

# Exploratory Boosted Feature Selection and Neural Network Framework for Depression Classification

Vanishri Arun<sup>1\*</sup>, Murali Krishna<sup>2</sup>, Arunkumar B.V.<sup>3</sup>, Padma S.K.<sup>1</sup>, Shyam V.<sup>4</sup>

<sup>1</sup> Department of Information Science and Engineering, SJCE, JSS S&T University, Mysuru (India)

<sup>2</sup> Department of Psychiatry, CSI Holdsworth Memorial Hospital, Mysuru (India)

<sup>3</sup> Department of Anaesthesiology, Apollo BGS Hospital, Mysuru (India)

<sup>5</sup> Director and President, Forus Healthcare Pvt. Ltd., Bengaluru (India)

Received 11 July 2018 | Accepted 19 September 2018 | Published 5 October 2018



## ABSTRACT

Depression is a burdensome psychiatric disease common in low and middle income countries causing disability, morbidity and mortality in late life. In this study, we demonstrate a novel approach for detection of depression using clinical data obtained from the on-going Mysore Studies of Natal effects on Ageing and Health (MYNAH), in South India where the members have undergone a comprehensive assessment for cognitive function, mental health and cardiometabolic disorders. The proposed model is developed using machine learning approach for classification of depression using Meta-Cognitive Neural Network (McNN) classifier with Projection-based learning (PBL) to address the self-regulating principles like how, what and when to learn. XGBoost is used for feature selection on the available data of assessments with improved confidence. To improve the efficiency of McNN-PBL classifier the best parameters are found using Particle Swarm Optimization (PSO) algorithm. The results indicate that the McNN-PBL classifier selects appropriate records to learn and remove repetitive records which improve the generalization performance. The study helps the clinician to identify the best parameters to analyze the patient.

## KEYWORDS

XGBoost, Meta-Cognitive Neural Network, Projection-based Learning, Particle Swarm Optimization, Depression, MYNAH Cohort.

DOI: 10.9781/ijimai.2018.10.001

## I. INTRODUCTION

**D**EPRESSION is a non-communicable psychiatric disorder which is burdensome to people of all ages. According to psychiatrists depression is characterized by insomnia, tearfulness, anhedonia, suicidality, anorexia, self-depreciation, fatigue and lack of concentration. These characteristics will impede a person's capability to perform her or his daily activities normally. Due to lack of culturally adapted and validated assessments in low and middle income countries, assessment of the prevalence of depression becomes difficult. Psychiatrists diagnose depression using various parameters based on Diagnostic and Statistical Manual of Mental Disorders (DSM) [1]. Early diagnosis of depression helps in prevention of onset of depression related mental disorders like dementia, memory deterioration, Alzheimer's, etc., by treating at right place and time [2].

The risk of depression in elderly who have crossed the age of 60 years, which is called geriatric depression, increases due to factors like dependency on children for their living, physical disability, strokes, hypertension, diabetes, obesity, cancer, chronic pain, and also due to certain medicines. Depression affects an individual's personal and professional life that leads to decrease in productivity, increase in healthcare cost and exclusion from family and friends. Due to poor resource setting, lack of skilled people and increase in population depression is grossly undertreated. There is a very little data available

or recorded on geriatric depression and very less research has been done on identification of depression using machine learning approaches. Some of the researchers have worked on small datasets using tools like Weka, SPSS, etc. The reason behind this area not being well explored is that the datasets are not properly maintained by the hospitals or the clinicians. Most of the datasets available are incomplete with either missing parameters or missing values. Therefore MYNAH cohort is used in our study which has good dimensions of 1201 features and 1321 patient records with phenotypic data. The values in the cohort are recorded after a comprehensive assessment of the patients for various mental and physical disorders.

Machine learning techniques help in detection of disorders faster with good accuracy and reduced misclassification rate. Knowledge acquisition for identification of depression from dataset is done using machine learning approaches. A novel model is presented in this paper for depression classification using XGBoost technique for feature selection and McNN-PBL for classification by finding the best parameters for McNN using PSO to improve the efficiency of the classifier. Using the XGBoost technique on data for feature selection makes the classifier learning and training faster, also reduces the misclassification error rate and improves the overall accuracy of the model. The algorithm is scalable and can handle large datasets. Therefore XGBoost technique was chosen for feature selection. Depending on each feature of the patient record in the cohort, McNN decides which record it should choose to learn or the record to be deleted so that redundancy is avoided addressing what-to-learn. When a patient record is input to the model, it is not always necessary for the model to learn immediately. It may be reserved for future learning

\* Corresponding author.

E-mail address: vanishriarun@gmail.com

which addresses when-to-learn. The records may also be used for either addition of a neuron or updating of weights of the output which addresses how-to-learn component of meta-cognition. Initially, McNN starts with zero hidden neurons. It then adds sufficient neurons so that the decision surface is approximated. The principle adopted by the projection based learning algorithm is minimization of energy function. It finds the network output parameters which have a minimum energy function. The best parameters to improve the efficiency of McNN-PBL classifiers are found using PSO algorithm. McNN classifier discards repeated records from the training input data records. This reduces the memory requirement, also minimizes the computation time of the model and avoids over training. Another advantage of McNN is that it helps to reduce misclassification error [3].

In the past, researchers have worked on feature selection and disease classification using various approaches. To develop a scalable and efficient classification model, XGBoost has been incorporated with gradient boosting [4]. Boosting technique has been widely used across different domains for both classification and regression. Prediction of bioactive molecule facilitates the computer-aided drug discovery and XGBoost has shown a good performance on various datasets used [5]. XGboost has been used for the classification of DDoS attack which has shown significant performance compared to SVM and Random Forest [6]. Diabetes detection on a larger dataset using boosting approach has shown to be easily scalable [7]. Various Machine learning approaches are used to identify patients with depression. Depressed patients have been identified through classification of patients into subtypes based on the syndromes using the Beck Depression Inventory (BDI) item scores implementing categorization algorithm [8]. Depression diagnosis and feature reduction have been done using Support Vector Machine (SVM) with Voxel based morphometry and Filter method using ANOVA [9]. SVM which is a linear kernel with Principal Component Analysis (PCA) was used with image modality during depression related functional MRI tasks [10]. PCA was also used to minimize the number of attributes in the dataset and an ensemble classification framework has been used with Hierarchical Majority Voting (HMV) for disease classification and prediction [11]. Robust Spatial Kernel FCM (RSKFCM) segmentation method to detect Diabetic Retinopathy (DR) has been used for optic disk elimination in the retinal image and McNN approach has been used for DR classification [12]. Relevance Vector Regression with filtering out voxels from brain regions has been used to evaluate the BDI and HRSD scores. Non-linear Gaussian Kernels like Relevance Vector Machines and Support vector machines have also been used [13]. Effective Machine learning techniques are applied for detection of depression which has yielded an average accuracy of 80% and the size of the dataset used for the study is small which ranges between 18 and 62 patient records using mainly neuro-imaging. Image analysis uses a wide range of data for the assessment of depression. The previous studies on classification and prediction using machine learning approaches are done on various areas. But no study has yet been conducted on mental health cohort.

Hence we propose a machine learning model using a novel approach to classify depressed patients in the well populated cohort with good number of parameters using exploratory feature selection using boosting technique and McNN-PBL for classification.

The organization of the paper is as follows: Section II describes the Exploratory Feature selection using XGBoost technique. Section III presents the proposed methodology for depression detection using boosting technique for feature selection and McNN-PBL approach for depression classification. Section IV evaluates the performance of the proposed methodology and section V summarizes the conclusion and the scope for future study.

## II. EXPLORATORY FEATURE SELECTION USING BOOSTING TECHNIQUE

### A. Exploratory Feature Selection

Feature selection is a method used to choose a small set of parameters among a large parameter set. In this study, we have employed exploratory feature selection since the number of significant parameters required for identification of depression was unknown. The psychiatrist identified 45 out of 1201 parameters as significant parameters for detection of depression in the cohort based on his expertise. In exploratory feature analysis on SPSS statistical tool, Principal Component Analysis (PCA) method was used to reduce the complexity in the dataset and the Kaiser rule which is based on distribution theory of Eigen values was used as stopping criterion [14]. PCA involves a mathematical concept to transform a large number of correlated parameters into a small set of uncorrelated parameters called principal components [15]. These principal components serve as the significant features which act as predictors for depression detection. Eigen value is computed by examining the relationship between the parameters. The feature with Eigen value greater than 1 is chosen as the significant feature.

Using SPSS tool, 13 features were identified as significant features for detection of depression as shown in the Scree plot in Fig 1.

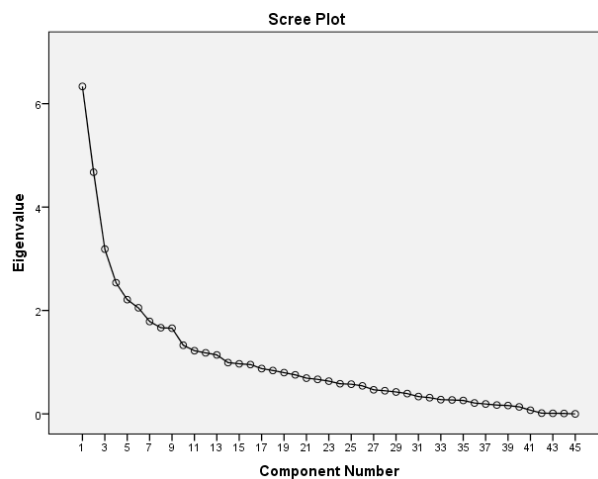


Fig. 1. Scree plot for Factor Analysis using PCA.

In our study, we used boosting technique to select the significant parameters by comparing each parameter applying decision rules to a decision tree model. The parameters are represented as internal nodes of the tree and the branching or the decision path is judged based on the parameters of the node. To get a single outcome at every leaf, a tree is created for the entire data. The model consists of a group of trees called an ensemble. In this ensemble, each tree is a decision tree called a weak learner or shallow tree based on the depth. XGBoost stands for Extreme Gradient Boosting and is an implementation of gradient boosted decision trees. XGBoost learns the data using the ensemble of boosted trees. It also handles the trade-off between the complexity of the model which is caused due to factors like number of trees, depth, etc. XGBoost technique for feature selection makes the learning and training faster. It also reduces the error rate during classification that in turn improves the overall accuracy of the model. Using boosting technique the significant factors chosen by the model were 10 out of 45 parameters. The 10 features chosen using XGBoost were given as input to the McNN-PBL model for depression classification.

### B. Depression Detection Using the Features Selected from the Cohort Using XGBoost and McNN-PBL Classifier Model

The McNN is trained with the input and the desired output after which the training occurs. Weights at both the hidden and output layers are adjusted so that the actual output corresponds to the desired output. Once trained, the neural network takes a new patient record and gives either 0 or 1 as output where 0 denotes absence of depression and 1 denotes presence of depression. If the classifier is uncertain it will produce a value somewhere in between 0 and 1.

When a patient record is input to the model, the learning process in McNN addresses the self-regulating principles like what, when and how to learn in a manner similar to the cognition principle of a human brain. The learning process in McNN uses estimated class label, maximum hinge error and class wise significance. McNN chooses the patient record from the cohort to be used for learning to address what-to-learn principle depending on the parameter of the record. It also chooses the patient record to be deleted to avoid redundancy. The model does not have to start the learning process as soon as the patient record is presented to it. It may reserve the record for future learning addressing when-to-learn principle. To address how-to-learn component, the input patient records may be used for either addition of a neuron or updating of weights of the output. To start with, McNN will have zero hidden neurons. It then adds sufficient neurons to approximate the decision surface. Projection Based Algorithm adopts the principle of minimization of energy function in which it finds the network output parameters having a minimum energy function. To find the best parameters to improve the efficiency of McNN-PBL classifiers, we used PSO algorithm. Another advantage of McNN classifier is that it discards the same records from the training input data records which in turn reduces the memory requirement and computation time of the model. It avoids over training and helps in reduction of misclassification error rate.

### III. DEPRESSION DETECTION USING BOOSTED FEATURE SELECTION AND McNN-PBL APPROACH - PROPOSED METHODOLOGY

The detection of depression using MYNAH cohort is done in two phases. The first phase consists of feature selection using XGBoost technique. The second phase involves detection of Depression using McNN-PBL.

Fig. 2 represents the overall proposed model for the detection of depression using MYNAH cohort.

The model developed can be used as a classifier with the following functionality:

#### A. Data Acquisition and Assessment of Patients

##### 1) The MYNAH Cohort

MYNAH is an abbreviation of MYsore studies of Natal effect on Ageing and Health, a cohort consisting of a total of 3,427 men and women born between 1934 and 1966 as singletons at Holdsworth Memorial Hospital, Mysore, Southern India. These people were traced by matching their birth records through a house-to house survey of the area of Mysore city surrounding the hospital between the years 1993 and 2001. A comprehensive examination of lung function, relationship between size at birth, adult cardio-metabolic disorders, etc., was conducted in which a total of 1069 people participated and this constituted the Mysore Birth Records Cohort. This cohort study was one of the first of its kind in low and middle income countries aiming to test developmental origins of health and disease (DOHaD) concepts with the predictions of associations between small size at birth and adult heart disease, resistance to insulin and reduced lung function and so on [16]. The surviving members in the cohort are aged

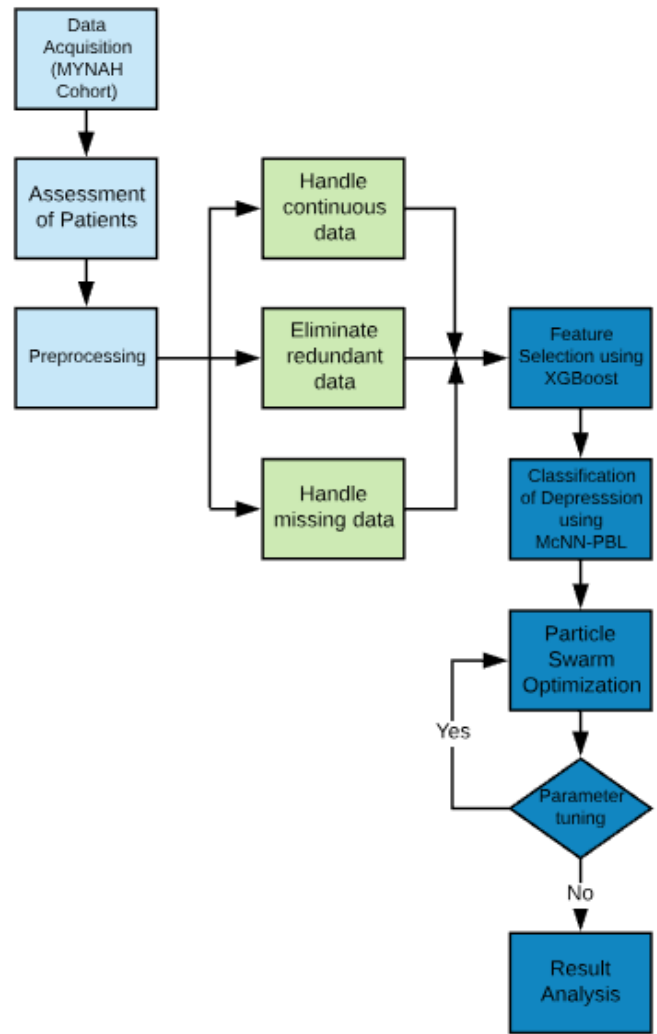


Fig. 2. Depression Classification Model.

over 60 years and the cohort serves as a unique resource for studies in epidemiology on old age.

##### 2) Assessment

Examination was conducted on 721 surviving members in the cohort between the years 2015 and 2017 for mental and cardio-metabolic disorders. Based on the symptoms depression was assessed using Geriatric Mental State (GMS) Examination which is internationally used for assessment of geriatric mental health [17]. A computerized diagnostic algorithm called Automated Geriatric Examination for Computer Assisted Taxonomy (AGECAT) is used to group the symptoms to form patterns recognized by a clinician as illness or syndrome class [18]. International Classification of Diseases (ICD) and the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria are used to add the items together and to generate affective disorder diagnoses. In-patient, out-patient and community samples are used to demonstrate the reliability and validity of the GMS. Validity of GMS/AGECAT algorithm has been investigated in several studies [19]. A Socio-demographic questionnaire was developed to collect information on age group, gender, Body mass index (BMI) group, level of education, Standard of living index based on standard indicators, Job, Job category, and others.

##### B. Preprocessing

In preprocessing stage, three issues relative to the cohort were addressed. Firstly, dealing with missing data, secondly, handling

the redundant features and finally, mapping of continuous data to categorical data or encoding of categorical features. The cohort consists of 1201 parameters recorded during the assessment. Based on feature analysis done by the psychiatrist, 45 parameters were chosen for the study. In some of the observations the assessment for the required parameters was missing and some observations were redundant. Such records were dropped which resulted in 270 patient records out of the total 1321. Continuous data features were encoded into categorical data like age group (0-below 65 years, 1- above 65years), gender (0-male, 1-female), Body mass index (BMI) group (0-Under weight, 1-Healthy, 2-Over weight, 3-Obese), level of education (0-Illiterate, 1-Secondary, 3-College, 4-Diploma, 5-Graduate, 6- Post graduate), and so on.

To avoid dominance of some parameters it is essential to normalize input features for the objective function to perform well and for the algorithm to converge during optimization. Euclidean norm was used to rescale the data into the range [0,1] by dividing each feature vector by the Euclidean length of the vector as shown in Eq. (1).

$$x' = \frac{x}{\|x\|} \tag{1}$$

In Eq. (1),  $x'$  is the normalized value,  $x$  is the original feature value and  $\|x\|$  is the Euclidean length of the vector.

Stratified Sampling method has been used in this study to handle the class imbalance by dividing the dataset into training and testing datasets in the ratio 75:25. The sampling function makes a split so that the proportion of values in the sample will be same as the proportion of values provided to parameter. It helps in dividing the training data into homogeneous groups and the ratio of samples from all the classes will be proportional. This will help in avoiding over fitting and bias towards the dominant class. Considering the stratified samples for training will help the algorithm to learn in an optimized way and it saves from further tuning or regularization.

### C. Feature Selection Using XGBoost Technique

The cohort consists of a large number of features. Hence to find the best set of features, we employed feature selection based on the feature relevance to identify significant factors as predictors for detection of depression. At each iteration of boosting, the best feature is searched and then added to the ensemble. This forms a combination of selected feature with new features. Model is fit to the training data and the importance of each feature is measured using a weight. Weight is taken as the number of times each feature appears in a tree and using these weights, feature selection is done. The top 10 features are retained based on the accuracy of the classifier. We conducted a number of trails including different number of features. But 10 features were found to be optimum after which, addition of more number of features enhances the accuracy of the classifier negligibly. For instance when the number of features was chosen as 22 there was an increase of 0.26 % in the overall accuracy. If 22 features were chosen the resultant model would have been too complex. With 10 features, the accuracy of the classifier was 97.22% and with 22 features the accuracy of the classifier was 97.48% after which the accuracy became constant.

The feature accuracy graph considering different number of features is shown in Fig. 3. Out of the 45 features identified by the psychiatrist, 10 features were considered to be of importance based on feature selection process as shown in Fig. 4. The features used for the study are named automatically based on their index in the input array from f1 through f45.

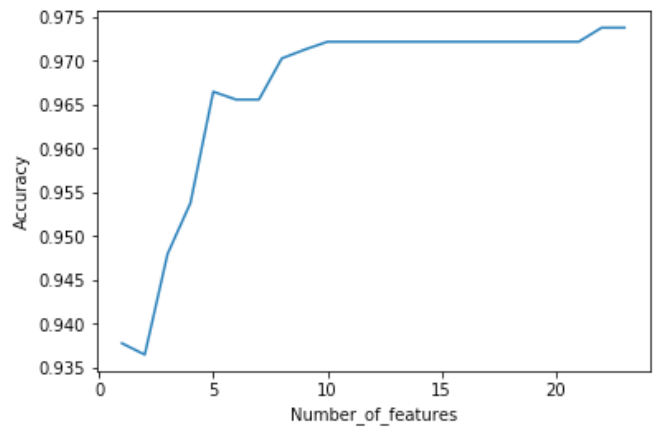


Fig. 3. Feature Accuracy graph.

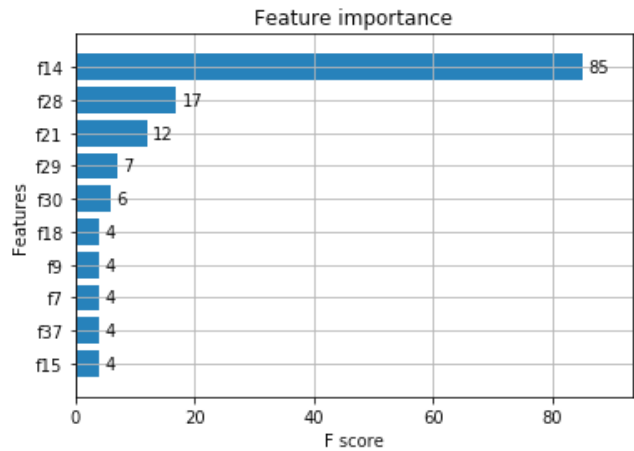


Fig. 4. Features of importance.

To identify the features of importance we manually mapped the indices to the feature names in the problem description and found that Eurotot has the highest importance followed by other features which are shown in Table I.

The importance of each feature considered for the study based on the psychiatrist’s analysis was identified by computing the F-score using Eq. (2)

$$F\text{-score} = \frac{2 * precision + Recall}{precision + Recall} \tag{2}$$

where precision and recall are based on confusion matrix in Table II.

Therefore the classifier model was built with 10 features given in Table I. The description of the features selected is given below:

- The Eurotot symptom scale is developed in Europe to compare the symptoms of geriatric depression across the continent. There are 12 eurotot items namely depressed mood, pessimism, wishing death, guilt, sleep, interest, irritability, appetite, fatigue, concentration, enjoyment and tearfulness which represent the depression scores taken from GMS. Each item is scored 0 for absence of symptom and 1 for presence of symptom. Therefore an ordinal scale is generated with a maximum score of 12.
- avggrip indicates right and left hand grip strength measured as the amount of force a hand can squeeze around a dynamometer [20].
- HTN or hypertension is one of the strong predictors of Depression [21].
- Frifailtytot is related to frailty which reduces a person’s ability to endure environmental stress. Frifailtytot is the sum of frailty scores where 0 indicates unintentional weight loss, 1 indicates exhaustion,

2 indicates muscle weakness, 3 indicates slowness while walking and 4 indicates low levels of activity [22].

- avgspifev is recorded through Spirometry Forced expiratory volume test and is the total amount of air exhaled in 1 second [23].
- cobese is the central obesity. [24].
- pjobgroup2 is associated with paid employment. The job categories used in this study are 0 indicating manager/administrator, 1 indicating professional, 2 indicating associate professional, 3 indicating clerical worker, 4 indicating shop keeper, 5 indicating skilled laborer, 6 indicating semi-skilled, 7 indicating laborer and 8 indicating agricultural worker.
- lencms is the length of the baby recorded at birth in cms. If the length of the baby at birth is less than the standard length, the foetal brain growth is shunted [25].
- m13ldlsi indicates the low density lipoprotein cholesterol measured in SI units. There will be an increased risk of depression if there are low cholesterol levels [26].
- bmi or the body mass index is calculated using the height and weight of a person which is a measure of body fat. There is a strong association between both underweight and obesity and depression [27].

TABLE I. FEATURES SELECTED USING XGBOOST TECHNIQUE

Feature Index	FEATURE
f14	Eurotot
f28	Avggrip
f21	HTN
f29	Frifrailtytot
f30	Avgspifev
f18	Cobese
f9	pjobgroup2
f7	Lencms
f37	m13ldlsi
f15	Bmi

#### D. Tuning Hyper Parameters

- Choosing depth indicates the height of tree that is used as an estimator and is computed using log loss function. It is run over multiple values and the best is chosen where loss is minimal as shown in Fig. 5.
- Choosing the number of estimators is an important task as we use an ensemble technique to make a collection decision for the model. We chose the number of estimators by computing the minimum loss as shown in Fig.6.

From the different trails we conducted, we set the number of estimators to be 50 with the depth of 3.

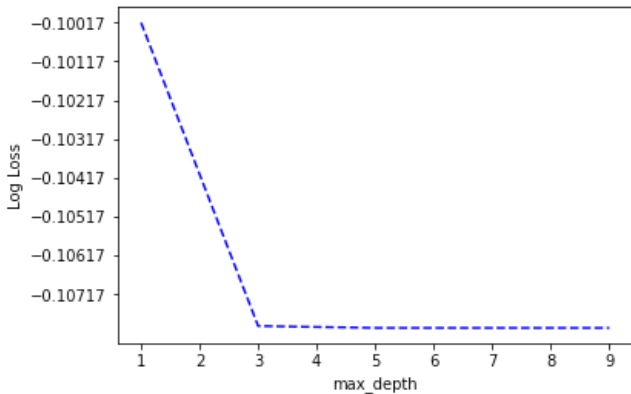


Fig. 5. Depth of the tree.

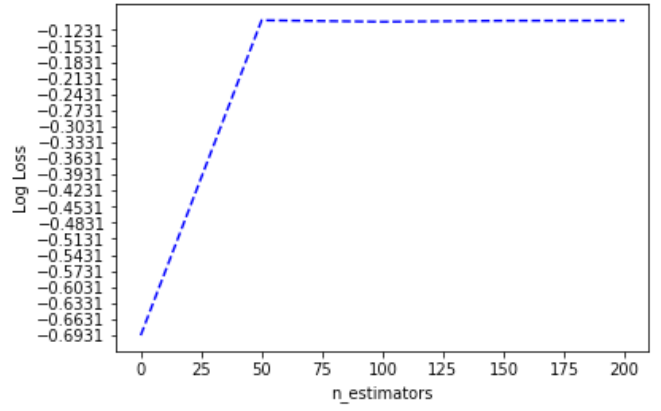


Fig. 6. Estimator setting.

The feature set obtained using boosting technique is given as input to the classification model built using McNN-PBL framework for detection of depression.

#### E. McNN-PBL Classifier

The two components of McNN are cognitive component and meta-cognitive component. The cognitive component consists of a three layered feed forward radial basis function network which takes each input  $x^t$  and produces an output  $\hat{y}_j^t$ . The meta-cognitive component consists of a dynamic model of the cognitive component. When a new training record  $x^t$  arrives, it takes the output  $\hat{y}_j^t$  and the knowledge present in the new training record is estimated. Now based on this information, the meta-cognitive component selects an appropriate strategy to control the cognitive component in learning from the training record. By this way, the meta-cognitive component addresses the principles namely what-to-learn, when-to-learn, and how-to-learn properly.

##### 1) Cognitive Component of McNN:

The three layered cognitive component of McNN consists of linear input, hidden and linear output layers in the network. The hidden layer of the cognitive component employs Gaussian activation function [28]. An assumption is made that the McNN builds  $K$  Gaussian neurons from  $t - 1$  training records without loss of generality. Let  $x^t$  be the given input, then  $\hat{y}_j^t$  which is the predicted output of the  $j^{th}$  output neuron of McNN is as in Eq. (3).

$$\hat{y}_j^t = \sum_{k=1}^K w_{kj} h_k^t, j = 1, 2, \dots, n \quad (3)$$

In Eq. (3),  $w_{kj}$  represents the weight connecting the  $k^{th}$  hidden neuron to the  $j^{th}$  output neuron.  $h_k^t$  represents the response of the  $k^{th}$  hidden neuron to the input  $x^t$  as in Eq. (4).

$$h_k^t = \exp\left(-\frac{\|x^t - \mu_k^l\|^2}{(\delta_k^l)^2}\right) \quad (4)$$

In Eq. (4),  $\mu_k^l \in R^m$  is the center and  $\delta_k^l \in R$  is the width of the  $k^{th}$  hidden neuron.  $l$  represents the class of the hidden neuron to which it belongs.

For learning process, the cognitive component uses Projection Based Learning (PBL) algorithm.

##### 2) Projection Based Learning (PBL) Algorithm:

The principle used by projection based learning algorithm is minimization of energy function. The algorithm finds the network output parameters with minimum energy function [29]. The energy function that is considered is specified based on the error at McNN output neurons and shown in Eq. (5).

$$J_j = \sum_{i=1}^n (y_j^i - \hat{y}_j^i)^2; i = 1, \dots, N \quad (5)$$

As we assume above, the McNN is processing the  $t^{th}$  input, therefore, the overall energy function is defined as in Eq. (6).

$$J(W) = \frac{1}{2} \sum_{i=1}^t J_i = \frac{1}{2} \sum_{i=1}^t \sum_{j=1}^n (y_j^i - \hat{y}_j^i)^2 \quad (6)$$

Using Eq. (2) we substitute the predicted output ( $\hat{y}_j^i$ ) and the energy function reduces to Eq. (7).

$$J(W) = \frac{1}{2} \sum_{i=1}^t \sum_{j=1}^n (y_j^i - \sum_{k=1}^K w_{kj} h_k^i)^2 \quad (7)$$

In Eq. 7  $h_k^i$  indicates the response of the  $k^{th}$  hidden neuron for  $i^{th}$  training record. The optimal output weights ( $W^* \in \mathbb{R}^{K \times n}$ ) are estimated in such a manner that the total energy reaches its minimum and is shown in Eq. (8).

$$W^* := \arg \min_{W \in \mathbb{R}^{K \times n}} J(W) \quad (8)$$

Let  $W \in \mathbb{R}^{K \times n}$  If  $J(W^*) \leq J(W) \forall W \in \mathbb{R}^{K \times n}$  then  $W^*$  is the optimal output weight corresponding to the minimum of the energy function.  $W^*$  has minimum value when the first order partial derivative of  $J(W)$  with respect to the output weight is 0 as in Eq. (9).

$$\frac{\partial J(W)}{\partial w_{pj}} = 0, p = 1, \dots, K; j = 1, \dots, n \quad (9)$$

The first derivative is as shown in Eq. (10).

$$\frac{\partial J(W)}{\partial w_{pj}} = \sum_{i=1}^t h_p^i [y_j^i - \sum_{k=1}^K w_{kj} h_k^i] p = 1, \dots, K; j = 1, \dots, n \quad (10)$$

Equating the derivative to 0, we get Eq. (11).

$$\sum_{k=1}^K \sum_{i=1}^t h_k^i h_p^i w_{kj} = \sum_{i=1}^t h_p^i y_j^i \quad (11)$$

Let matrix  $A \in \mathbb{R}^{K \times K}$  be as in Eq. (12).

$$a_{kp} = \sum_{i=1}^t h_k^i h_p^i, k = 1, \dots, K; p = 1, \dots, K \quad (12)$$

and the matrix  $B \in \mathbb{R}^{K \times n}$  be as in Eq. (13)

$$b_{pj} = \sum_{i=1}^t h_p^i y_j^i, j = 1, \dots, n; p = 1, \dots, K \quad (13)$$

Equation 10 can be rewritten as Eq. (14).

$$\sum_{k=1}^K a_{kp} w_{kj} = b_{pj}, p = 1, \dots, K; j = 1, \dots, n \quad (14)$$

or in a matrix form as Eq. (15).

$$A W = B \quad (15)$$

Therefore, the optimal  $W^*$  is as in Eq. (16).

$$W^* = A^{-1} B \quad (16)$$

With this optimal value of  $W^*$ , the energy function reaches its minimum value.

### 3) Meta-cognitive Component of McNN :

As measures of knowledge in the new training record, the meta-cognitive component uses estimated class label ( $\hat{c}^t$ ), maximum hinge error ( $E^t$ ) and spherical potential based class-wise significance to control the learning process of the cognitive component. The definitions of these measures are as below:

#### 4) Estimated Class Label ( $\hat{c}^t$ ):

The estimated class label ( $\hat{c}^t$ ) is obtained using the predicted output

( $\hat{y}^t$ ) as shown in Eq. (17).

$$\hat{c}^t = \arg \max_{j \in \{1, 2, \dots, n\}} \hat{y}_j^t; t = 1, \dots, N \quad (17)$$

#### 5) Maximum Hinge Error ( $E^t$ ):

The main objective of the classifier is to reduce the error between the predicted output ( $\hat{y}^t$ ) and the actual output  $y^t$  such that the error is minimized. The classifier that uses hinge loss function estimates more accurate posterior probability compared to the classifier developed using mean square error function. Therefore, in McNN, the hinge loss error ( $e^t = [e_1^t, \dots, e_j^t, \dots, e_n^t]^T$ )  $\in \mathbb{R}^n$  is used and is defined as in Eq. (18).

$$e_j^t = \begin{cases} 0 & \text{if } y_j^t \times \hat{y}_j^t > 1 \\ y_j^t - \hat{y}_j^t & \text{otherwise} \end{cases} j = 1, 2, \dots, n \quad (18)$$

The maximum absolute hinge error ( $E^t$ ) is as in Eq. (19).

$$(E^t = \max_{j \in \{1, 2, \dots, n\}} |e_j^t|) \quad (19)$$

#### 6) Class-wise Significance ( $\Psi_c$ ):

The input feature ( $x^t$ ) is mapped on to a hyper-dimensional spherical feature space  $S$  by using  $K$  Gaussian neurons, i.e.,  $x^t \rightarrow \phi$ . Hence, all  $\phi(x^t)$  lie on a hyper-dimensional sphere.

In McNN, the feature space  $S$  is described by center ( $\mu$ ) and width ( $\sigma$ ) of the Gaussian neurons. Let the center of the  $K$ -dimensional feature space be  $\phi_0 = \frac{1}{K} \sum_{k=1}^K \phi(\mu_k)$ .

The potential of the new data  $x^t$  in original space is the knowledge present in it. It is the squared distance from the  $K$ -dimensional feature space to the center  $\phi_0$ . The potential is as shown in Eq. (20).

$$\Psi = \|\phi(x^t) - \phi_0\|^2 \quad (20)$$

Eq. (20) can be expressed as in Eq. (21).

$$\Psi = \phi(x^t, x^t) - \frac{2}{K} \sum_{k=1}^K \phi(x^t, \mu_k^1) + \frac{1}{K^2} \sum_{k,r=1}^K \phi(\mu_k^1, \mu_r^1) \quad (21)$$

In the Gaussian function, the first term  $\phi(x^t, x^t)$  and last term  $\frac{1}{K^2} \sum_{k,r=1}^K \phi(\mu_k^1, \mu_r^1)$  in Eq. (21) are constants. These constants are discarded because potential is a measure of novelty. The potential can be reduced to Eq. (22).

$$\Psi = -\frac{2}{K} \sum_{k=1}^K \phi(x^t, \mu_k^1) \quad (22)$$

The class-wise distribution influences the performance of the classifier significantly. Therefore the measure of the spherical potential of the new training record  $x^t$  belonging to class  $c$  with respect to the neurons associated to same class is used. That is  $l=c$ .

The class-wise spherical potential or class-wise significance ( $\Psi^c$ ) is defined using  $K^c$  which is the number of neurons associated with the class  $c$  as shown in Eq. (23).

$$\Psi_c = -\frac{2}{K^c} \sum_{k=1}^{K^c} \phi(x^t, \mu_k^c) \quad (23)$$

The knowledge contained in the records is directly indicated by the spherical potential. A smaller potential means that the record is similar to the existing knowledge in the cognitive component. A higher potential that is close to zero means that the record is novel.

The main objective of McNN is to approximate the underlying function that maps  $x^t \in \mathbb{R}^m \rightarrow y^t \in \mathbb{R}^m$ . Initially the neuron network starts with zero neuron. Addition of new neurons or update of existing

neurons is done during the processing of records [28]. The 4 strategies used by the McNN are:

1. Sample/Record deletion:

In our study the dataset has to be classified as Depressed or Not Depressed. Therefore the model has to be trained to classify as Depressed and Not Depressed using some significant features when a new patient record is given for classification. Suppose the new record is the same as the trained one then that record is deleted.

2. Sample/Record reservation:

Suppose the input record is very much similar to the trained record with slight change then that feature of the new record is reserved for further use.

3. Parameter updating:

Suppose all the features of the input record are the same as trained record and a new feature is recognized then the parameter is updated in the training set.

4. Neuron Addition:

As the new features are updated along with the new input records, the neuron addition or growth happens.

7) PBL Algorithm for McNN Classifier:

PBL algorithmic steps for McNN Classifier are as shown below:

1. Using Eq. (3) & (4), compute the cognitive component output ( $\hat{y}^t$ ) for each new training record input ( $x^t$ ).
2. Using Eq. (17),(19) & (23), for the new training record ( $x^t$ ), the meta-cognitive component finds ( $\hat{c}^t$ ), ( $E^t$ ) and ( $\Psi_c$ ) which are estimated class label, maximum hinge error and class-wise significance measures respectively.
3. One of the strategies like Deletion of record, Growth of Neuron, Updating of parameters and Reservation of record are selected by the meta-cognitive component.
4. The above selected strategy is executed by the cognitive component.
5. Until there are no more records in the training data set iterate steps 1 to 4.

F. Particle Swarm Optimization (PSO)

To maximize the performance efficiency, optimized parameter values must be set for each dataset. The user can find the optimized parameter after choosing data files where each of the 9 parameters lies in a range which can be set manually. Otherwise default values for the 9 parameters are set based on several experimentation. The default ranges set are as shown in Fig. 7.

The dataset consists of two files namely training file and testing file containing training data and testing data respectively. The training file is used to train the McNN classifier by creating and growing its neurons network using PBL and the testing file is used to test the classifier to assess the efficiency of the algorithm. The performance of the classifier depends on 9 parameters namely Skip threshold ( $\beta_b$ ), Adding error threshold ( $\beta_a$ ), Learning error threshold ( $\beta_l$ ), Self-adaptive decay factor ( $\delta$ ), Overlap factor ( $K_2$ ), Overlap factor for first neuron ( $K_1$ ), Maximum number of reserved records, Spherical potential threshold ( $\phi_c$ ) and Center shifting factor ( $\zeta$ ). Hence for the classifier to perform better these parameters need to have best values fixed. We use a computational method called PSO in our study to fix best values for these parameters. The implementation of PSO consists of a class that provides methods to optimize parameters for a neuron network. It also keeps the default values and the ranges of the 9 parameters of a neuron network. For MYNAH data set, PSO is used to achieve the best parameters as shown in Fig. 8.

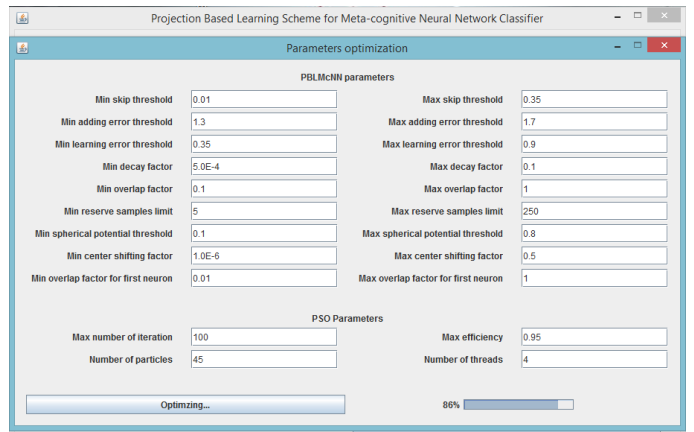


Fig. 7. Default range for 9 parameters.

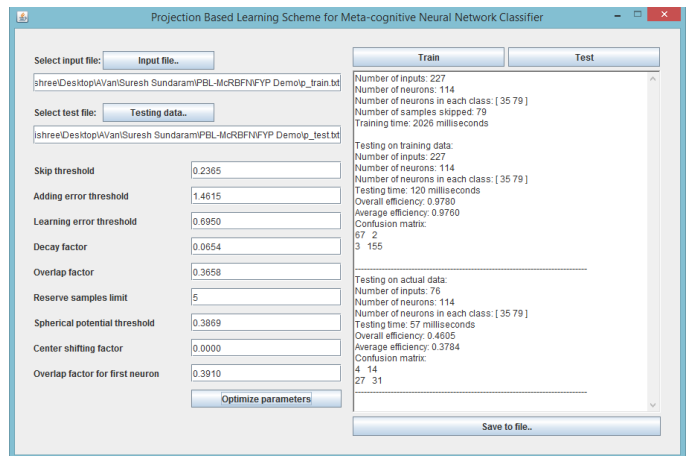


Fig. 8. Optimized Parameters for MYNAH cohort.

PSO algorithm contains a vector of the parameters. A parameter lies within a range, and it is initialized as in Eq. (24).

$$p = p_{min} + r \times (p_{max} - p_{min}) \quad (24)$$

Where  $p$  is the parameter initialized value,  $p_{min}$  is the minimum value of the parameter,  $p_{max}$  is the maximum value of the parameter and  $r$  is a random number which can be created by Math function.

Each particle of PSO algorithm also has a certain velocity, which is a vector of the 9 parameters. In McNN, the velocity is initialized as in Eq. (25).

$$v_i = r \times 0.1 \times (p_{i\ max} - p_{i\ min}) \quad (25)$$

Where  $v_i$  is the velocity of dimension  $i$  ( $i=l, \dots, 9$ ),  $r$  is a random number which is generated using the function  $\text{Math.random}()$ ,  $p_{i\ max}$  is the max value of  $i^{th}$  parameter ( $i=l, \dots, 9$ ) and  $p_{i\ min}$  is the min value of  $i^{th}$  parameter ( $i=l, \dots, 9$ ).

The main objective function of PSO algorithm is to enhance the overall testing efficiency. The program creates N particles to optimize parameters. In each iteration, train() and test() methods are invoked using the parameter set for each particle. To evaluate the particle current position we use the overall testing efficiency of the model. To terminate the application we used two stopping criteria namely number of iterations and efficiency. If the number of iterations is greater than a certain value, we have set 100 as default, then the application terminates. The second stopping criterion is the efficiency. If the overall training efficiency is less than 0.95 or if the overall testing efficiency is greater than 0.95 then the application is terminated. The stopping criterion for the optimization process is the global best value. If the global best value

is greater than 1.95 then optimizing process terminates. Since PSO takes more time to optimize because it requires iterations with many particles, optimizeParameters() method was implemented to temporarily store the best set of parameters into a file after each iteration where the file contains the current global best value, the current number of iteration, the current data set, and the current set of parameters. Therefore best results can be kept in case unexpected errors occur.

We conducted experiments to fix best values for the 9 parameters of McNN-PBL classifier using other meta-heuristic methods like Genetic algorithm and Ant Colony optimization algorithm. With the implementation of these methods there was increase in misclassification error compared to PSO. Hence we used PSO to fix the best values to improve the performance of McNN-PBL classifier.

G. Result Analysis

In our study, based on the selection of estimators and selection of depth, gender-wise variation in prevalence of depression is reported. As expected depression was more common in women compared to men and among those who were not in paid employment and is shown in Fig. 9 and variable 2 in Fig. 11. Unintentional weight loss is found more in men with BMI < 18.5 kg/m2. Exhaustion and muscle weakness are higher in men because 62% are in paid job in our study when compared to women. Slowness while walking is prevalent in women due to higher body mass. Frailtytot graph is as shown in Fig. 10. Pattern of depressive symptoms varied by gender: The levels of Guilt, appetite, irritability, interest and impaired concentration were more common among women; the levels of insomnia or sleep, tearfulness and enjoyment were prominent symptoms among men. Suicidality or wishing death are more prevalent in men than women as shown in Fig. 11. According to Fig. 12 parameter avggrip is found higher in men since the BMI of men fall within the healthy range. In the Fig. 13, 64% of men and 79.62% of women are obese. The rate of obesity in women is high since their BMI is higher compared to men. HTN is found to be more prevalent in women since the BMI does not fall under healthy range as in Fig. 13 and Fig. 14.

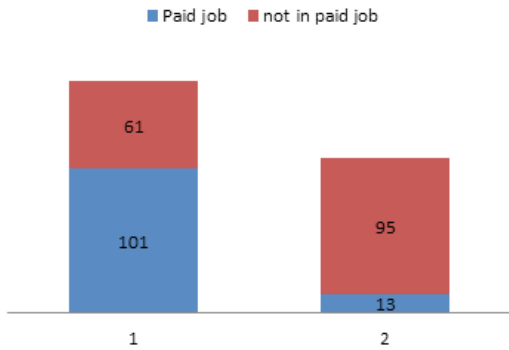


Fig. 9. Class wise distribution of Pjobcat2.

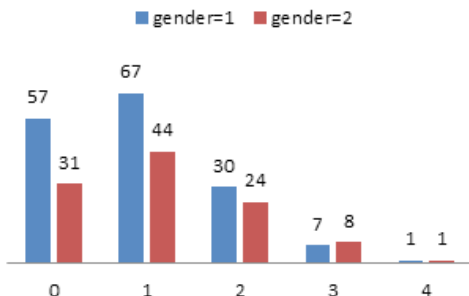


Fig. 10. Class wise distribution of Frifrailtytot.

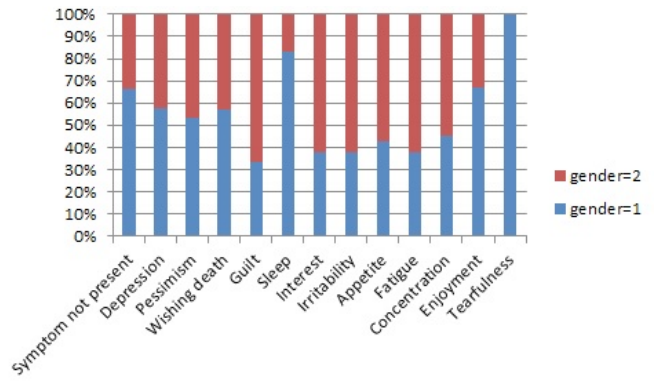


Fig. 11. Class wise distribution of Eurotot.

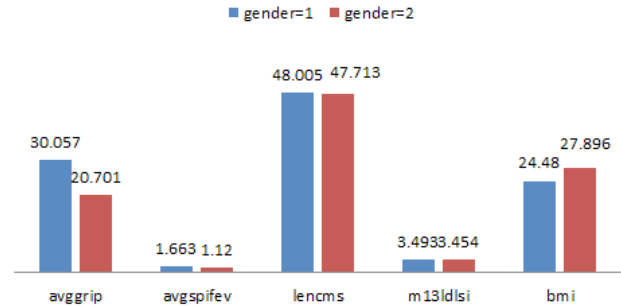


Fig. 12. Class wise distribution of other parameters.

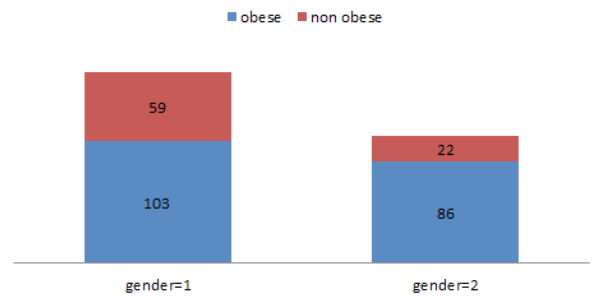


Fig. 13. Class wise distribution of Cobese.

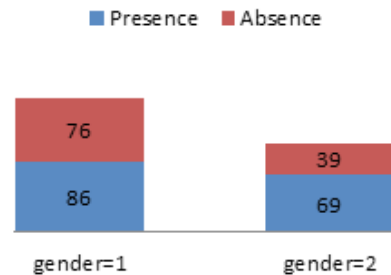


Fig. 14. Class wise distribution of HTN.

A 10 fold-Stratified cross validation evaluation technique was used to enforce the same distribution of classes in each fold of training, which handles the class imbalance problem. The average accuracy in this method is 97.80% and standard deviation of 2.98%. For each trial in the study, a ratio 75:25 of training and testing samples is randomly chosen.

IV. PERFORMANCE EVALUATION

The model was built using Java and the experiment on feature selection was conducted using Python running on Intel i3 processor. To evaluate the classifier performance, misclassification rate and correct classification rate play important roles. To find these we use sensitivity



and specificity values which are defined as in Table II. The parameters Eurotot, paid employment, Education and Obesity contribute to better classification that helps clinician to identify patients with depression. Adding these parameters to the model improved the efficiency of the classification.

TABLE II. CONFUSION MATRIX

Confusion Matrix Positive		Target			
		Negative	Positive		
Model	Positive	a	b	Positive Predictive value	$a/(a+b)$
	Negative	c	d	Negative Predictive value	$d/(c+d)$
		Sensitivity	Specificity	Accuracy $(a+d)/(a+b+c+d)$	

In the confusion matrix a is True Positive, the number of patients correctly classified as Depressed, b is True Negative, the number of patients correctly classified as Not Depressed, c is False Positive, the number of Not Depressed patients classified as Depressed and d is False Negative, the number of Depressed patients classified as Not Depressed. In disease diagnosis, sensitivity plays a significant role. Classifying a patient who is Not Depressed as Depressed is better than classifying a Depressed patient as Not Depressed.

The performance of the McNN-PBL classifier with and without boosted feature selection and feature selection using PCA on MYNAH Cohort is shown in table III and Fig. 15. McNN-PBL classifier picks out the best sequence of training records and produces a sensitivity of 94.55% on the testing records, 95.40% when PCA was used for feature selection and 97.8% when XGBoost technique was used for feature selection which indicates that the classifier performance improved with reduced misclassification rate. McNN-PBL algorithm used 70 neurons of which 28 were for class-1(Depressed) and 42 for class-2 (Not Depressed) in both training and testing phases. The model used 599 milliseconds while training phase and 16 milliseconds while testing phase to produce training accuracy of 98.45 percent and testing accuracy of 97.80 percent.

TABLE III. PERFORMANCE OF MCNN-PBL CLASSIFIER ON MYNAH COHORT

Testing Phase	Without feature selection (%)	With feature selection using pca (%)	With feature selection using xgboost (%)
Efficiency	89.40	91.00	94.57
Specificity	54.11	90.45	94.10
Sensitivity	93.40	94.55	97.80
Execution Time (milliseconds)	156ms	31ms	16 ms

After testing, the results were compared with the results of other algorithms including Self-adaptive Resource Allocation Network (SRAN), Extreme Learning Machines (ELM) and Support Vector Machines (SVM) and the results of the application of McNN-PBL with boosted feature selection using MYNAH cohort. All the results were compared based on properties like training efficiency, testing efficiency and number of neurons as shown in table IV.

From Table IV and Fig. 16, we can see that McNN-PBL classifier performed better than the best performing SRAN classifier on MYNAH cohort. McNN-PBL classifier also performed better than

ELM and SVM classifiers. Also as shown in Table III, the McNN-PBL classifier with boosted feature selection performed much faster than the McNN-PBL classifier without feature selection and feature selection using PCA.

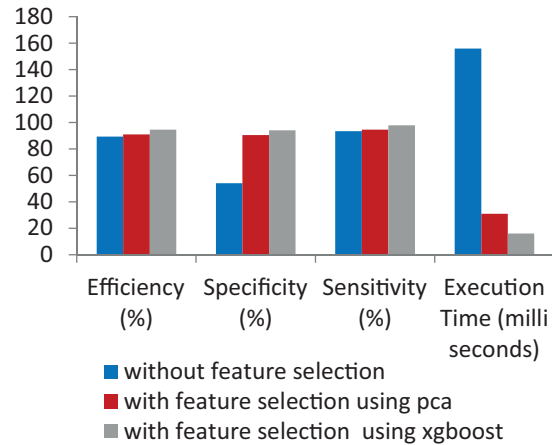


Fig. 15. Performance chart of the classifier.

TABLE IV. COMPARISON OF RESULTS USING MYNAH COHORT

Classifier	Training efficiency (%)	Testing efficiency (%)	Number of neurons	Training time in seconds
McNN-PBL	98.45	97.80	70	0.059
SRAN	98.00	92.29	69	0.088
ELM	100.00	90.67	120	0.150
SVM	100.00	90.62	110	0.140

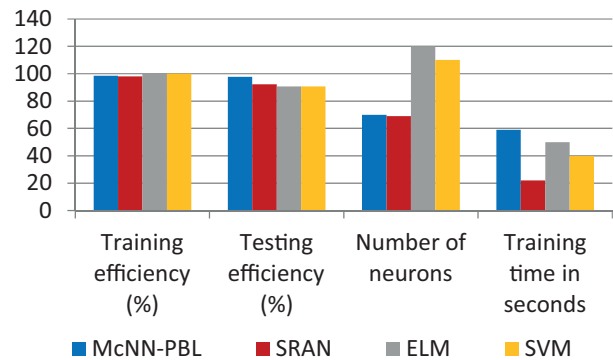


Fig. 16. Result comparison between different classifiers.

The limitation of the proposed model is that the process of optimization of parameters usually takes more time. Hence there is a need to store the optimized parameters. Therefore the optimized parameters are stored in training files so that the program only needs to run the parameter optimizing process once and reuse the optimized parameters in later executions for each training data file which slightly improves the execution time. To reduce this time complexity, other meta-heuristic methods have to be implemented to optimize the parameters.

## V. CONCLUSION AND FUTURE ENHANCEMENT

A novel approach is presented in this paper for an accurate classification of depression using XGBoost technique, McNN-PBL classifier and PSO algorithm. For performance analysis, MYNAH cohort with 1321 patient records and 1201 parameters was used. Using the boosted feature selection technique, the McNN-PBL classifier

showed significant testing efficiency. The model specifically helps the clinician identify depressed patients resulting in improved treatment and prevention of progression of depression leading to self harm, schizophrenia, obsessive-compulsive disorder, etc. Using this machine learning approach relative to other approaches like SRAN, ELM and SVM, we can better identify the significance of parameters and can define whether the patient has depression or not with less memory requirement and computation time and reduced misclassification error. An experiment on Particles Swarm Optimization (PSO) algorithm was done to use it to find the best parameters for McNN-PBL classifier which significantly improved the efficiency of the classifier. To improve the application performance in terms of speed and usability, as a future work, experiments may be conducted using meta-heuristic optimization methods namely Grey-wolf optimizer, Moth-flame optimizer and others. Another problem for future work would be to identify different types of depression based on the symptoms and other psychiatric diseases using MYNAH cohort and classify them using McNN-PBL classification model. Psychiatric disorders can be predicted using other classifiers and a comparative study can be done. The model can also be trained to predict other cognitive disorders using MYNAH cohort which helps in improved mental health of the people. For the large dataset we can run the model using parallel threads for better performance.

#### ACKNOWLEDGMENT

This study was funded by the Wellcome DBT India Alliance and supported through a core programme grant to the MRC Lifecourse Epidemiology Unit, University of Southampton by the Medical Research Council, UK and Department for International Development, UK to Dr. Murali Krishna. We are greatly thankful to the research assistants at CSI Holdsworth Memorial Hospital, Mysore for data management. We sincerely thank Dr. Suresh Sundaram, Associate Professor, School of Computer Engineering, Nanyang Technological University, Singapore and Dr. B. S. Mahananda, Associate Professor, Department of Information Science and Engineering, SJCE, JSS S&T University, Mysuru, India for their valuable suggestions during the execution of this study.

#### REFERENCES

[1] Vahia, Vihang N. "Diagnostic and Statistical Manual of Mental Disorders 5: A Quick Glance." *Indian Journal of Psychiatry*, May 2018.

[2] B.S. Mahanand, S. Suresh, N. Sundararajan, M. Aswatha Kumar, "Identification of brain regions responsible for Alzheimer's disease using a Self-adaptive Resource Allocation Network", *Neural Networks*, Volume 32, August 2012.

[3] G.Sateesh Babu, S.Suresh, "Meta-cognitive Neural Network for classification problems in a sequential learning framework", *Neurocomputing*, Volume 81, 1 April 2012.

[4] Tianqi Chen, Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System", *KDD '16 The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, CA, USA, August 13 - 17, 2016.

[5] Ismail Babajide Mustapha and Faisal Saeed, "Bioactive Molecule Prediction Using Extreme Gradient Boosting", *Molecules*, 2016.

[6] Zhuo Chen, Fu Jiang, Yijun Cheng, Xin Gu, Weirong Liu, et al., "XGBoost Classifier for DDoS Attack Detection and Analysis in SDN-based Cloud", *IEEE International Conference on Big Data and Smart Computing*, 2018.

[7] Xun Wei, Fan Jiang, Feng Wei, Jiekui Zhang, et al., "An Ensemble Model for Diabetes Diagnosis in Large-scale and Imbalanced Dataset", *Proceedings of the Computing Frontiers Conference*, Siena, Italy, May 2017.

[8] Haslam, Beck, "Categorization of major depression in an outpatient sample", *J. Nerv. Ment.*, 2012.

[9] Costafreda, S.G., Chu, C., Ashburner, J., Fu, C.H., "Prognostic and diagnostic potential of the structural neuroanatomy of depression", *PLoS One*, 2009.

[10] Hahn, T., Marquand, A.F., Ehlis, A.C., Dresler, T., Kittel-Schneider, S., Jarczok, T.A., Fallgatter, A.J., "Integrating neurobiological markers of depression", *Arch. Gen. Psychiatry*, 2011.

[11] V Arun, Arunkumar B.V., Padma S.K., Shyam V, "Disease Classification and Prediction using Principal Component Analysis and Ensemble Classification Framework", *International Journal of Control theory and applications*, ISSN:0974-5572, Vol 10, No. 14, 2017.

[12] Rubeena Banu; V Arun; N. Shankaraiah; Shyam V., "Meta-cognitive neural network method for classification of diabetic retinal images", *Second International Conference on Cognitive Computing and Information Processing (CCIP) 2016*.

[13] Mwangi, B., Matthews, K., Steele, J.D., "Prediction of illness severity in patients with major depression using structural MR brain scans", *J. Magn. Reson.*, 2012.

[14] Braeken, J., & Van Assen, An empirical Kaiser criterion. *Psychological Methods*, <http://dx.doi.org/10.1037/met0000074>, 2017.

[15] Liton Chandra Paul, Abdulla Al Suman and Nahid Sultan, *Methodological Analysis of Principal Component Analysis (PCA) Method*, *IJCEM International Journal of Computational Engineering & Management*, Vol. 16 Issue 2, March 2013.

[16] Bhagyashree, S.R., Nagaraj, K., Prince, Muralikrishna., "Diagnosis of Dementia by Machine learning methods in Epidemiological studies: a pilot exploratory study from south India", *Soc Psychiatry Psychiatr Epidemiol.* 2018.

[17] John RM Copeland, Martin Prince, K C M Wilson, "The Geriatric Mental State Examination in the 21st century", *International Journal of Geriatric Psychiatry*, August 2002.

[18] John RM Copeland, M. R. E. Dewey, Scott Henderson, "The Geriatric Mental State (GMS) used in the community: Replication studies of the computerized diagnosis AGE-CAT", *Psychological Medicine*, March 1988.

[19] Martin Prince, Daisy Acosta, Helen Chiu, "Effects of education and culture on the validity of the Geriatric Mental State and it AGE-CAT algorithm", *The British Journal of Psychiatry*, December 2004.

[20] Massy-Westropp, N. M., Gill, T. K., Taylor, A. W., Bohannon, R. W., & Hill, C. L., "Hand Grip Strength: age and gender stratified normative data in a population-based study", *BMC Research Notes*, 2011.

[21] Meurs M, Groenewold NA, Roest AM, et al., "The associations of depression and hypertension with brain volumes: Independent or interactive?" *NeuroImage : Clinical*, 2015.

[22] Lally F, Crome P. "Understanding frailty". *Postgraduate Medical Journal*, 2007.

[23] <https://en.wikipedia.org/wiki/Spirometry>.

[24] Agarwal A, Agarwal M, Garg K, Dalal PK, Trivedi JK, Srivastava JS. "Metabolic syndrome and central obesity in depression: A cross-sectional study". *Indian Journal of Psychiatry*, 2016.

[25] Broekman BF1, Chan YH, Chong YS, Kwek K, et al., "The influence of anxiety and depressive symptoms during pregnancy on birth size", 2014.

[26] <https://www.greatplainslaboratory.com/articles-1/2015/11/13/the-implications-of-low-cholesterol-in-depression-and-suicide>.

[27] Leonore M. de Wit, Annemieke van Straten, Marieke van Herten, Brenda WJH Penninx, and Pim Cuijpers, "Depression and body mass index, a u-shaped association", *BMC Public Health*, Longitudinal assessment of neuropsychological functioning major depression. *Aust. N. Z. J. Psychiatry*, 2009.

[28] Ramasamy Savitha, Sundaram Suresh, Narasimhan Sundararajan, "Projection-Based Fast Learning Fully Complex-Valued Relaxation Neural Network", *IEEE Transactions On Neural Networks And Learning Systems*, Vol. 24, No. 4, April 2013.

[29] G. Sateesh Babu, R. Savitha, S. Suresh "A projection based learning in Meta-cognitive Radial Basis Function Network for classification problems", *International Joint Conference on Neural Networks (IJCNN)*, 2012.



Vanishri Arun

Vanishri Arun received Bachelor of Engineering degree in Computer Science and Engineering in the year 1995 and Master of Technology in Software Engineering in the year 2009 from Sri Jayachamarajendra College of Engineering, Mysuru, India and currently pursuing Ph.D from Visvesvaraya Technological University, Belagavi, India in Biomedical Informatics under the guidance of Dr.

Shyam V and Dr. S K Padma and medical experts Dr. Arunkumar B V and Dr. Murali Krishna. She is working in the capacity of Assistant Professor in the Department of Information Science and Engineering at JSS Science and Technology University, Mysuru, India. She has teaching experience of about 12 years at technical institutes and has industrial experience of about 8 years in software development. Her research interests include Bio-Medical Informatics, Machine Learning and Data Analytics.



Shyam V.

Dr. Shyam holds a doctorate in Real Time Embedded Systems, specialising in Parallel Computer Architecture from Indian Institute of Science (gold Medal) & has been with CG Smith, Ericsson, Tata Consultancy Services and as Director of Technology at Philips. He was instrumental in devising one of the first hardware based antivirus solution in 90s. He currently holds the posts of President and CTO at

Forus Health (P) Ltd, Director at MYMO Wireless Technology (P) Ltd at SID – IISc, Bangalore and Technical Director at Maastricht University Medical Center in Netherlands. His innovation 3nethra, all in one eye screening device, at Forus has won several laurels like DST Locked Martin Gold medal, Perimal Award, Sankalp Award, CNBC TV18 Award, Samsung Innovation Quotient Award, NASCOM Innovation Award and Anjani Mashelkar Inclusive Innovation Award for 2011. He has filed more than 20 international patents, published around 50 papers in international conferences and journals.



Murali Krishna

Dr Murali Krishna is a Consultant in Old Age Psychiatry and Psychiatric Epidemiologist. He graduated from the Government Medical College, Mysuru, India and completed his specialist training in Psychiatry from University of Cambridge and Liverpool UK. He finished his higher specialist training in Old Age Psychiatry from University of Liverpool UK. He is a member of the Royal

College of Psychiatrists UK. He has an advanced Masters Degree in Research from University of Sheffield. He is a honorary Senior Lecturer at Faculty of Health and Social Care, Edge hill University in England and a honorary Research Fellow from School of Social Sciences, University of Bangor, Wales. His current research work examining life course risk factors for late life neurocognitive disorders is supported by a prestigious award from Welcome Trust UK and DBT India.



Arunkumar B V

Dr. Arunkumar received MBBS in the year 1996 and Diploma in Anaesthesiology in 1999 from JSS Medical College, Mysuru, India, Diplomate of National Board in Anesthesia (DNB) from National Board of Examination, Ministry of Health, New Delhi in 2004 and EGMP on Business Administration from IIM-Bangalore in 2016. He is currently heading the Department of Anesthesiology,

Apollo BGS Hospitals, Mysuru, India. His research interests are Medical Informatics, Data Analytics and Cognitive science.



S K Padma

Dr. Padma has pursued Ph.D. in Computer & Information Sciences. Currently a professor in the Department of Information Science and Engineering, JSS Science and Technology University, Mysuru, India and has been guiding many research scholars in the areas of Biomedical Informatics, Natural Language Processing, Machine Learning, Data Analytics and others. She is a life member

of Indian Society for Technical Education (ISTE), New Delhi, India, Institution of Engineers, India and Computer Society of India.