https://doi.org/ 10.33472/AFJBS.6.10.2024.728-737

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Automated Staging of Lupus by collected symptom dataset using MLP

P. Saha[1] *, Dr. R.K. Mandal[2]

[1] Department of Computer Science, North Bengal St. Xavier's College, West Bengal, India
[2] Department of Computer Science & Technology, University of North Bengal, West Bengal, India
e-mail: payel17.10@gmail.com, rakesh_it2002@yahoo.com

Article History
Volume 6,Issue 10, 2024
Received:17 Apr 2024
Accepted : 05 May 2024
doi: 10.33472/AFJBS.6.10.2024.728-737

Abstract

Lupus is one of the most serious disease found mostly in the women. Most of the patients of Lupus are infected withLupus nephritis. Many studies have been carried out so far by the medical experts for proper diagnosis and treatment manually. Some studies are intended towards diagonally the disease at an early stage by using gene dataset and mostof the diagnosis is based on symptoms of the disease. But it is very difficult for the medical experts to diagnose the disease at an early stage to treat proper treatment and management could be provided to the patient. This paper is an approach towards the formation of a new dataset by augmenting the already existing gene dataset with the symptom dataset that has been collected from various medical institutions. The new formed dataset is presented to a MLP which has been developed here as this is capable of automating the staging of Lupus. This method minimizes the time and effort to diagnose the disease with great accuracy, when compared with the already existing ML methods viz SVM, CNN. It has been found that recommendation wise the method developed here is better.

Keywords SLE, Perceptron, ML, CNN, ANN

1 INTRODUCTION

Systemic Lupus Erythematosus
(SLE) is a disease which is chronic and autonomous in nature (Katherine G. Skocelas, 2018). SLE mainly affects younger women but it may occurs in 20% of patients who are 50 years or above. It may affect almost all the symptoms in a human body and the diagnosis is based on America College of Rheumatology (Gurevitz et al,2013). Systemic Lupus Erothematosus (SLE) has a disease course which is not predictable (Janri et al, year). British Isles Lupus Assessment Group (BILAG) is a computerized index which measures chemical activities of disease in SLE(Hay et.al, 2013). There are six symptoms which defines and evaluates disease activity viz The Ropes System, The New York Hospital for Special Surgery System, BILAG Scale, the University of Toronto SLE Disease Activity Index [SLE-DAI] and the systemic Lupus Activity Measure [SLAM] (LIANG et al, 1989). Logistic Regression was used to find out the mobility to prescribe major immune suppressive drugs (Bello et al, 2016).
An early diagnosis of SLE helps in SLE major management and treatment of SLE. Different researchers have used many computational of SLE. A novel ML solution was constructed to generate a model which predicts SLE symptom

severity levels (skocelas et al, 2018). Artificial Neural Network (ANN) was used to predict lupus nephritis in patients with SLE (Rajmehr et al, 2002). Artificial Neural Network Ngyyen Window algorithm was implemented to predict lupus(Aisyah et al, 2019). An RNN-LSTM based deep learning methodology was used to predict hospital radmission for lupus patients (Reddy et al, 2018)

Systemic Lupus Erythematosus is known as SLE or lupus. It is an autoimmune disorder which affects multiple organs. Research works have been carried out to predict Lupus by using various machine learning methods. Identifying different stages of Lupus is very effective in clinical decision making and patient management system. The Artificial Neural Network (ANN) has been applied in different fields. The present research work has developed a multi class perceptron model of neural network. To predict the stages of Lupus the accuracy of the developed model is better than the already developed models. Since this is rare disease, the data set is limited. 150 data have been collected from individual lupus patient through survey process. Different symptoms have been considered in the lupus stage identification process such as weight loss (W_l), Fatigue (F_a), Fever (F_e), Cytoid bodies (C_y), Hemorrhages (H_e), Papillitis (P_a). Stages of lupus patient have been considered in this research paper as normal, moderate and severe. The outcome is compared to other models.

II. RELATED WORK

During the time of this research work it was observed that research work has already been done to suspect various diseases like cancer, dengue etc using ANN. G.K.W. Lam, M.Petri worked on lupus disease[1]. Payel Saha et al have used ANN to detect dengue disease[5]. Farhad Soleimanian Gharehchopogh et al have used ANN to diagnose thyroid disease[2]. Dey et al [3] has used ANN techniques to diagnose Diabetes disease. The applied data in this paper have been collected from Sikkim Manipal Institution of Medial Science Hospital which includes530 patients. The output includes 2 classes of 0 or 1. They suggested two feed forward ANN architectures where the first one includes the number of neurons in three layers (6-10-1) and the second one involves two hidden layers and the number of neurons in (6-14-14-1) layers. F.S.Gharehchopogh et al. have used ANN to diagnose heart disease[4]. They used MLP ANN with 60 nodes in input layer, 4 nodes in hidden layer and 2 nodes in output layer which is back propagation learning algorithm. Matthew H. Liang et al worked on reliability and validity of six systems of disease activity in SLE[6]. Another study was conducted by Thomas Stoll et al. on BILAG disease activity index in SLE patients[7]. E.M. Hay et al studied the BILAG index[8]. Garys Firestein et al have done research on Lupus[9]. D.P.M Symmons et al worked on clinical disease activity in SLE[10]. Annegret Kuhn et al. have done research work on the classification and diagnosis of cutaneous lupus erythematosus[11]. Claire Bombardia studied on SLEDAI disease activity index of SLE[12]. Basic study on ANN was conducted by Laurene Fausett[13].

III. METHODOLOGY

a. Multi-Class Perceptron: The multi-class perceptron algorithm is a supervised learning algorithm for classification of data into one of a series of classes. This algorithm is built in such a way that it can be generalized to any use-case, with details on how to format data in the sections below. It is meant to be easy to use and understand, without any significant performance issues. For added benefit, this module also contains functions to facilitate training, building, and testing the classifier, providing useful metrics and statistics to judge performance.

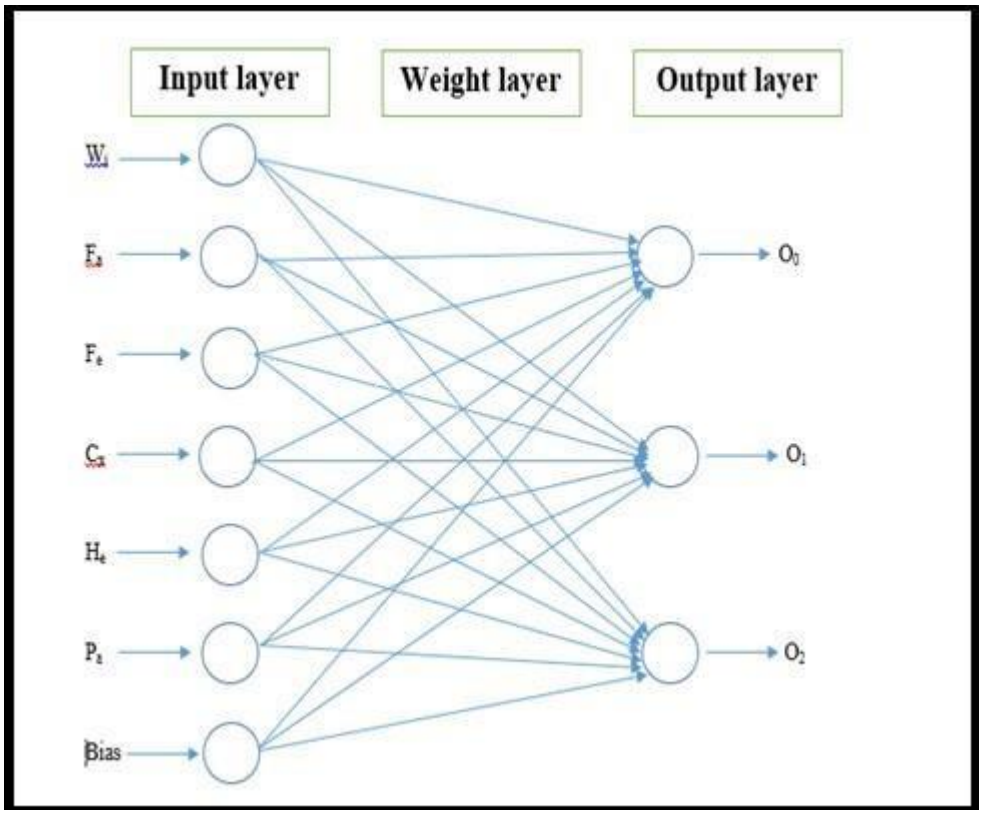


Fig 1: Neural Network for Multiclass perceptron

b. Algorithm Summary: This algorithm, like most perceptron algorithms is based on the biological model of a neuron, and its activation. In the case of a normal perceptron (binary classifier), the data is broken up into a series of attributes, or features, each with a specific value. When this feature vector is received by the artificial neuron as a stimulus, it is multiplied (dot product) by a weight vector, to calculate the activation value of the specific data point. If the activation energy is high enough, the neuron fires (the data meets the classification criteria).

In the case of a multi-class perceptron, things are a little different. The data comes in the same way, but instead of the respecting feature vector being multiplied by a single weight vector (for a single class), it is multiplied (dot product) by a number of weight vectors (a separate vector of weights for each unique class). Whichever weight vector that yields the highest activation energy product is the class the data belongs to. This decision process is known as the Multi-Class Decision Rule.

$$\mathbf{Y = W_y . f(x)\ W_y = W_y}$$

$$\mathbf{- f(x)\ W_y = W_y + f(x)}$$

W_y is the weight vector, f(x) is the feature vector

c. Training Process: To train the algorithm, the following process is taken. Unlike some other popular classification algorithms that require a single pass through the supervised data set (like Naive Bayes), the multi-class perceptron algorithm requires multiple training iterations to fully learn the data. The iteration count can be easily set as a parameter.

During each iteration of training, the data (formatted as a feature vector) is read in, and the dot product is taken with each unique weight vector (which are all initially set to 0). The class that yields the highest product is the class to which the data belongs. In the case this class is the correct value (matches with the actual category to which the data belongs), nothing happens, and the next data point is read in. However, in the case that the predicted value is wrong, the weight vectors are corrected as follows: The feature vector is subtracted from the predicted weight vector, and added to the actual (correct) weight vector. This makes sense, as we want to reject the wrong answer, and accept the correct one.

After the final iteration, the final weight vectors should be somewhat stable (it is of importance to note that unlike the assumptions of the binary perceptron, there is no guarantee the multi-class perceptron will reach a steady state), and the classifier will be ready to be put to use.

d. Classifier Building: The following sections detail how to format the data for use with the classifier builder, as well as how to train and save the classifier for later use. The last section deals with how to build an analytics report for the data.

e. Formatting the Data: The bulk of the classifier is abstracted away into a Python class that takes the following parameters as input:

f. Classes: List of categories/classes that data is divided into. This should be a Python list of strings, and each string should be an exact match of the class tag in the actual feature data.

g. Feature_list: List of features, as strings in a Python list.

h. Feature_data: A python list of tagged feature data, in the following format:

Feature_data = [(“class1”, { “feature1”: 0, “feature2”: 0, “feature3”: 0 }), (“class2”, { “feature1”: 0, “feature2”: 0, “feature3”: 1 }), (“class3”, { “feature1”: 0, “feature2”: 1, “feature3”: 0 })]

i. Train_test_ratio: Ratio of data to be used in training vs. Testing is set to 75% .

j. Iterations: Number of iterations to run training data through set to 100.

k. Work Flow: Here the inputs are $I_1,I_2,\ldots I_n$ which are also known as feature vectors and $W_1,W_2,\ldots W_n$ are the weight vectors which are helping us to find out the outcome from this implemented the algorithm. The flow of work has been described in the Figure 2.

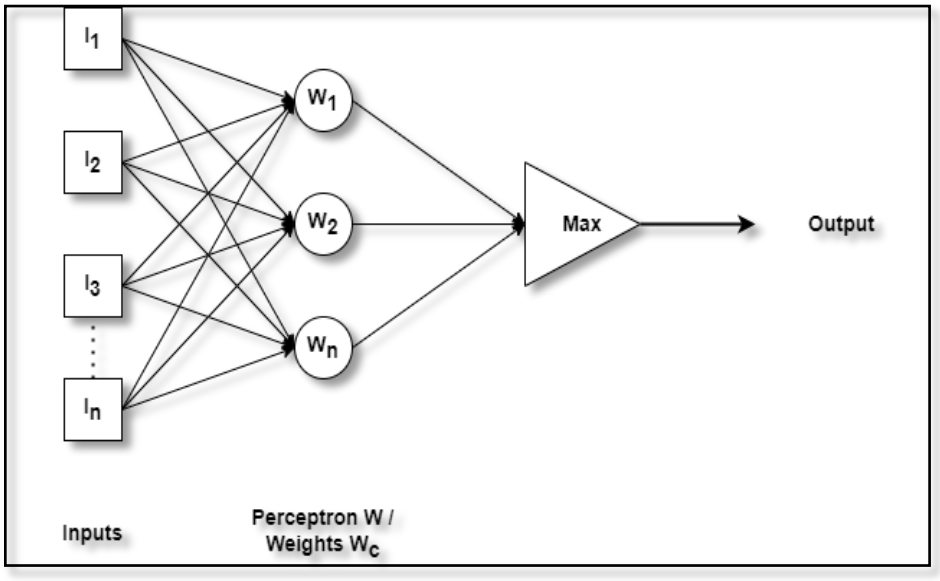


Fig 2: Workflow of Multiclass Perceptron Model

Apart from that there are other models which are also implemented in this scope of work where a comparative study has been provided. This implemented models in this scope of work are all helping in classifying the input data and these models are – Support Vector Machine where two of the major kernels like polynomial and linear, Stochastic Gradient Descent (SGD) and Quadratic Discriminant Analysis (QDA).

The analysis of the mentioned classifiers is working like the following flowchart (Figure 3) and all of these are providing us the classification results with their performance.

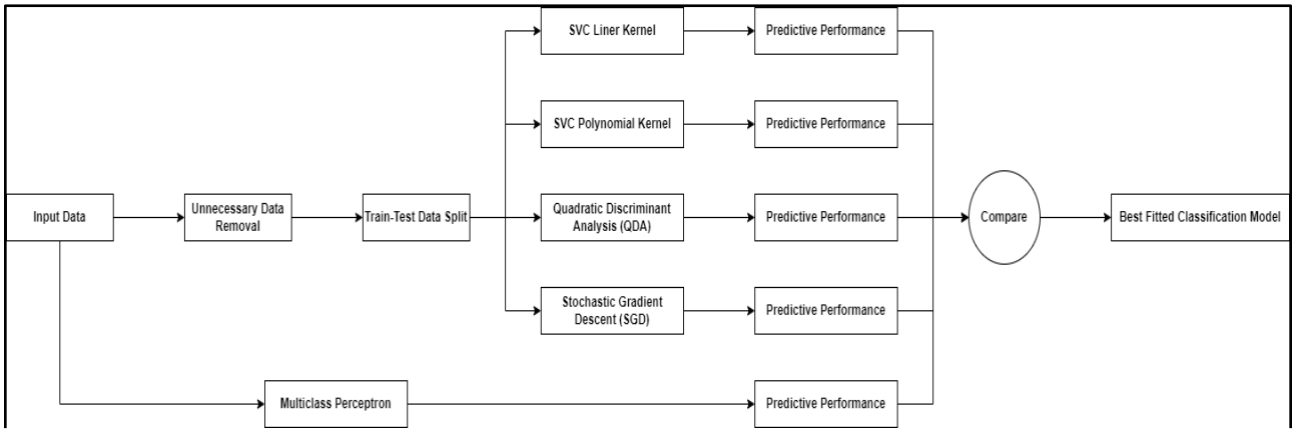


Fig 3: Workflow of entire proposed classification work

So, each of the model's performances are hereby gathered and later these are being compared so that we can have a conclusion for the best fitted model among all of the implemented models.

IV. RESULT AND DISCUSSION

This work has been performed with python and its libraries and dataset of Lupus disease. First dataset is loaded. The following dataset has been used to train the neural network. The dataset has been prepared by a survey conducted at North Bengal Medical College and Hospital. The dataset consists of 150 records. The field of the

dataset consists of different symptoms found in the Lupus patients. These symptoms are very useful to detect the Lupus stage. Table 1 depicts the glimpse of the dataset.

Table 1: Glimpse of the clinical dataset used in the study

Case	Wl	Fa	Fe	Cy	He	Pa	Class
1	0	0	0.5	1	0	0	0
2	0	0.5	0.5	0.5	1	0	1
3	0.5	1	1	1	0.5	1	2
4	1	0.5	0	0	0	0	0
5	0	0	0.5	0.5	0.5	0.5	1
6	0.5	1	1	0.5	0	0	2
7	0	0.5	0.5	0	0	1	1
8	0	1	0.5	1	1	0	2
9	0.5	1	0.5	1	0.5	0	2
10	1	0	0	0	1	0	1
11	0	0	0	0	1	0	0
12	0.5	0	0	0	0	0	0
13	1	0	0.5	0	0	0	0
14	0.5	1	0.5	0	0	0	1
15	0	1	1	0.5	0.5	0.5	2

Figure 4 shows the results achieved by the proposed model. An application of the python named jupyter has been used here to identify the stage of lupus using the dataset taken into consideration.

```
In [17]: classifier.run_analytics()

         CLASSIFIER ANALYSIS:

         PRECISION STATISTICS:
         1 Class Precision: 0.8571428571428571
         0 Class Precision: 1.0
         2 Class Precision: 1.0

         RECALL STATISTICS:
         1 Class Recall: 1.0
         0 Class Recall: 1.0
         2 Class Recall: 0.6666666666666666

         F-BETA SCORES:
         0 Class F-Beta Score: 1.0
         1 Class F-Beta Score: 0.923076923076923
         2 Class F-Beta Score: 0.8

         ACCURACY:
         Model Accuracy: 0.9375
```

Fig 4: Glimpse of result of the experiment

Figure 5 depicts the plot representation of the output. Here 100 epochs have been observed. The plot shows till 20 epochs the accuracy was not stable. After 20 epochs the stability in accuracy is achieved.

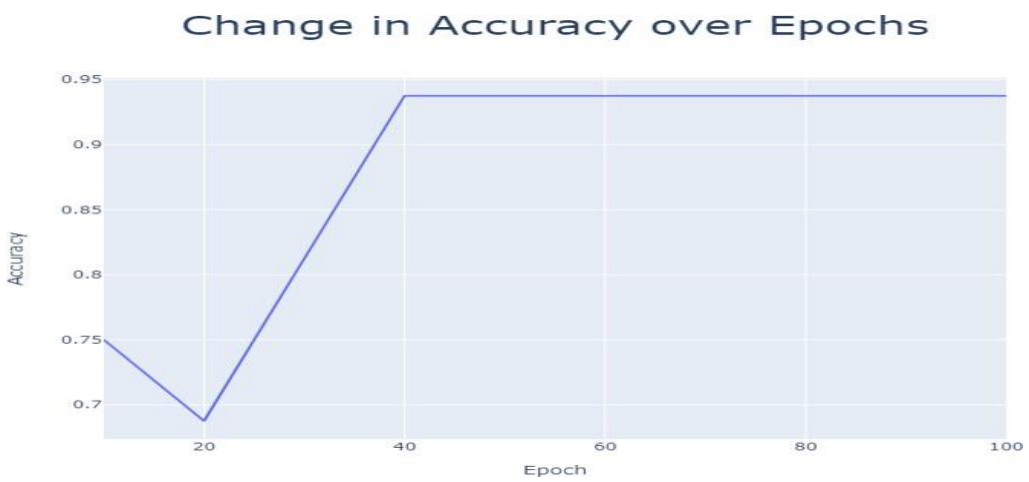
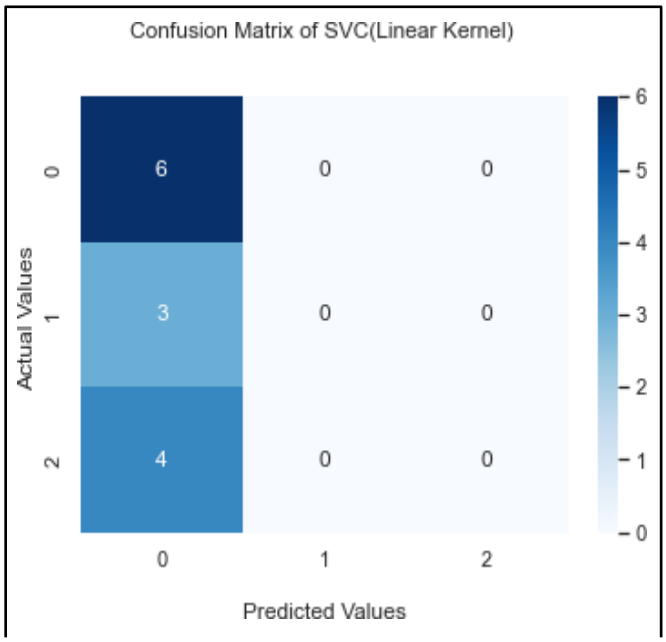
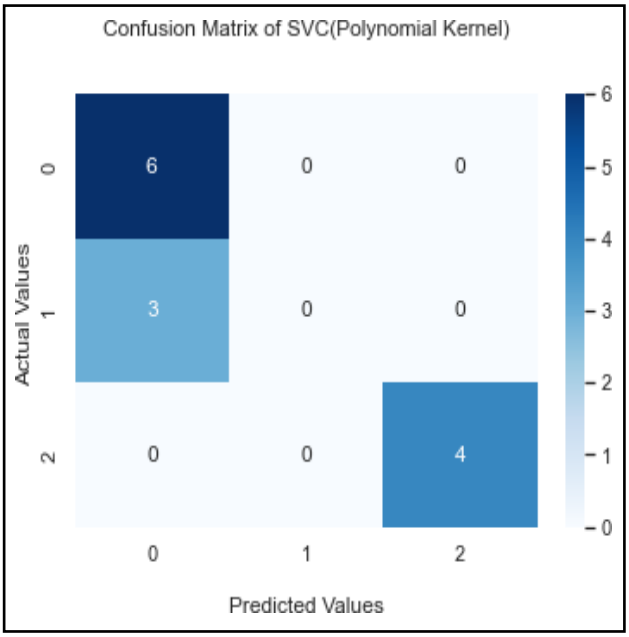


Fig 5: Plot representation of output.

Now we also have the outcome from other models as well from where we have plotted all of the confusion matrix from the prediction results of the models. We got the accuracy of each model like 46%, 77%, 46% and 92% for the Support Vector Machine where two of the major kernels like linear and polynomial, Stochastic Gradient Descent (SGD) and Quadratic Discriminant Analysis (QDA) respectively. The confusion matrix plots of individual classification models are mentioned below (Figure 6).





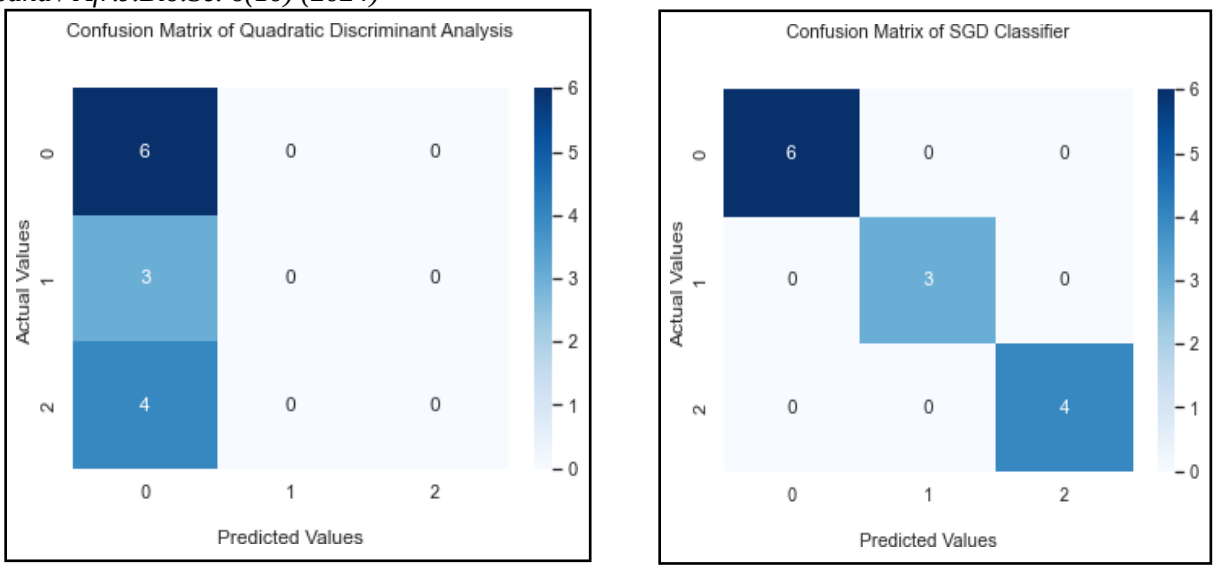


Fig 6: Confusion matrix plot of Support Vector Machine where two of the major kernels like linear and polynomial, Stochastic Gradient Descent (SGD) and Quadratic Discriminant Analysis (QDA) respectively

V. COMPARISON

The ANN model developed here, has been compared with Linear SVM, Polynomial SVM, QDA and SGD models. It has been found that the accuracy achieved by the multiclass perceptron developed here is the best among the other models taken into consideration which is 93.75%. Figure 7 depicts the result generated in python. It consists of the accuracy of the developed ANN along with other models. Figure 8 depicts the bar graph of different models.

	name	score
0	Linear_SVM	0.461538
1	Polynomial_SVM	0.769231
2	QDA	0.461538
3	SGD	0.923077

Fig 7: Glimpse of output of experiment

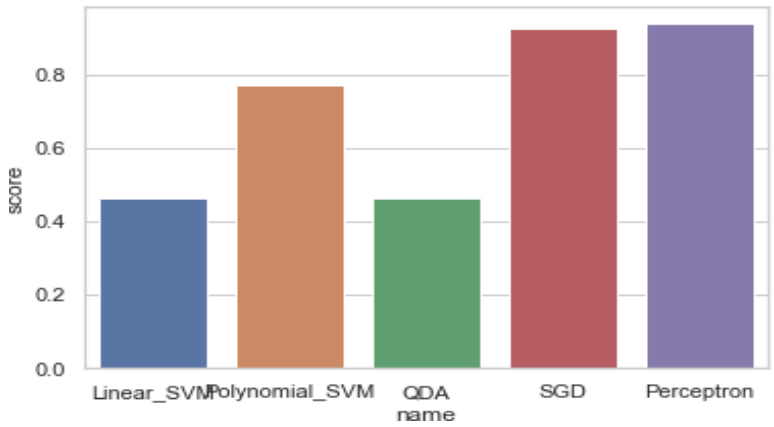


Fig 8: Graphical representation of comparison of different models

VI. CONCLUSION

The model developed here has been trained and tested with the dataset taken into consideration. Various libraries of python have been utilized. The developed model has been compared to four models. The accuracy of the proposed model is above 90%. This is very helpful for early diagnosis of Lupus disease. On the basis of the accuracy of the developed model other features can be taken into consideration for future development.

REFERENCES

[1]G.K.W.Lam,M.Petri"Assessment of Systemetic lupus erythmatosus", http://www.clinexprheumatol.org.

[2] Farhad Soleimanian Gharehchopogh, Maryam Molany and Freshte DabaghchiMokri, "Using Artificial Neural Network in diagnosis of thyroid disease": a case Study, International Journal Computational Sciences & Applications (IJCSA) Vol.3, No.4, August 2013.

[3] R. Dey, V. Bajpai, G. Gandhi, B. Dey, "Application of Artificial Neural Network (ANN) technique for Diagnosing Diabetes Mellitus", the Third international Conference on Industrial and Information Systems(ICIIS), PP.1-4, IEEE, Kharagpur, India,2008..

[4] F.S. Gharehchopogh, Z.A. Khalifelu," Neural Network Application in Diagnosis of Patient: A Case Study", International Conference on Computer Networks and Information Technology (ICCNIT), PP- 245 – 249, Abbottabad, Pakistan, 2011.Study", International Conference on Computer Networks and Information Technology (ICCNIT), PP. 245 – 249, Abbottabad, Pakistan, 2011.

[5] Payel Saha, Rakesh Kumar Mondal "Detection Of Dengue Disease using Artificial neural Network" International Journal of Computer Sciences and Engineering , May 12, 2016.

[6] Matthew H. Liang, Steven A Socher, Martin G. Larson and Peter H. Schur " Reliability and Validity of Six Systems for the Clinical Assessment of disease activity in systemic Lupus Erythematosus", Artheritis Rheum. September,1989; 32(9):1107-18.

[7] Thomas Stoll, Gerold Stucki, Javid Malik, Stephen Pyke, David A Isenberg " Further validation of the BILAG disease activity index in patients with systemic lupus erythematosus", 756-760,1996

[8] E.M. Hay, P.A. Bacon, C. GorDon, D.A. Isenberg, P. Maddison, M.L. Snaith, D.P.M. Symmons, N. Viner and A. Zoma "The BILAG index: a reliable and valid instrument for measuring clinical disease activity in systemic lupuserythematosus", Quaterly Journal of Medicine,1993; 86:447-458.

[9] Gary S.Firestein, Ralph Budd, Sherine E Gabriel, Iain B. Mclnnes, James R O'Dell " Kelly and Firestein's Textbook of Rheumatology" .

[10] D.P.M. Symmons, J.S.Coppock, P.A. Bacon, B. Bresnihan, D.A. Isenberg, P.Maddison, N. Mchugh, M.L.Snaith and A.S.zoma "Development and Assessment of a Computerized Index of Clinical Disease Activity in Systemic Lupus Erythematosus",Quarterly Journal of Medicine, New Series 69, No.258, PP.927-937, November 1988.

[11] Annegret Kuhn, Asyche Landmann " The Classification and diagnosis of cutaneous lupus erythematosus", Journal of Autoimmunity, 2014.

[12] Claire Bombardier, Dafna D. Gladman, Murray B. Urowitz, Dominique Caron, Chi Hsing Chang and the Committee on Prognosis Studies in SLE " Derivation Of The SLEDAI A Disease Activity Index for /lupus Patients", Arthritis and Rheumatism, Vol.35, No.6 June,1992.

[13] Laurene Fausett, " Fundamentals of Neural Networks",1994.

[14] http://timesofindia.indiatimes.com/life-style/health-fitness/health-news/Living-with-lupus/articleshow/29909854.cms (Available online)

[15] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3380630/#R11(Available)

[16] Katherine G. Skocelas "Systemic Lupus Erythematosus Symptom Prediction Using Recursive Neural Network", 2018.

[17] Siti Aisyah, Mawaddah Harahap, Abdi Dharma, Mardi Turnip "Implementation artificial neural network nguyen widrow algorithm for lupus prediction" 2019.

[18] E.M. HAY, P.A. BACON, C. CORDON, D.A. ISENBERG, P. MADDISON, M.L. SNAITH, D.P.M. SYMMONS, N. VINER and A. ZOMA From the ARC Epidemiology Research Unit, Stopford Building (University oi Manchester), Oxford Road, Manchester "The BILAG index: a reliable and valid instrument for measuring clinical disease activity in systemic lupus erythematosus" 13 March 1993 .

[19] Ghalib A Bello,1 Michael A Brown,2 Jennifer A Kelly,2 Aikaterini Thanou,3 Judith A James,2,4 Courtney G Montgomery " Development and validation of a simple lupus severity index using ACR criteria for classification of SLE" ,2016.

[20] Annegret Kuhn , Aysche Landmann "The classification and diagnosis of cutaneous lupus erythematosus" ,2014.