# Summary and Analysis of given Research Paper

The paper given to us, “Construction of Feed Forward MultiLayer Perceptron Model For Genetic Dataset inn Leishmaniasis Using Cognitive Computing”, presents a new approach for diagnosing Leishmaniasis.

Leishmaniasis is a tropical protozoan parasitic skin disease, transmitted through the bite of sand flies, and mainly affects the poor localities of Asia, Africa and America. Leishmanial parasites can modulate the human immune system for survival. Some species of leishmaniasis are becoming virulent and drug resistant too, which made researchers focus on the comprehensive identification and analysis of its metabolic pathway and genome.

The early diagnosis of this disease would reduce the mortality rate and control the infectious stages of the disease. A Recurrent Neural Network was used to process the genetic dataset, as it can be used for recognising the variance in the dataset to form a robust dataset defining mode using normalisation process.

**Feed Forward Multilayer Perceptron Model**

The Genetic datasets of leishmanias were all collected from the Gene Express Omnibus database, which is one of the public data repositories of functional genomic data. 33 separate datasets were collected, 5 of healthy people, and 28 of infected people, having total RNA and double stranded cDNA concentration of leishmaniasis in humans, with the same number of variables (45033 rows).

Each dataset was then labelled and compiled in a single dataset on CSV format. 80% of the dataset was used for training, and 20% for testing. The model was trained with the learning rate kept at 0.001. And the training was repeated for 500 epochs. Using 500 epochs is justified because when we see the graph of the loss function on page 23, the loss is getting minimised as the epochs increase, and near 500 only, the curve starts to flatten out.

The perceptron involved 4 hidden layers with 45 neurons each, and the accuracy, loss function, and mean square error of the model were studied for each iteration. The variations in the data were normalised using log2 normalisations. Normalising the data like this enables the changes to be interpreted proportionally, and doubling or halving of data can be seen as a change of -1 or +1. So it makes the analysis more efficient.

The Gradient Descent Optimiser is used to optimise the change in the threshold value. Gradient Descent takes into account the first derivative when performing the updates on the parameters. On each iteration, we update the parameters in the opposite direction of the gradient of the objective function J(w) w.r.t the parameters where the gradient gives the direction of the steepest ascent. The size of the step we take on each iteration to reach the local minimum is determined by the learning rate α, which is 0.001 in this case. Therefore, we follow the direction of the slope downhill until we reach a local minimum.

The output of each neuron depends on the weight matrix, given by

$$[Y] = \Sigma\ [x] * [w]$$

[Y] is the output matrix of the corresponding layer, [x] is the input matrix, [w] is the weight matrix. The iterative change in the weight call is carried out through back propagation. The back propagation algorithm make the ANN keep learning from the errors, hence improving the model.

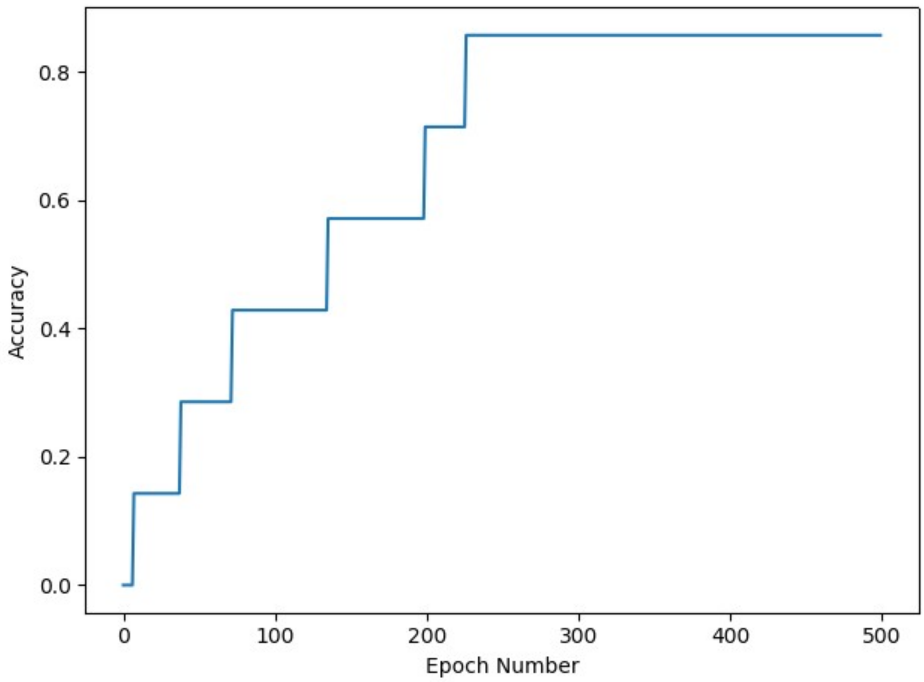
The output of each layer is the sum of the multiplied matrix of input and weight matrix with the bias matrix, given by

$$[Z] = [Y] + [B]$$

where [Z] is the output layer and [B] is the bias matrix.
The back propagation induced feed forward network is activated using the sigmoid function, which has the output range from 0 to 1. Output from sigmoid function is unnormalised probability, which simplifies the classification task. The model classifies two different leishmaniasis classes. These two classes are characterised in the output layer. The value of the output class crossing above the threshold value is given as an output of the given test data. The random structure of the the hidden layers increase the efficacy of the model. The increased accuracy of the model is due to the use of increased neutrons in the hidden layer. The increased number of hidden layers, decreases the error of the model.

**Accuracy**

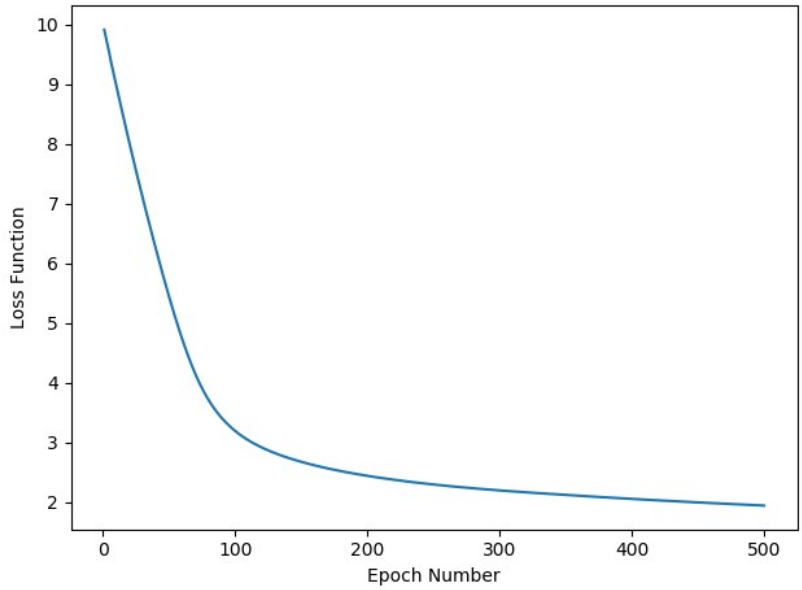
The final accuracy of the data is 85.71%, which shows that the model is a good one for predicting the experimental result for the given dataset of leishmaniasis. It also denotes the decreasing error state in the model, along with increase in the regression coefficient. The usage of back propagation along the feed forward neural network model over spike timing-dependency- plasticity model for biological datasets is because gradient descent used on the latter has been shown to be computationally expensive.
The accuracy stays at a certain level after 25-50 epochs in the beginning, and then goes higher after that. After the 250th epoch the accuracy stays constant at 85.71%. A variation of a step graph is obtained for the accuracy of the model as shown on page 22.




**Loss Function**

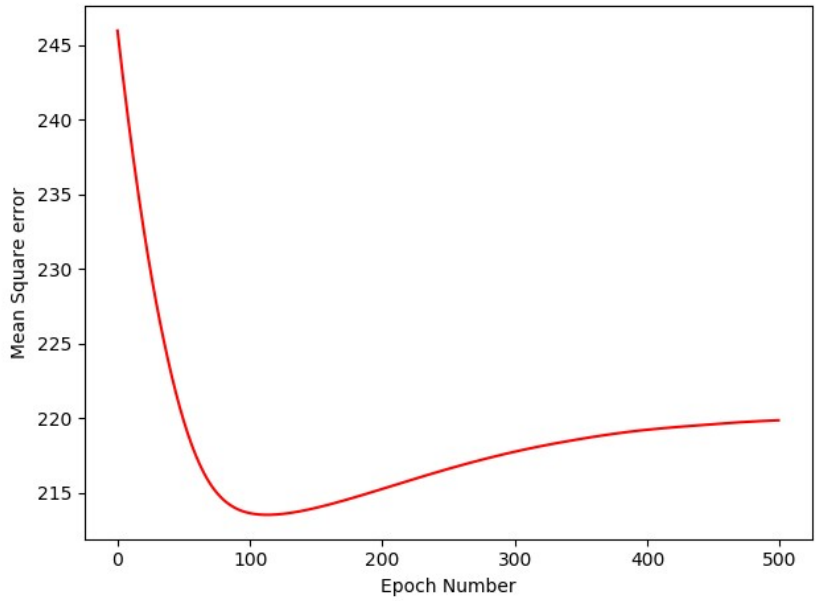
The loss function shows a decrease in value with constant increase in the number of epochs. It is 9.91 at the start of training, and 1.94 as the training ends. The increase in the number of epochs decreases the cost. While plotting the graph of the loss function against the number of epochs, the curve starts to flatten out as the epochs approach 500, telling us that training the model for 500

epochs is justified. The graph is shown on page 23.




**Mean Square Error**

The mean square error of the process is 245.96 initially. After that there is a gradual decrease int the MSE till 100 epochs, and then a slight increase till 300 epochs. This is observed because after 100 epochs, the model starts to overfit the data, and after 300 epochs, the MSE is standardised to the value of 219. The graph of the MSE against the number of epochs is shown on page 24.




**Conclusion**

This paper used techniques which we were already largely familiar with. Using back propagation in the Artificial Neural Network makes the model keep learning from the errors, and it gave an accuracy of 85.71%. This did help in bringing a new solution for predicting leishmaniasis in humans. The paper is written in a very concise and readable format, and is easy to follow. It explains all the methods used very clearly, but in some places the grammar is a bit off, but the idea is largely un-

derstood. The results which they have got are very promising, and this is a really great step in the direction of predicting and treating Leishmaniasis.