ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# ROBUST FEATURE SELECTION WITH CHICKEN SWARM INTELLIGENCE IMPROVED MULTILAYER PERCEPTRON FOR EARLY DETECTION OF MENTAL ILLNESS DISORDER

**S. SARANYA [1], N. KAVITHA[2]**

[1]Research Scholar, Dept. of Computer Science,
Nehru Arts and Science College/Assistant Professor,
Dr. N. G. P. Arts and Science College, Coimbatore, India
[2]Associate Professor and Head, Dept. of Computer Science,
Nehru Arts and Science College, Coimbatore, India
Email: [1]ssaranyamphil@gmail.com

**ABSTRACT**

Mental illness is a kind of health condition highly impacts a person's emotions, mind and their behavior. When mental disorder is not diagnosed at its earlier condition it results in depression, anxiety, schizophrenia, autism etc., are extremely widespread today. Recently many machine learning methods have been developed to help experts such as psychologists' and psychiatrists' for making decision regarding mental health of patient based on historical data of the patients. However conventional machine learning methods require optimal performance of feature engineering to obtained better performance and reduce computation complexity. This paper focuses on developing improved multilayer perceptron its objective is to handle over fitting, fine-tuning parameter values by integrating optimized feature selection and adopting mimetic meta heuristic model. LASSO (Least Absolute Shrinkage and Selection Operator) model-based feature selection is used for discovering best feature subset because it controls over fitting and prediction error with the help of ridge regression. In improved MLP instead of using trial and error based random parameter value assignment it inherits the Intelligence of Chicken Swam behavior. The best fittest value is assigned to the parameters of MLP (Multi-Layer Perceptron) to improve the accuracy rate of mental illness disorder detection. The simulation results are conducted on OSMI 2019 dataset and the outcome proved that CSI-MLP accomplished highest rate of accuracy in early detection of mental illness disorder compared to other conventional models.

**Keywords:** *Chicken Swarm Intelligence, LASSO Regression, Mental Illness Disorder, Multilayer Perceptron, Optimization Ridge Regression.*

## 1. INTRODUCTION

Mental illness of a person is the state of imbalanced outcome in brain chemistry. A person with healthy mind can able to realize their abilities, handle stress that occurs in a normal life, able to work productively and can contribute more to their community [1]. According to the report of WHO due to mental health problems nearly one trillion dollars in productivity loss around the world. Most individuals are affected by depression due to different reasons [2]. In India one in seven persons suffered from mental disorders of different severity with depression being the most common mental disorder. Depression contributed 33.8 % of all mental disorder [3]. Mental illness is increasing at a tremendous rate and it is time to upgrade mental health services, remove social stigma, creating awareness and generate access to treatment. It is very essential to diagnose and offer treatment to the peoples with mental illness by detecting them at its earlier stages.

In reason days usage of data mining and machine learning is greatly increased in medical care services, particularly in mental health. Data can be collected from different sources to maintain the patient's history which can assist the data scientist to detect mental disorder in its preliminary stages. These prediction models can be used to discover likelihood of patients seeking treatment with a highest accuracy by utilizing a self-reported questioner. But handling voluminous data may leads to increased cost to overcome this problem it is essential to recognize and utilize relevant features in the dataset. Detecting mental issue is a challenging task because misdiagnosis can result in serious problem. Thus, appropriate construction of prediction/classification model is vital to detect and treat mental disorder more accurately.

## 2. RELATED WORK

TukaAlhanai et al [4] in their study performed automated detection of depression using Long Short-Term Memory. They conducted interview with 142 individuals and agents by conducting interactions with text and audio features. The explicit interaction using LSTM neural network screens the depression detection.

YangLe et al [5] developed a multi modal depression model using deep neural network and convolutional neural network. They used text streams, audio and video to design the modality to receive compact dynamic information.

Alonso et al [6] in their work performed survey of mental health prediction using various algorithms such as support vector machine, association rule mining, K-nearest neighbor and Randomization.

Karthik and Sudha [7] designed a rank-based gene biomarker detection and classification to discover the overlapping as well as non-overlapping patterns of gene for bipolar disorder detection. In this work Gene expression omnibus dataset is used.

Chancellor and De Choudhury [8] explained about the data annotation, quality enhancement, feature subset selection and model development and validation for mental health illness. They discovered concern trend and lack of reflection in the technique of mental disorder prediction.

Enrique et al [9] reported the survey using sensor and machine learning model for mental health monitoring system. They concentrated on condition of mental disorder like stress, depression, bipolar disorder, anxiety, etc. Developed a classification nomenclature to direct the overall phases of the mental health monitoring.

Mirandaet al [10] designed a markov chain model for detection of anxiety with short term wearable based on the heart rate and dermal activity of both male and female.

Siirtola [11] signed a wearable sensor for measuring the migraine chronic using sleep data using various metrics such as blood volume pulse, acceleration, heart rate, its variability and temperature. They used quadratic discriminant analysis for classification.

Wawer et al [12] in their work used text mining to detect anxiety by applying bag of words, vectors based on dictionary and machine learning models. Deep learning methods with text representation infers the anxiety based on context level analysis.

## 3. METHODOLOGY

The major challenge in accurate prediction of mental state is due to its inconsistency and vague information presented in the dataset. When the redundant and irrelevant attributes in the dataset is not properly treated, the prediction model will be greatly influenced and results in lack of accurate classification. Hence, it is very essential to identify potential attributes which contributes more information in classification of mental state. In standard Multilayer perceptron, the detection rate and error rate mainly depend on the hyper parameters such as learning rate, weight and bias values are assigned in a random manner. These parameters based on the error rate obtained they are adjusted in a trail and error basis. This results in overfitting and lapse in performance of the prediction model. To improve the learning rate of the MLP, in this work the parameters are fine-tuned by applying the intelligence of chicken swarm intelligence. These are main motivation of this paper, to design and develop robust feature selection with chicken swarm intelligence improved multilayer perceptron for early detection of mental illness disorder.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



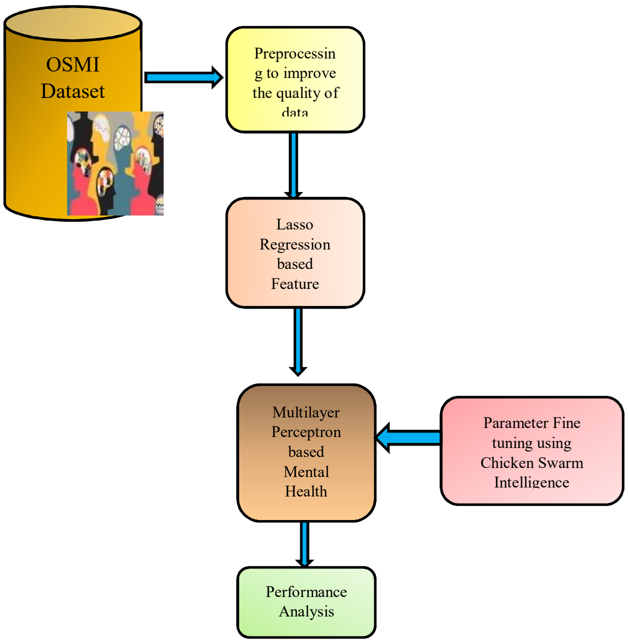


*Figure 1: Overall Framework of the Chicken Swarm Intelligence Improved Multilayer Perceptron*

The main objective of this research work is to predict the mental health disorder at the earliest to enhance the life quality of the patients. The data is preprocessed to enrich the data quality by performing normalization and discretization. The dataset used in this work is collected from OSMI 2019 Mental Health in Tech Survey Results of Kaggle repository [13]. The dataset comprised of 4218 instances with 20 attributes are available. The significant attributes which contribute more in classification of presence of mental disorder is determined by applying Lasso Regression based feature subset selection. The prediction of healthy or presence of mental disorder is done by designing improved multilayer perceptron with meta heuristic algorithm to fine tune the parameters. The overall architecture is depicted in the Figure 1.

### 3.1 Dataset Description

The dataset used for predicting the mental disorder is collected from Open Sourcing Menta Illness (OSMI) Mental Health in Tech Survey 2019, with 4218 instances and 20 attributes [13]. The dataset comprised of details about the working people's aims to provide awareness about mental illness at its earliest which will be beneficial to both employees and employers. The list of attributes in OSMI dataset is shown in the Table 1.

*Table 1: Description of the Dataset*

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

### 3.2 Data Preprocessing

The features are divided into categorical and numerical features. Using Z-Score normalization the numerical values are scaled and the categorical values are encoded by OneHotEncoder. The 60% of the dataset is used for training the model and remaining 40% is used for testing the performance of the trained model. The mathematical representation of Z-Score Normalization is

$$\text{ZSN}\ (\text{Att}_{i=1..m}) = \frac{Att_i - \overline{Att}}{\sigma(Att)} \quad (1)$$

Where $\text{Att}_i$ is the individual attribute, $\overline{Att}$ is the mean value of the concern attribute column values and $\sigma(Att)$refers to standard deviation of the concern attribute values.

### 3.3 Feature Selection using LASSO Regression Method

To discover relationship among one or more independent attributes the regression model uses statistical method. Least Absolute Shrinkage and Selection Operator (LASSO) [14] is kind of regression model which control the problem of overfitting in prediction model is used in this research work for feature selection. The main reason behind feature selection using LASSO regression is because of its ability to remove variables which are redundant and does not produce any valuable information. Before Lasso model, stepwise selection is used to select covariates. While using that approach only in certain combination of covariates it will produce improved accuracy of prediction, in all other cases it results in increased prediction error. Ridge regression shrinks the sum of squares to reduce the overfitting but it does not perform perfect covariate selection.

This concept is adapted in Lasso which incorporates both the goals by making sum of absolute value of the regression coefficient to be less than a fixed value, which consequences in obliging specific coefficient to zero, by successfully not including them.

In LASSO based feature selection it chooses the variables that have non-zero coefficient even after applying the shrinking process. This will minimize the prediction error. By performing shrinking and eliminating the coefficients can reduce difference without a considerable increase of bias [15]. The tuning parameter λ plays a vital role because when its value increases the bias will be increased and variance will be decreased. By eliminating irrelevant attributes that are not associated with response attribute will results in reduction of overfitting. This is the major reason to use LASSO as feature selection method in this work for mental illness disorder detection

***Procedure for LASSO based Significant Feature Selection***

OSMI Dataset: N cases; E: No of covariates (attributes), R Output (class)

$X_i = \{x_1, x_{2..}, x_E\}$// covariate vector of ith instance, $R_i$ - output of $X_i$

The LASSO model estimates each attribute as follows

$$\beta(\lambda) = \underset{\beta}{\operatorname{argmin}} \left( \frac{\left\|R - X_\beta\right\|_2^2}{n} + \lambda_2\|\beta\|^2 + \lambda_1\|\beta\|_1 \right) \quad (2)$$

Where$\beta$ refers to regression coefficient, regularization parameters are $\lambda_1$and $\lambda_2$

Output:$\beta(\lambda_{i=1\ldots E})$// significant parameters

Thus, using the LASSO Regression method attributes in the OMSI dataset is removed when their $\beta(\lambda)$coefficient is equal to zero.

### 3.4 Multilayer Perceptron

The perceptron is an algorithm which classifies the input by discriminating two categories using a straight line which is termed as linear classifier [16]. The input is set of attributes (a) of the dataset represented in the vector format and it is multiplied by the weights (wt) and summed with the bias (bs) value. It is represented as

$$Y = a * wt + bs \quad (3)$$

It produces a single output depending on many inputs by forming a linear combination using the input weights and passing it through the non-linear activation function denotes as

$$Y = \phi(\textstyle\sum_{i=1}^{n} wt_i\, v_i + bs) \quad (4)$$

Where wt is the weight vector, V is the input vectors, bs refers to bias and $\phi$ is the activation function. The single layer perceptron is shown in the Figure 2.

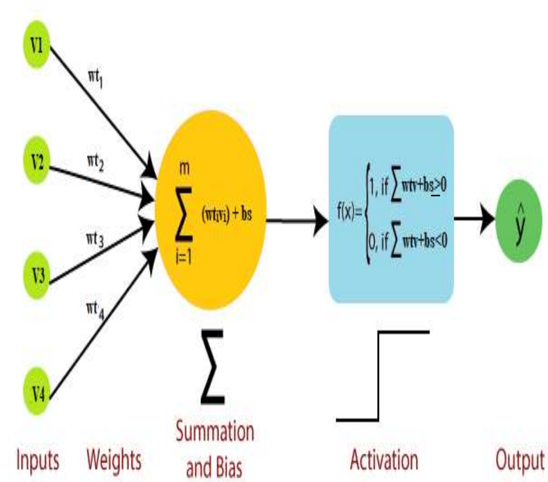


*Figure 2: Single Layer Perceptron*

Journal of Theoretical and Applied Information Technology
31st May 2022. Vol.100. No 10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A multilayer perceptron (MLP)is a kind of deep artificial neural network, which comprised of many perceptron as illustrated in the Figure 3. It is structured with an input layer that receive the attribute values, an output layer that involves in prediction about the corresponding input and in-between these two layers are random number of hidden layers which are the real computation engine of the MLP. It belongs to supervised learning model as it trains with set of input-output records and learns to model the dependences or correlation among them. During this learning phase, it involves in parameters adjustments (i.e) weights and bias values, to reduce the error. In conventional MLP, the weights and bias parameter values are adjusted using Backpropagation method based on the error. The errors are measured in different ways including Root Mean Square Error (RMSE).

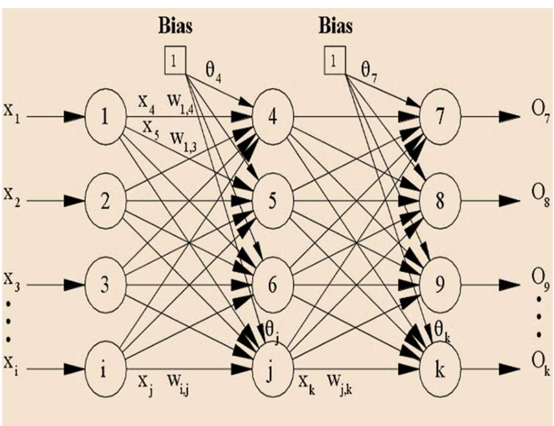


*Figure 3: Multilayer Perceptron*

Hence, this research work concentrates to improve the performance of the MLP instead of using backpropagation-based parameter adjusting in a trial-and-error manner. The chick swarm intelligence is adapted to discover best optimal values that can be assigned to the weights and bias of the hidden layer during the learning phase of the MLP. The working principle of Chicken Swarm intelligence enhanced MLP is discussed in the following section.

### 3.5 Chicken Swarm Intelligence

The chicken Swarm Intelligence is developed based on the inspiration of the social lives of chickens which plays a vital role by framing hierarchical order [17]. The preeminent chickens in a flock will dominate the weak chickens. The behaviors of chickens vary based on the gender. In each group rooster will the head that positively search food, fight with other chickens from different groups when the invade their territory. More dominant hens would be more consistent with the rooster head to food foraging. The submissive hens or chickens will be at the peripheral of the group to pursuit for food. Chicks search for food around their mother hen, and they are cooperative to each other. Thus, in the chicken swarm hierarchy rooster is the leader, mother hens are co-leaders, hens are elders and members of the swarm are chicks, depending on the fitness value there are sorted from top to bottom hierarchy respectively as shown in the Figure 4.

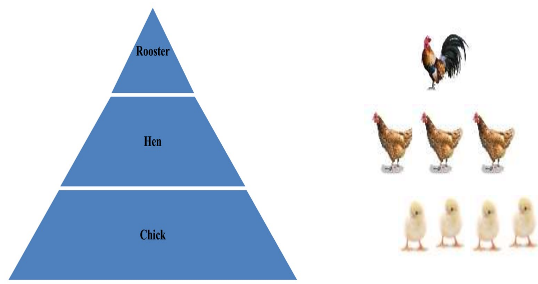


*Figure 4: Chicken Swarm Intelligence Hierarchical Order*

#### 3.5.1 Movement of Chick for searching food

The Rooster with the best fitness value is treated as highest priority and assigned as the header, the worst fitness value will be assigned with least priority, the best fittest roosters have more searching space while worst fittest hens will be searching with the limited space. It is mathematically characterized as

$$c_{i,j}^{t+1} = c_{i,j}^{t} * (1 + rd(0, \sigma^2)) \tag{5}$$

$$\sigma^2 = \begin{cases} 1, & if ft_i \le ft_m, \\ exp\left(\frac{ft_m - ft_i}{|ft_i| + \varepsilon}\right), & else, m\epsilon[1, N], m \ne i \end{cases} \tag{6}$$

Where $\sigma^2$ is the standard deviation, the gaussian distribution is $rd(0, \sigma^2)$ whose mean value is 0, $\varepsilon$ is the small constant value used to avoid divide by zero error. m is the arbitrary rooster chosen from the group of roosters, ft signifies the fitness value of relevant rooster c.

The food searching nature of the dominant hens are expressed as

$$c_{i,j}^{t+1} = c_{i,j}^{t} + v1 * rd * \left(c_{r1,j}^{t} = c_{i,j}^{t}\right) + v2 * rd * (c_{r2,j}^{t} = c_{i,j}^{t}) \tag{7}$$

$$v1 = \exp((ft_i - ft_{r1})/(abs(fi_i) + \varepsilon)) \tag{8}$$

$$v2 = \exp((fit_{r2} - fit_i)) \quad (9)$$

Where r1 $\in$[1, …, N] is the rooster index of the i[th] hen's group mate, r2 $\in$[1, …, N] is the rooster or hen's index, which is indiscriminately chosen from the swarm such that r1 $\neq$ r2. Random uniform number is denoted by rd whose value lies in the interval [0,1].

Noticeably, $ft_i > ft_{r1}, ft_i > ft_{r2}$, thus v2 <1<v1, v1 and v2 have different formula due to the presence of competition within a group. The fitness value of chickens relative to the rooster fitness value are treated as competition among them in a group. If v2 = 0 then the i[th] hen would search for the food within its territory. The rooster fitness value will be unique in each group.

The movement of chicks for searching of food will be around their mother hen. It is mathematically modeled as

$$c_{i,j}^{t+1} = c_{i,j}^{t} + CL * (c_{m,j}^{t} - c_{i,j}^{t}) \quad (10)$$

Where i[th] chick's position of the mother is denoted as $c_{m,j}^{t}$, CL represents chick would follow its mother to forage for food. CL of every chick would be arbitrarily selected from 0 and 2.

This chicken swarm intelligence is considered as the objective problem to be solved to reach optimization in parameter fine tuning in Multilayer Perceptron for prediction of Mental illness Disorder.

**Algorithm 1: Chicken Swarm Intelligence Enhanced Multilayer Perceptron to Predict the Mental Illness Disorder**

Input: OSMI dataset DS, X is the input attributes, Y class variable of the relevant record
Output: Classification of presence or absence of mental illness

1. Select the initial parameters for weight vectors wt
2. While termination condition is not reached do
3. For all (X, Y) )$\in$ DS do
    a. Fed the input X to the network
    b. Compute the network output Y
    c. Compute the difference between expected and observed output
4. End for
5. Calculate the RMSE of the E(Y)
6. Entire training pattern weights are summed
7. Update the weights using the chicken swarm intelligence
    a. Call the Chicken Swarm Intelligence Algorithm
8. For all weights summing over all training patterns
9. Perform one update step of the minimization approach
10. End while

**Algorithm 2: Chicken Swarm Intelligence for Parameter Optimization**

Input: Weight vectors and bias vectors
Output: Best Weight and bias values

- Assign the N number of chicken's population and their relevant parameters
- Compute entire chicken's (N) fitness value and set t = 0
- While t < max-gen

If (t % G == 0) then

- Rank the fitness value of all the chickens
- Establish a hierarchal order in the swarm
- Split the swarm into various groups
- Discover the relation among chicks and mother hens within a group.

Endif
For j = 1 to N
IF j == rooster then

- Update the location of the rooster using the equation

$$c_{i,j}^{t+1} = c_{i,j}^{t} * (1 + rd(0, \sigma^2)) \quad (11)$$

Else if j == hen then

- Update the location of the dominate hen using the equation

$$c_{i,j}^{t+1} = c_{i,j}^{t} + v1 * rd * \left(c_{r1,j}^{t} = c_{i,j}^{t}\right) + v2 * rd * (c_{r2,j}^{t} = c_{i,j}^{t}) \quad (12)$$

$$v1 = \exp((ft_i - ft_{r1})/(abs(fi_i) + \varepsilon)) \quad (13)$$

$$v2 = \exp((fit_{r2} - fit_i)) \quad (14)$$

Else if j == chick then update position of chick using the equation

$$c_{i,j}^{t+1} = c_{i,j}^{t} + CL * (c_{m,j}^{t} - c_{i,j}^{t}) \quad (15)$$

End if
Compute the new search solution
IF the new solution is better than the previous, update it
End for
End while
End

The algorithm illustrates that after reducing the features using LASSO regression feature

Journal of Theoretical and Applied Information Technology
31st May 2022. Vol.100. No 10


ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

selection, they are passed to the improved multilayer perceptron known as chicken swarm intelligence enhanced MLP (CSI-MLP). The conventional MLP parameter are weights and bias whose values are assigned using backpropagation and based on the resultant prediction error, these parameter values are adjusted in the hidden layers in an iterative manner. To overcome this issue of random assigned of values to the parameters in MLP, the proposed CSI-MLP uses the chicken swarm intelligence, where the chickens in the swarm search for the best values to be assigned for the parameters such as weights and bias and thus it results in optimized prediction of mental illness disorder with the objective of reduce prediction error.

## 4. RESULTS AND DISCUSSIONS

This section discusses about performance analysis of mental illness disorder detection using proposed Chicken Swam Intelligence Enhanced Multilayer Perceptron (CSI-MLP) it is implemented using python code. The dataset is collected from OSMI 2019 KAGGLE repository. It comprised of 4218 records with 20 attributes. By applying LASSO regression-based feature selection it is reduced to 7 attributes and these are involved in prediction process using CSI-MLP as shown in the Table 2. The proposed model performance is assessed with three different existing classifications models namely K-Nearest Classifier, Support Vector Classifier and Decision Tree.

*Table 2: Performance Analysis of Feature Selection Models*

| | Accuracy | Precision | Recall | RMSE |
|---|---|---|---|---|
| KNC | 78.2 | 77.5 | 77.1 | 0.0153 |
| DT | 72.4 | 71.5 | 70.9 | 0.0294 |
| SVC | 80.3 | 80.7 | 80.2 | 0.0105 |
| MLP | 82.9 | 83.1 | 83.7 | 0.0101 |
| CSI-MLP | 96.5 | 97.1 | 97.3 | 0.0019 |

*Table 3: Performance Comparison of Classification Models for Mental illness Disorder Detection*

| Feature Selection Methods | No. of. Feature Selected | Features list |
|---|---|---|
| Best First Search | 13 | 7,3,1,12,6,8,2,14,9,11,10,5,4 |
| Ranker Search Algorithm | 12 | 14,6,4,3,2,5,7,12,8,11,10,9 |
| Lasso Regression | 7 | 1,3,7,9,10,11,12,14 |

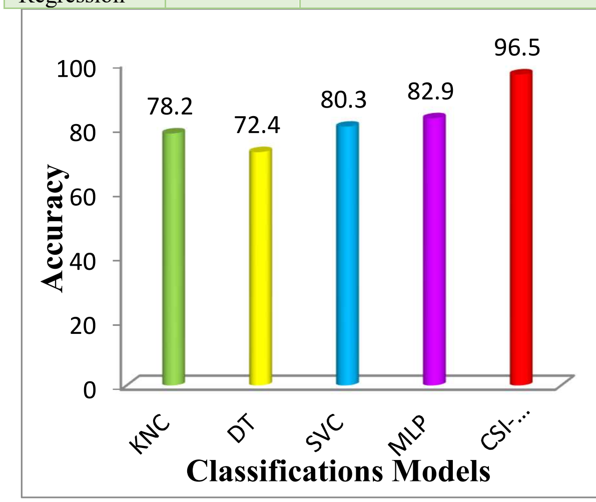


*Figure 5: Performance Comparison based on Accuracy for Mental illness disorder detection*

The Table 3 and Figure 5 illustrate accuracy comparison of four different classifications models to predict presence of mental illness. From the results generated it is observed that the proposed CSI-MLP produced highest accuracy compared to other three conventional models. The reason is problem of class imbalance is not well treated by existing models. While using CSI-MLP it determines highly significant attributes in the dataset using Lasso Regression Model. The MLP parameters are fine-tuned using Chicken Swam Intelligence. Thus, it achieves highest accuracy compared to other models.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

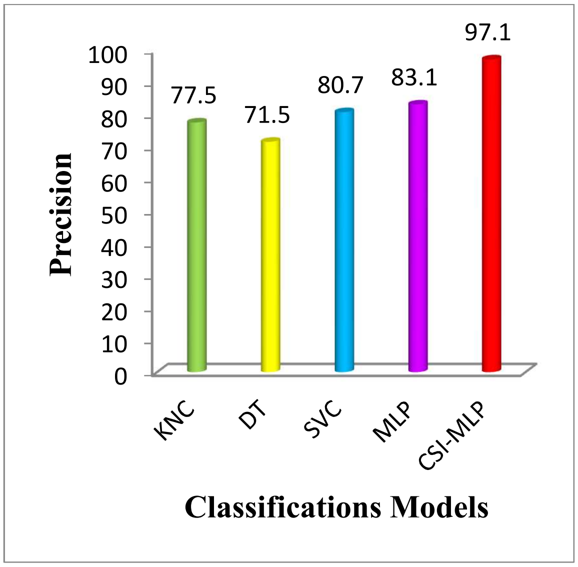


*Figure 6: Performance Comparison based on Precision for Mental illness disorder detection*

The precision rate obtained by four different prediction models for discovering mental illness is depicted in the Figure 6. It is observed form the assessment that the proposed CSI-MLP achieves its highest precision rate as it integrates LASSO feature selection along with mimetic heretical order of Chicken Swam Intelligence to optimize mental illness disorder.

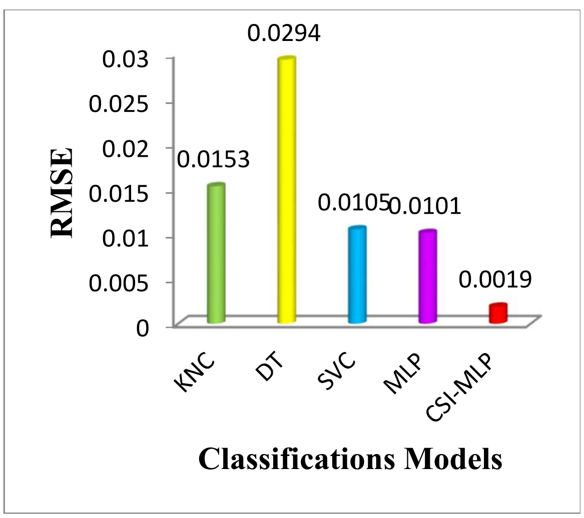


*Figure 7: Performance Comparison based on RMSE for Mental illness disorder detection*

The Figure 7 explores the comparison based on RMSE obtained by classifications models involved in mental illness disorder detection. The error rate of proposed CSI-MLP is very less as it treats both over fitting and class imbalance by focusing on highly contributed attributes and the optimal assignment of parameter values such as weight and bias in MLP. The conventional models suffer from low instances of classes with mental illness. Thus, they produce highest error rate compared to CSI-MLP.

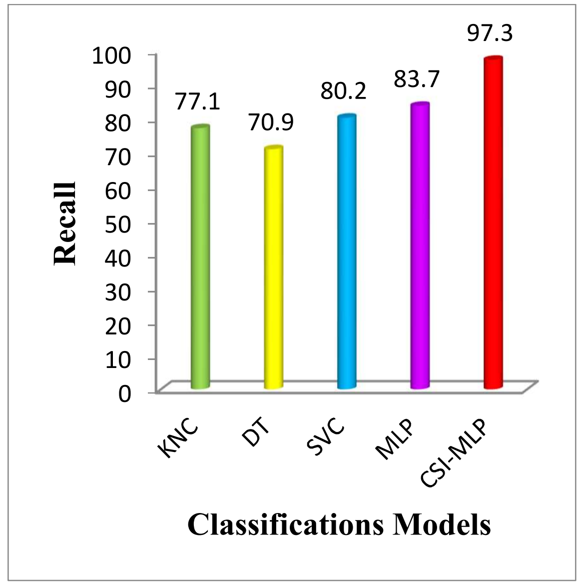


*Figure 8: Performance Comparison based on Recall for Mental illness disorder detection*

The recall metric comparison to detect mental illness disorder using four various prediction models is illustrated in the Figure 8 The attributes in the dataset are reduced by applying LASSO regression which controls the over fitting problem by shrinking penalty parameters. The MLP during its training stage instead of using random assignment of parameter values this proposed work adapt Intelligence of Chicken Swam based on its fitness value.

## 5. CONCLUSION

The existing algorithms lacks in handling the vague and inconsistent information presented in the mental dataset. This research work performs early detection of mental illness disorder by developed an improved multilayer perceptron which handles the voluminous dataset based on maximal relevancy and minimal redundancy. This newly constructed model involves significant attributes assessed using lasso feature selection method and fine-tuning its parameters with the knowledge of behavior-based metaheuristic approach. The proposed CSI-MLP overcomes the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

problem of over fitting by applying lasso regression algorithm and the attributes which doesn't contribute for predicting mental illness is removed from the dataset. The efficiency of multilayer perceptron is entirely dependent on the parameters which are highly influential for classification. These parameter values such as weight and bias are assigned optimally by adapting Chicken Swam Intelligence which fine tunes its performance. The result explores that the proposed CSI-MLP produced predominant contribution with highest accuracy to detect mental illness disorder in OSMI dataset 2019 compared to other three conventional classifications models with reduced misclassification error by integrating lasso regression and chicken swarm optimization.

**REFERENCES:**

[1] Alonso. S. G, I. Torre-Diez S. Hamrioui, M. l Lopez-Coronado, D. C. Barreno, L. M. Nozaleda, M. Franco. (September 2018) "Data Mining Algorithms and Techniques in Mental Health: A Systematic Review". *J. Med. Syst.* Vol. 42, 9, 1– 15.

[2] Chancellor, S., De Choudhury, M. (2020) "Methods in predictive techniques for mental health status on social media: a critical review". *npj Digit. Med*. 3, 43.

[3] EnriqueGarcia-Ceja, MichaeRieglera, Tine Nordgreen, Petter Jakobsen, Ketil J. Oedegaard, JimTorresen, (December 2018) "Mental health monitoring with multimodal sensing and machine learning: A survey", *Pervasive and Mobile Computing* Volume 51, , Pages 1-26

[4] Hanisch, S.E., Twomey, C.D., Szeto, A C H. (2016) "The effectiveness of interventions targeting the stigma of mental illness at the workplace: a systematic review". *BMC Psychiatry* vol 16(1),

[5] Hastie, Trevor. Tibshirani, Robert. Friedman, Jerome. (2009) "The Elements of Statistical Learning: Data Mining, Inference, and Prediction". *Springer*, New York, NY.

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

[7] Huang X, C. Ye, and J. Zheng, (Mar. 2018) ''Chicken swarm optimization algorithm of hybrid evolutionary searching strategy,'' *Comput. Eng. Appl.*, vol. 54, no. 7, pp. 176–181.

[8] Jiang, Yuan (2016). "Variable selection with prior information for generalized linear models via the prior lasso method". *Journal of the American Statistical Association*. 111 (513):

[9] Karthik, S., Sudha, M. (2020) "Predicting bipolar disorder and schizophrenia based on non-overlapping genetic phenotypes using deep neural network". *Evol*. Intel.

[10] Miranda. D, J. Favela, B. Arnrich, (2017) "Detecting anxiety states when caring for people with dementia, Methods" *Inf. Med*. 56 (1) 55–62

[11] Reddy U. S, A. V. Thota, A. Dharun, "Machine Learning Techniques for Stress Prediction in Working Employees," *2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, Madurai, India, 2018, pp. 1-4.

[12] Siirtola. P, H. Koskimäki, H. Mönttinen, J. Röning, "Using sleep time data from wearable sensors for early detection of migraine attacks", *Sensors* 18 (5) (2018).

[13] Theorell T, Hammarstrom A, Aronsson G, Bendz LT, Grape T, Hogstedt C, Marteinsdottir I, Skoog I, Hall C (2015) "A systematic review including meta-analysis of work environment and depressive symptoms". *BMC Public Health* 15(1):738

[14] TukaAlhanai, Mohammad Ghassemi, James Glass, "Detecting Depression with Audio/Text Sequence Modeling of Interviews" *Proc. Interspeech* 2018, pp 1716-1720.

[15] Wawer, A., Chojnicka, I., Okruszek, L, LukaszOkruszek, Justyna Sarzynska-Wawer, "Single and Cross-Disorder Detection for Autism and Schizophrenia". *CognComput* (2021).

[16] YangLe, Dongmei Jiang, Xiaohan Xia, Ercheng Pei, Meshia Cedric Oveneke, Hichem, "Detecting Depression Using Machine Learning, Multimodal Measurement of Depression Using Deep Learning Models" *In Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge – AVEC 2017*, pp 53–59.

[17] Zou, Hui; Hastie, Trevor (2005). "Regularization and Variable Selection via the Elastic Net". *Journal of the Royal Statistical Society. Series B (statistical Methodology). Wiley*. 67 (2): 301–20.