A with Min(A), and Max(A) respectively.

$$v^{|} = \frac{v - min(A)}{max(A) - min(A)}\left(new_{max(A)} - new_{min(A)}\right) + new\_min\ (A) \quad (1)$$

where v is the initial value and the $v^{|}$ is the new value of every image. The range of boundary values of min and max ranges between new_Min(A) and new_Max(A), respectively.

The inconsistent and irrelevant samples are eliminated from the input dataset, resulting in 1292 efficient samples. The resulting clean data samples are segregated into distinct categories of DAS. The calculation of severity takes place based on the score evaluation in the DASS-42 dataset, which consists of 42 questions. The severity is estimated by merging the scores of the related questions where the input 42 questions are categorized into three categories, each with 14 questions, and presents the results into severe, normal, and mild categories. The ranges of various levels of DAS values are presented in Table 1.

Table 1. Ranges of DAS datasets at various levels

Severity	Depression	Anxiety	Stress
Normal	0-9	0-7	0-14
Mild	10-13	8-9	15-18
Moderate	14-20	10-14	19-25
Severe	21-27	15-19	26-33
Extremely severe	28+	20+	34+

In this section, data is pre-processed and severity levels of the mental illness are calculated. Then, the outcome of pre-processed data is forwarded to the hyperparameter tuning optimization. The detailed explanation about hyperparameter tuning optimization is provided in the next section.

3. HYPERPARAMETER TUNING OPTIMIZATION

The pre-processed data is classified into two categories as testing and training. A large-scale group decision making (LSGDM) method is proposed with the ensemble classifier model by combining CNN, RNN, and LSTM for effective classification of depression, anxiety and stress (DAS) levels. Hyperparameter tuning using the grid search optimization technique is employed for identifying the best operating conditions of the classifier models. The ensemble model is trained by using the collected data and then provided it to testing process. The error rate is reduced by selecting best classifier output using the LSGDM method and the parameters of the algorithm are improved using grid search-cross validation (grid-search CV) algorithm. The grid search CV is used to optimize the model for extracting high performance during the training of the proposed classifier algorithms at multiple times with various hyper-parameters, at an optimal balance between bias and variance. The performance of the model is improved by hyperparameters, or the grid search CV calculates the performance for every combination and results in the best value for hyperparameters. The cross-validation takes place while training the model in grid search CV and is time consuming based on the hyper-parameters. The grid search algorithm mainly helps to avoid the model from under-fitting and over-fitting the data. With this approach, a new model is automatically set to an entire training dataset using the parameters that give superior cross validation performance. The generalization performance is estimated more precisely with this technique. Using cross validation, the prediction error is estimated as represented in (2):

$$CV(f) = \frac{1}{n}\sum_{i=1}^{n} T(y_i, f^{-k(i)}(xi)) \quad (2)$$

Where k is the number of subsets, n is the size of the dataset, T is the loss function, and $f^{-k(i)}$ is the fitted function.

3.1. Ensemble classifier model

The extracted features are now provided as input to the ensemble classifier model in which CNN, RNN, and LSTM. The outcome of each base layer is given to the next base layer. This process of integrating all weak learners causes the framework to concentrate on every situation. The framework of this ensemble classifier is presented in two sections. Firstly, each individual classifier's ability is described and secondly, the LSGDM is presented to classify the data.

3.2. Convolutional neural network

CNN is the first base learner considered for classification of extracted features in the ensemble model. The input data is classified to reduce the noise existing in the training dataset and optimizes the weights initializations including the backpropagation. An outcome of every layer is taken as an input for next layer where the connected layer, classification layer, and pooling layer are involved in the classification of the data. The pooling takes place after convolution to optimize the overfitting and training time. The relevant extracted features are classified efficiently by using the CNN algorithm, where the training of the CNN takes place in a feedforward manner from input to output which is measured using (3) and (4). The output from the CNN is then given to the 2nd base learner which is considered to be RNN.

$$In_p^h = \sum_{q=1}^{n} W_{pq}^h X_q + b_p \quad (3)$$

$$Out_p^h = Max(0, In_p^h) \quad (4)$$

Where p represents the neural nodes in the layer h, where the input is taken from nodes of neural network q in the forward approach of a layer of h-1. b_p, W_{pq}^h are the bias terms and h layer weight vector, respectively. The output of CNN is computed by using the ReLu function.

3.3. Recurrent neural network

The RNN gives efficient predictions using the patterns, while the data sequential characteristics are recognized. The input sequence which is output from CNN is considered as x_t, and by keeping the hidden layer h_t of step time data, output y_t is predicting. The input dataset of the previous and the current time step of hidden variables are calculated using (5). The output from RNN is then given to LSTM.

$$h_t = \phi(x_t w_{xh} + h_{t-1} w_{hh} + b_h \quad (5)$$

Where ϕ is an activation function, w_{xh}, w_{hh} represent the weights, and b_h represents the deviation of the hidden layer.

3.4. Long short-term memory

LSTM is a RNN type to reduce long-term dependency by using the three layers of input, output, and forget gates of RNN. The current time step input is taken from the input of LSTM as x_t with the preceding time step of the hidden layer being h_{t-1}, while the fully connected layer is utilized to evaluate an output. The input gate I_t, forget gate F_t, and output gate O_t are calculated by using (6) to (8):

$$I_t = \sigma(W_{i^*}[h_{t-1}, x_t] + b_i) \tag{6}$$

$$F_t = \sigma(W_{f^*}[h_{t-1}, x_t] + b_f) \tag{7}$$

$$O_t = \sigma(W_{o^*}[h_{t-1}, x_t] + b_o) \tag{8}$$

Where σ is an activation function; W_{i^*}, W_{f^*}, W_{o^*}, b_i,b_i , b_o are the weights of network; h_{t-1} is the hidden layer of the previous time step; and x_t is the input for the specified time step.

3.5. Large scale group decision making

The group decision making is illustrated below to select the best classifier output. An m alternative such as $A_1, A_2, .., A_m$, committee of K decision makers such as $K_1, K_2, .., K_3$, and n decision criteria such as $C_1, C_2, ..., C_n$ are all considered in building a group decision matrix. Each decision maker has a chance of evaluating alternatives and its corresponding criteria individually. After the evaluation, the performance rating called x_{ij}^k is obtained for every alternative with respect to its criteria within a range of 0 to 1. Other important factors like each criterion weight $(w_1, w_2, .., w_n)$ and each decision maker $(W_1, W_2, .., W_k)$ are also considered. The decision making data is collected from the matrices x_k as shown in (9):

$$X_k = \begin{matrix} x_{11}^k & x_{1j}^k & x_{1m}^k \\ x_{i1}^k & x_{ij}^k & x_{im}^k \\ x_{n1}^k & x_{nj}^k & x_{nm}^k \end{matrix} \tag{9}$$

Where $i = 1,2, .., n$, $j = 1,2, .., m$ and $k = 1,2, .., K$. The large scale decision making approach is utilized to combine these matrices with important weights to facilitate decision making. A LSGDM based EL framework is used to produce the best classification result by merging different classification methods under the classification issue. Each classification method such as CNN, RNN, and LSTM is referred to as weak learners. This framework views weak learners as decision-makers, different categories as alternatives, and various categorization outcomes obtained by distinct base learners as performance ratings.

The performance of the classification is reflected in precision (P), recall (R), and accuracy (A). Precision is the most important performance measure that calculates similarity measure of the obtained values, calculated using in (10). The rate of accuracy performed to find the positive parameters are measured by the parameter Recall, measured using in (11). Accuracy represents the closeness between measurements and the true value as given by in (12). In case of unwanted data, the measure of accuracy increases the performance of the model.

$$P = \frac{TP}{TP+FP} \tag{10}$$

$$R = \frac{TP}{TP+FN} \tag{11}$$

$$A = \frac{TN+TP}{TN+TP+FP+FN} \tag{12}$$

Where TP and FP are the true positive and false positive values. The P, R, and A of base learner k for category i are represented by (13) to (15):

$$P_k = [p_1, p_2, \ldots.., p_m]^T \tag{13}$$

$$R_k = [r_1, r_2, \ldots, r_m]^T \tag{14}$$

$$A_k = [a_1, a_2, \ldots, a_m]^T \tag{15}$$

The weight of each decision maker is measured by in (16) which shows that larger weights come with larger indicators. The final classification result is measured using in (17).

$$W_k = P_k + R_k + A_k \tag{16}$$

$$H(x) = arg\,max \sum_{k=1}^{K} X_k W_k(x) \tag{17}$$

Where $H(x)$ is considered as an alternative x with the highest score.

4. RESULTS AND DISCUSSION

The ensemble classifier model results in effective predictions by validating the software environment of Python language using scikit-learn library. The configuration of the system with Windows-11 operating system with i5 processor, and 8 GB random access memory. The performance of the model is measured with true positive (TP), false positive (FP), true negative (TN), and false negative (FN) for measuring the metrices like accuracy, recall, precision, error rate, and F1-Score.

The number of correctly predicting classifiers to the ratio of predictions done overall is the accuracy of an model. The accuracy results in an efficient statistical rate in case the prediction is done efficiently. The ratio of the FP rate to the total of FP and TP is the result of error rate. The MCC is a significant individual value that ranges from 0 to 1, with 1 denoting the highest degree of agreement between the actual and anticipated values and 0 denoting no agreement. MCC's mathematical expression is provided in (18). The harmonic value of precision (P) and recall (R) is F1-Score, where the equivalence between P and R is evaluated. The F1-Score is measured using (19).

$$MCC = \frac{TN \times TP - FP \times FN}{\sqrt{(TP+FP)\times(FP+TN)\times(FN+TP)\times(FN+TN)}} \quad (18)$$

$$F1 - Score = \frac{2(Precision \times Recall)}{Precision + Recall} \quad (19)$$

4.1. Quantitative evaluation

It is evident that, still most models find it challenging in estimating their findings due to the lack of sufficient and validated evidence. In addition, not all the techniques perform equally well on all issues. Depending on the data samples obtained and the data's properties, the models' performance changes. Additionally, preliminary tasks like data cleaning and parameter tuning have an impact in acquiring the best outcomes. Therefore, it is crucial for researchers to explore and analyse the data using different algorithms in order to select the one with the highest accuracy.

Here, in this research, the proposed ensemble algorithm using the DASS-42 dataset as input is compared with existing algorithms like CNN, RNN, GNN, and LSTM by means of accuracy, sensitivity, specificity, MCC, and error rate. The adagrad, adam, and gradient descent loss optimization functions which does not result in efficient accuracy are used in the existing model, hence the ensemble techniques deploy the LSGDM loss optimization function for efficient results. The performance of the LSGDM function with the existing optimization functions is presented. The proposed ensemble classifier model results in the accuracy of 98.88%, sensitivity of 99.21%, specificity of 98.62%, matthew's correlation coefficient (MCC) of 99.41%, and F1-Score of 99.98%. The ensemble classifier techniques give commendable results in reducing complexity and runtime.

The proposed ensemble classifier model using the LSGDM optimization algorithm results in an accuracy of 98.88%, precision of 99.21%, recall of 98.62%, MCC of 99.41%, and an F1-Score of 98.90%, which evidences an improved performance compared to the existing loss optimizers. The ensemble method results in superior performance in comparison to the existing algorithms like CNN, GAN, RNN, and LSTM. The ensemble model improves the overall accuracy of classification and minimizes overfitting. The graphical representation of Table 2 with gradient descent, adam, adagrad, and LSGDM are represented in Figures 2 to 5, respectively.

Figure 2 depicts the graphical representation of the outcomes evaluated using the gradient descent loss optimizer. The ensemble model using gradient descent loss optimizer results in efficacious performance in terms of accuracy of 96.83%, sensitivity of 97.21%, specificity of 97.71%, F1-Score of 98.34%, and MCC of 99.06%. Figure 3 represents the graphical representation of the outcomes evaluated using the Adam loss optimizer. The ensemble model using Adam loss optimizer results in an accuracy of 96.08%, sensitivity of 96.97%, specificity of 97.93%, F1-Score of 97.01%, and MCC of 98.74%. Figure 4 represents the graphical representation of the results evaluated by using the Adagrad loss optimizer. The ensemble model using Adagrad loss optimizer results in an accuracy of 97.60%, sensitivity of 98.33%, specificity of 98.84%, F1-Score of 98.42%, and MCC of 99.39%.

Table 2 describes the results of ensemble models with different loss optimizer like Gradient descent, Adam, Adagrad and LSGDM. The effectiveness of the ensemble model is compared with the individual DL approaches like CNN, GAN, RNN and LSTM. The effectiveness of the ensemble approach is validated through the different performance metrices like accuracy, sensitivity, specificity, F1-Score, and MCC. The ensemble model acheives maximum accuracy results of 96.83%, 96.98%, 97.60% and 98.88% with the Gradient descent, Adam, Adagrad and LSGDM loss optimizer. As shown in Table 2, the ensemble model with LSGDM loss optimizer attains best results as compared to the other loss optimizers like Gradient descent, Adam and Adagrad.

Table 2. Experimental results of various loss optimizers

Algorithms	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)	MCC (%)
Results of ensemble models using Gradient descent loss optimizer					
CNN	89.16	89.96	90.98	90.94	91.62
GAN	91.65	91.88	92.36	92.46	92.96
RNN	93.19	93.77	94.25	94.42	94.47
LSTM	94.81	95.21	94.59	96.23	95.25
Ensemble	96.83	97.21	97.71	98.34	99.06
Results of ensemble models using Adam loss optimizer					
CNN	89.59	89.74	89.92	90.64	90.52
GAN	90.17	91.98	91.11	91.00	92.44
RNN	92.79	92.94	93.22	93.67	93.34
LSTM	94.12	94.48	94.63	95.86	95.33
Ensemble	96.08	96.97	97.93	97.01	98.74
Results of ensemble models using Adagrad loss optimizer					
CNN	90.83	90.79	90.99	91.44	91.16
GAN	91.73	92.84	92.09	93.99	93.92
RNN	93.59	94.23	94.31	95.17	95.77
LSTM	95.14	96.33	96.02	97.68	97.17
Ensemble	97.60	98.33	98.84	98.42	99.39
Results of ensemble models using LSGDM loss optimizer					
CNN	92.48	92.73	92.06	93.69	93.20
GAN	93.08	93.31	93.40	93.18	93.92
RNN	94.60	94.65	94.00	95.24	95.41
LSTM	95.25	96.78	96.50	97.61	97.95
Ensemble	98.88	98.21	98.62	99.98	99.41

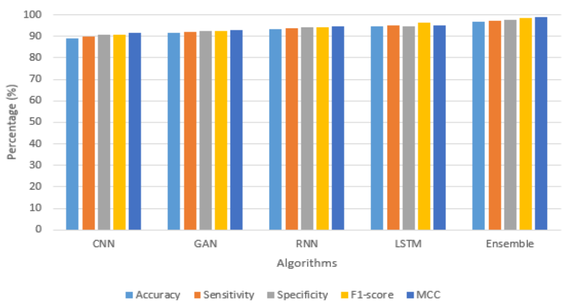

Figure 2. Results of ensemble models using Gradient descent loss optimizer

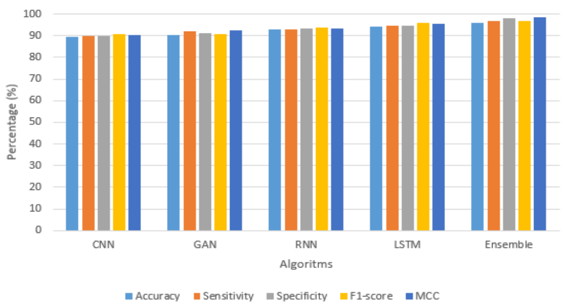

Figure 3. Results of ensemble models using Adam loss optimizer

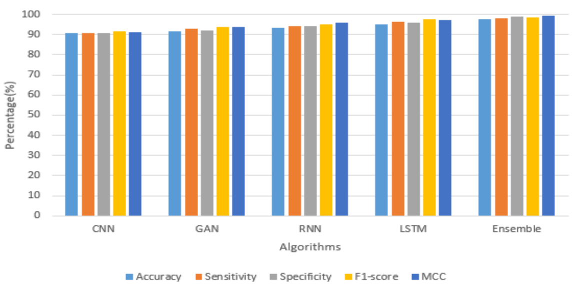

Figure 4. Results of ensemble models using Adagrad loss optimizer

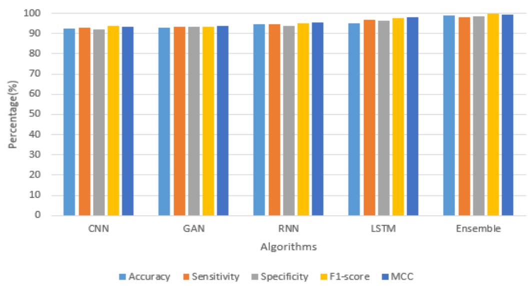

Figure 5. Results of ensemble models using LSGDM loss optimizer

The proposed ensemble classifier model using an LMBFS loss optimizer presents improved results compared to the existing loss optimizers. Figure 5 represents the graphical representation of the results evaluated using the LMBFS loss optimizer. The ensemble model using LMBFS loss optimizer results in an enhanced performance with an accuracy of 98.88%, sensitivity of 98.21%, specificity of 98.62%, F1-Score of 99.98%, and MCC of 99.41%.

4.2. Comparative analysis

The comparative analysis of the ensemble classifier model with the previous works like the ML based ensemble model [23] and PCA-mFCBF [24] on evaluation with the DASS-42 dataset of stress level classification is displayed in Table 3. To overcome the issues of data imbalance, small data size, and vanishing gradient in the existing methods, the ensemble classifier uses the advantage of decision weights on selecting the predicted outcomes of classifier models in the best suitable performance as output with prediction performance metric as the objective. The ensemble classifier model results in accuracy of 98.88%, sensitivity of 98.21%, specificity of 98.62%, F1-Score of 99.98%, and MCC of 99.41%. An ensemble classifier presents improved results in contrast to the existing models. The effectiveness of the model is improved using various ML algorithms like CNN, GNN, RNN, and LSTM which minimizes the variance, bias, and noise. The accuracy and stability of the model is improved using the machine learning algorithms by involving the boosting technique. The weight of observation is adjusted depending on the previous classifications by using the iterative technique of boosting. The proposed method solves the issues of gradient reduction, data misclassification, and data imbalance by identifying the relationship between non-linear and linear data.

Table 3. Comparative analysis of proposed ensemble model on DASS-42 dataset

Methods	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ML based ensemble model [23]	DASS-42	91.51	91.43	91.48	91.47
PCA-mFCBF [24]		82.5	-	-	-
Proposed model		98.88	98.21	99 .62	98.90

4.3. Discussion

This section discusses the advantages of the proposed LSGDM method over the limitations of the previous works for the prediction of mentall illness. The limitation of the previous work, ML based ensemble model [23] was that it was difficult to train the ML model for obtaining high robustness and accuracy due to data imbalance and small size data. In PCA-mFCBC [24], the results obtained during trinary detection were impressive only when compared to binary detection. To overcome these challenges, this research introduces the LSGDM with ensemble classifier model, combining CNN, RNN, and LSTM for effective classification of DAS levels. Hyperparameter tuning using the grid search optimization technique are employed for identifying the best operating conditions of the classifier models. The proposed method collects the DASS-42 dataset to estimate the efficiency of the proposed method. The proposed method results is compared with the existing methods like ML based ensemble model [23] and PCA-mFCBF [24] approaches on DASS-42 dataset. Here, the proposed method achieves an accuracy of 98.88%, precision of 98.21%, recall of 99.62% and F1-Score of 98.90%. It attains superior results when compared to the previous works. However, the proposed LSGDM method does not support for real-time applications, and furthermore improvements are required in social media diagnosis-based mental illness prediction. In future, these limitations will have been overcome to enhance the performance of the model.

5. CONCLUSION

The efficient machine learning approaches are presented to predict mental illnesses like depression, stress, and anxiety using the dataset of DASS-42. The mental disorder is a complex problem and its prediction in the early stages of the illness reduces the severity of the disease. The effectivenss of illness prediction using the proposed method is validated based on the various performance metrices. The proposed ensemble model using the LSGDM loss optimizer results in an accuracy of 98.88%, precision of 98.21%, recall of 98.62%, F1-Score of 98.90%, and MCC of 99.41%. The proposed LSGDM loss optimizer offers improved results, in contrast to the existing Gradient descent, Adam, and Adagrad optimizers. In future, the synthetic data samples will be deployed using data augmentation techniques, with dimensionality reduction for real-time predictions for improving the social media diagnosis-based mental illness predictions.

ACKNOWLEDGMENTS

We thank to "Centre of Excellence in AI&ML (COE-AIML)", which is funded by Govt. of Karnataka, India, located at Sahyadri College of Engineering and Management, Mangaluru, for providing support to this research work.

REFERENCES

[1] K. O. Asare *et al.*, “Mood ratings and digital biomarkers from smartphone and wearable data differentiates and predicts depression status: A longitudinal data analysis,” *Pervasive and Mobile Computing*, vol. 83, p. 101621, Jul. 2022, doi: 10.1016/j.pmcj.2022.101621.

[2] S. Mahato, N. Goyal, D. Ram, and S. Paul, "Detection of Depression and Scaling of Severity Using Six Channel EEG Data," *Journal of Medical Systems*, vol. 44, no. 7, p. 118, Jul. 2020, doi: 10.1007/s10916-020-01573-y.
[3] M. P. Uwimbabazi, J. d. A. Muziki, A. Muhayisa, T. Uwera, and J. Mutabaruka, "The mediating role of anxiety between negative feelings and depression among students with congenital physical disabilities," *PLoS ONE*, vol. 18, no. 3 March, p. e0281430, Mar. 2023, doi: 10.1371/journal.pone.0281430.
[4] R. Grover, P. Dua, S. Juneja, L. Chauhan, P. Agarwal, and A. Khurana, "'Depression, Anxiety and Stress' in a Cohort of Registered Practicing Ophthalmic Surgeons, Post Lockdown during COVID-19 Pandemic in India," *Ophthalmic Epidemiology*, vol. 28, no. 4, pp. 322–329, Jul. 2021, doi: 10.1080/09286586.2020.1846757.
[5] A. S. Uban, B. Chulvi, and P. Rosso, "An emotion and cognitive based analysis of mental health disorders from social media data," *Future Generation Computer Systems*, vol. 124, pp. 480–494, Nov. 2021, doi: 10.1016/j.future.2021.05.032.
[6] L. Wang, H. Liu, and T. Zhou, "A sequential emotion approach for diagnosing mental disorder on social media," *Applied Sciences (Switzerland)*, vol. 10, no. 5, p. 1647, Mar. 2020, doi: 10.3390/app10051647.
[7] A. A. Chamid, Widowati, and R. Kusumaningrum, "Graph-Based Semi-Supervised Deep Learning for Indonesian Aspect-Based Sentiment Analysis," *Big Data and Cognitive Computing*, vol. 7, no. 1, p. 5, Dec. 2023, doi: 10.3390/bdcc7010005.
[8] M. I. Khan and Y. M. Abbas, "Robust extreme gradient boosting regression model for compressive strength prediction of blast furnace slag and fly ash concrete," *Materials Today Communications*, vol. 35, p. 105793, Jun. 2023, doi: 10.1016/j.mtcomm.2023.105793.
[9] T. Richter, B. Fishbain, E. Fruchter, G. Richter-Levin, and H. Okon-Singer, "Machine learning-based diagnosis support system for differentiating between clinical anxiety and depression disorders," *Journal of Psychiatric Research*, vol. 141, pp. 199–205, Sep. 2021, doi: 10.1016/j.jpsychires.2021.06.044.
[10] M. T. Hawes, H. A. Schwartz, Y. Son, and D. N. Klein, "Predicting adolescent depression and anxiety from multi-wave longitudinal data using machine learning," *Psychological Medicine*, vol. 53, no. 13, pp. 6205–6211, Oct. 2023, doi: 10.1017/S0033291722003452.
[11] F. Javed, S. O. Gilani, S. Latif, A. Waris, M. Jamil, and A. Waqas, "Predicting risk of antenatal depression and anxiety using multi-layer perceptrons and support vector machines," *Journal of Personalized Medicine*, vol. 11, no. 3, p. 199, Mar. 2021, doi: 10.3390/jpm11030199.
[12] K. M. Dalmeida and G. L. Masala, "Hrv features as viable physiological markers for stress detection using wearable devices," *Sensors*, vol. 21, no. 8, p. 2873, Apr. 2021, doi: 10.3390/s21082873.
[13] H. El Hamdaoui, S. Boujraf, N. E. H. Chaoui, B. Alami, and M. Maaroufi, "Improving Heart Disease Prediction Using Random Forest and AdaBoost Algorithms," *International journal of online and biomedical engineering*, vol. 17, no. 11, pp. 60–75, Nov. 2021, doi: 10.3991/ijoe.v17i11.24781.
[14] S. Kaewchada, S. Ruang-On, U. Kuhapong, and K. Songsri-In, "Random forest model for forecasting vegetable prices: a case study in Nakhon Si Thammarat Province, Thailand," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 5, pp. 5265–5272, Oct. 2023, doi: 10.11591/ijece.v13i5.pp5265-5272.
[15] Al-Khowarizmi, R. Syah, M. K. M. Nasution, and M. Elveny, "Sensitivity of MAPE using detection rate for big data forecasting crude palm oil on k-nearest neighbor," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 3, pp. 2696–2703, Jun. 2021, doi: 10.11591/ijece.v11i3.pp2696-2703.
[16] F. Delmastro, F. Di Martino, and C. Dolciotti, "Cognitive Training and Stress Detection in MCI Frail Older People through Wearable Sensors and Machine Learning," *IEEE Access*, vol. 8, pp. 65573–65590, 2020, doi: 10.1109/ACCESS.2020.2985301.
[17] D. Shin, K. J. Lee, T. Adeluwa, and J. Hur, "Machine learning-based predictive modeling of postpartum depression," *Journal of Clinical Medicine*, vol. 9, no. 9, pp. 1–14, Sep. 2020, doi: 10.3390/jcm9092899.
[18] F. Ceccarelli and M. Mahmoud, "Multimodal temporal machine learning for Bipolar Disorder and Depression Recognition," *Pattern Analysis and Applications*, vol. 25, no. 3, pp. 493–504, Aug. 2022, doi: 10.1007/s10044-021-01001-y.
[19] K. S. Srinath, K. Kiran, S. Pranavi, M. Amrutha, P. D. Shenoy, and K. R. Venugopal, "Prediction of Depression, Anxiety and Stress Levels Using Dass-42," in *2022 IEEE 7th International conference for Convergence in Technology, I2CT 2022*, pp. 1–6, Apr. 2022, doi: 10.1109/I2CT54291.2022.9824087.
[20] K. S. Srinath, K. Kiran, A. G. Gagan, D. K. Jyothi, P. D. Shenoy, and K. R. Venugopal, "Enhancing Mental Illness Prediction using Tree based Machine Learning Approach," in *2022 IEEE International Conference on Electronics, Computing and Communication Technologies, CONECCT 2022*, pp. 1–5, Jul. 2022, doi: 10.1109/CONECCT55679.2022.9865689.
[21] P. E. Sumolang and W. Maharani, "Depression Detection on Twitter Using Bidirectional Long Short Term Memory," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 2, pp. 369–376, Sep. 2022, doi: 10.47065/bits.v4i2.1850.
[22] A. U. Rahman and Z. Halim, "Identifying dominant emotional state using handwriting and drawing samples by fusing features," *Applied Intelligence*, vol. 53, no. 3, pp. 2798–2814, Feb. 2023, doi: 10.1007/s10489-022-03552-x.
[23] Y. H. Sun, H. Luo, and K. Lee, "A Novel Approach for Developing Efficient and Convenient Short Assessments to Approximate a Long Assessment," *Behavior Research Methods*, vol. 54, no. 6, pp. 2802–2828, Jan. 2022, doi: 10.3758/s13428-021-01771-7.
[24] J. A. Nolazco-Flores, M. Faundez-Zanuy, O. A. Velázquez-Flores, C. Del-Valle-soto, G. Cordasco, and A. Esposito, "Mood State Detection in Handwritten Tasks Using PCA–mFCBF and Automated Machine Learning," *Sensors*, vol. 22, no. 4, p. 1686, Feb. 2022, doi: 10.3390/s22041686.
[25] A. Amanat *et al.*, "Deep Learning for Depression Detection from Textual Data," *Electronics (Switzerland)*, vol. 11, no. 5, p. 676, Feb. 2022, doi: 10.3390/electronics11050676.
[26] M. Açik, M. Altan, and F. P. Çakiroğlu, "A cross-sectionally analysis of two dietary quality indices and the mental health profile in female adults," *Current Psychology*, vol. 41, no. 8, pp. 5514–5523, Aug. 2022, doi: 10.1007/s12144-020-01065-9.
[27] I. Ramadhani and W. Maharani, "Predicting Depressive Disorder Based on DASS-42 on Twitter Using XLNet's Pretrained Model Classification Text," *Journal of Computer System and Informatics (JoSYC)*, vol. 3, no. 4, pp. 379–385, Sep. 2022, doi: 10.47065/josyc.v3i4.2157.
[28] M. J. Islam, S. Ahmad, F. Haque, M. B. I. Reaz, M. A. S. Bhuiyan, and M. R. Islam, "Application of Min-Max Normalization on Subject-Invariant EMG Pattern Recognition," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022, doi: 10.1109/TIM.2022.3220286.
[29] H. M. A. Fattah, K. M. A. Hasan, and S. Das, "A voting classifier for the treatment of employees' mental health disorder," *2021 International Conference on Automation, Control and Mechatronics for Industry 4.0, ACMI 2021*, pp. 1–6, 2021, doi: 10.1109/ACMI53878.2021.9528102.
[30] P. S. Topannavar and D. M. Yadav, "An effective feature selection using improved marine predators algorithm for Alzheimer's disease classification," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 5, pp. 5126–5134, Oct. 2023, doi: 10.11591/ijece.v13i5.pp5126-5134.

BIOGRAPHIES OF AUTHORS

Srinath Kudlapura Shivaiah received the B.E. degree in Information Science and Engineering and M.Tech. degree in Networks and Internet Engineering from Visvesvaraya Technological University, India. Currently working as research scholar in Department of Computer Science and Engineering in University Visvesvaraya College of Engineering, Bangalore University, Bengaluru, India. He can be contacted at email: solansrinath@gmail.com.

Kiran Krishnappa is currently working as an associate professor in the Department of Computer Science and Engineering, University of Visvesvaraya College of Engineering, Bangalore. He completed his Ph.D. in 2015 from Bangalore University. His areas of research include machine learning, natural language processing, cyber security, and wireless networks. He has published 25 papers in refereed international conferences and journals. He is also a senior member of IEEE and currently serving as execom member of IEEE Bangalore section. He can be contacted at email: kirank@uvce.ac.in.

Naveen Kumar Boraiah received the B.E. and M.Tech. degrees in computer science and Engineering from Visvesvaraya Technological University in 2014 and Ph.D. degree in Computer Science and Engineering from UIET, Chhatrapati Shahu Ji Maharaj University famously called as Kanpur University in 2019 (since from 2014 to 2016). Currently working as Associate Professor in Sahyadri College of Engineering and Management in the Department of Information Science and Engineering. He can be contacted at email: navkan24@gmail.com.

Punjalkatte Deepa Shenoy is currently working as a professor in the Department of Computer Science and Engineering, University of Visvesvaraya College of Engineering, Bangalore. She did her doctorate in the area of Data Mining from Bangalore University in the year 2005. Her areas of research include data mining, soft computing, biometrics, and social media analysis. She has published more than 200 papers in refereed international conferences and journals. She is also a senior member of IEEE and currently serving as chair of, IEEE Bangalore section. She can be contacted at email: shenoypd1@gmail.com,

Venugopal Kuppanna Rajuk is a former Vice Chancellor, Bangalore University, Bangalore. He obtained his Bachelor of Engineering from University Visvesvaraya College of Engineering. He received his Masters degree in Computer Science and Automation from Indian Institute of Science Bangalore. He was awarded Ph.D. in Economics from Bangalore University and Ph.D. in Computer Science from Indian Institute of Technology, Madras. He has authored and edited 79 books on Computer Science and Economics and has published over 1000 research papers. His research interests include computer networks, wireless sensor networks, parallel and distributed systems, digital signal processing, and data mining. He can be contacted at email: venugopalkr@gmail.com.