

CLASSIFICATION OF SICKLE CELL DISEASE USING FEEDFORWARD NEURAL NETWORK WITH BACK PROPAGATION

Rinku Naik¹, Dr Chandra Sekhar Panda²

¹Mphil Scholar, Dept. of Computer Science and Applications,

²Reader & Head, Dept. of Computer Science and Applications, Sambalpur University, Jyotivihar, Burla, Odisha, India

Abstract – This study is on Sickle Cell Disease (SCD) and Artificial Neural Network (ANN). SCD is a blood disorder and it needs at least 24 hours to get diagnosed and to know which type of SCD the patient does have by conventional method of diagnosis. Till now, it's a challenge to get diagnosed either in terms of Reliability for the tests done quickly like Blood Centrifugation method and/or cost-effectiveness for the costly HPLC method. So, for this study, I have taken the references of recent studies in the field of Healthcare, SCD and ANN techniques being used to diagnose the issue/ disease. And for this research I have used the same ANN techniques for classifying SCD and identify patient with/ without SCD, which in turn will help doctors to take decision for treating a patient with SCD.

In this study, the focused is only on popular Feed Forward Neural Network (FFNN). The inputs will be Blood Test parameters (i.e. Hb, RBC, MCV, MCH, MCHC, WBC, Platelet Count, Bilirubin-Total and Bilirubin-Direct). Again, a bias is added in hidden layer to the sum of the input parameter. FFNN selects 70% of input for best match and 30% for testing. Lastly, Sigmoid function is used to activate the NN to provide the output SCD with 0 or 1 value.

Key words: Sickle Cell, SCD, Artificial Neural Network, ANN, FFNN, Hb, RBC, MCV, MCH, MCHC, WBC

1. INTRODUCTION

Sickle Cell Disease (SCD) is found all over the world, particularly people who migrated from the Malaria affected region like Africa, South East Asia in particular India. SCD mainly occurs due to the mutation in Hemoglobin molecule. SCD occurs due to inherited abnormal hemoglobin (Hb) gene, which produce HbS (Hb-Sickle). Due to the presence of HbS and because of its abnormality converts RBCs into a half-moon shaped called as Sickle. This sickle shaped instead of normal soft round shape, which is the main cause of complication of Sickle Cell disease. Sickle gene is transmitted from parent to child. If both the parent are having Sickle gene, then it is sure that the child will have SCD. If one of the parents is Sickle gene and other parent is either Sickle Cell Trait (Sickle Cell Carrier) then, there is a probability of having child with SCD.

SCD can cause painful episodes like: Stroke, Chest Pain, Abdominal pain, Infections, Jaundice, and Anemia. So, early

diagnosis of SCD is very important because many complications can be prevented with early diagnosis and treatment.

The diagnosis is done using the available laboratory screening test and confirmatory test [1]. The screening test is done using sodium metabisulphite sickling test and sickle solubility tests to screen out that the patient is a probable candidate of Sickle Cell. Post screening test, widely available confirmatory test haemoglobin electrophoresis (HBE), isoelectric focusing (IEF), and high performance liquid chromatography (HPLC) done to confirm the SCD.

Although, the screening methods are inexpensive, are not reliable and not at all confirmatory test, rather it suggests for a confirmatory test. And although, Confirmatory tests are reliable, these tests are costly, take longer time and dependent on other parameters like: Alkaline and Acidic conditions, Visibility and Mobility. Under acidic conditions, Under haemoglobin electrophoresis (HBE), the HbF migrates relatively more fast and is therefore uniquely differentiable from both HbA and HbS. Isoelectric focusing(IEF) uses the same principles but is slightly more expensive than HBE. However, it is able to identify more Hb variants that would not be detected by HBE. It also has the advantage that it does not require commercial reagents. HPLC uses cation exchange chromatography to identify the various hemoglobins (Hb) in an individual subject of interest (here patient) and also accurately quantify the Hb levels.

In developed countries, screening has largely been replaced by HPLC and confirmation is then done by IEF or HBE because of the exhaustive nature of work, time consumption and the inability of not identifying abnormal bands or quantifies Hb. Additionally, the quantification of Hb fractions by HPLC is used to monitor patients who are on Hydroxyurea therapy or exchange blood transfusion [1]. Therefore, reliable and cost effective way of diagnosing SCD is still a challenge.

The purpose of this work is to help doctors to screen out the SCD patient using the FFNN-BP model and send for further confirmatory test and then treatment.

To explain this briefly, let's take an example of a patient with Yellowish and/ or pale skin color or weak comes to the doctor for the treatment. Then doctor will ask only for CBC

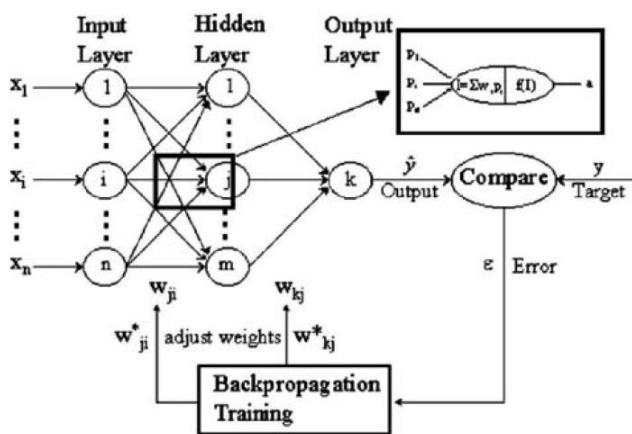
(Complete Blood Count) Test to know the level of Hb level and Bilirubin test to confirm Jaundice. If he found Hb level is below normal reference range and jaundice result positive, then doctor send the CBC and Bilirubin hematological data for testing with the proposed FFNN-BP model to confirm whether the patient is a positive case of SCD or negative case of SCD. And then doctor can start the treatment.

2. PROPOSED METHODOLOGY

One of the network structures that have been widely used is the feed forward neural network (FFNN) with Back-Propagation. Because of the Fault tolerance, Generalization and Learning from environment like capabilities of ANN, it's becoming more and more popular in medical diagnosis and healthcare sector.

2.1 FFNN with Back-propagation

Learning in FFNN with backpropagation occurs during the training phase in which each input pattern from the training set is applied to the input layer and then propagates forward. The pattern of activation arriving at the output layer is then compared with the correct (associated) output pattern to calculate an error signal. The error signal for each such target output pattern is then backpropagated from the output layer to the input neurons in order to adjust the weights in each layer of the network. After the training phase during which the NN learns the correct classification for a set of inputs, it can be tested on a second (test) set of samples to see how well it classifies new patterns. Thus, an important consideration in applying backpropagation learning is how well the network makes the generalization.



[Fig 2.1.1 A three layer FFNN with Back Propagation^{23,24]}

FFNN with BP as shown in Fig 2.1.1 above, i, j and k denote input layer, hidden layer and output layer neurons, respectively and w is the applied weight of the neuron [24]. The term "feed-forward" means that a neuron association only be present from a neuron in the input layer to other neurons in the hidden layer or from a neuron in the hidden

layer to neurons in the output layer and the neurons within a layer are not associated to each other.

The FFNN-BP expression is represented as below:

$$\hat{y}_k = f_o \left[\sum_{j=1}^M w_{kj} \cdot f_h \left(\sum_{i=1}^N w_{ji} x_i + w_{jo} \right) + w_{ko} \right]$$

[Fig 2.1.2 FFNN-BP output expression^{24]}

Where,

- w_{ji} - the weight in the hidden layer connecting the i^{th} neuron in the input layer and the j^{th} neuron in the hidden layer,
- w_{jo} - the bias for the j^{th} hidden neuron,
- f_h - the activation function of the hidden neuron,
- w_{kj} - the weight in the output layer connecting the j^{th} neuron in the hidden layer and the k^{th} neuron in the output layer,
- w_{ko} - the bias for the k^{th} output neuron,
- f_o - the activation function for the output neuron,
- x_i - the i^{th} input variable for input layer
- \hat{y} - the calculated output variable
- y - the observed output variables

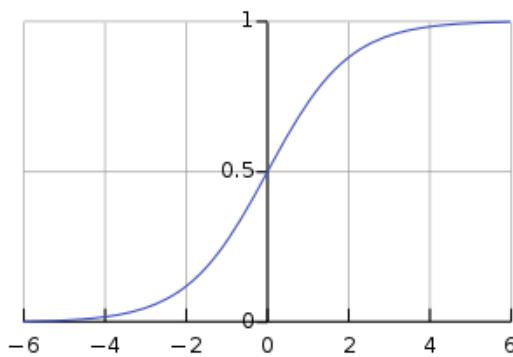
The weights are different in the hidden and output layers, and their values can be changed during the process of network training.

2.2 Sigmoid Function

Sigmoid function is widely used for a classification problem, where the problem needs to get output either true or false. And for this reason sigmoid activation function is also used in Medical Diagnosis problems. And as our problem is a classification one, in this thesis, we have focused only on Sigmoid activation function and we want the output to be either 0 or 1 [15] based on the CBC data of the patient to know whether the patient has sickle cell Disease or not i.e if it is SCD, the output is 1, else output is 0.

The sigmoid activation function is defined as below [2], [15] which has the range from 0 to 1.

$$S(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1}.$$



[Fig 2.3 Sigmoid function and graph]

Let's understand the Sigmoid function as below w.r.t the Fig.3 above:

- We know the output of Sigmoid function falls within 0 to 1. The specialty of this function is that the curve crosses at 0 of x-axis and 0.5 of y-axis and is the rule which drives the activation function. So, when Neural Network's is greater than or equal to 0.5, the output is considered to be near to 1 and positive else the output is considered to be negative.
- When the value at x-axis is more towards left side or negative, then the output is considered as 0. And when the value at x-axis is more towards right side or positive, then the output is considered as 1. But, when the value is around 0 at of x-axis, then we can see more deviation as it fluctuates relatively more near this point.

3. Experiment and Discussion

From the list of references, it was found that FFNN is most widely preferred and suitable for Classification of Medical Diagnosis. So, in this experiment FFNN is taken into consideration. However, comparison with other models is scope of future work.

3.1 Pre-Modelling

The experiment is carried out in Intel(R) Core i3-3217U, CPU 1.80GHz, 12 GB RAM, with Windows 10, 64 bit operating system.

CBC and Bilirubin Test results of 534 patients were studied and 9 hemogram input parameters such as Hb, RBC, MCV, MCH, MCHC, WBC, Platelet Count, Bilirubin-Total, Bilirubin-Direct were considered to get Sickling/ SCD as Output.

70% of the datasets selected for training and 30% datasets were selected for testing.

Sigmoid Activation Function is selected for this FFNN-BP model wherein, if the output is not as expected, then weight is increased/ decreased with bias parameters i.e randomly

by the classifier and the experiment goes on until get a desired output.

3.2 Modelling

Experiment is done using below programming language, Libraries:

Programming language = Python

- Python is the best tool for Machine Learning integration and deployment
- Simple, Emerging, and More popular with active community
- Easier to learn how to build a model from scratch
- Python includes great libraries to manipulate matrix or to code the algorithms

Python libraries = Numpy, Pandas, Scikit, Matplotlib, Seaborn, Tensorflow

- Numpy – NumPy (Numerical Python) is an important linear algebra library very useful for performing mathematical and logical operations on Arrays on which almost every data science or machine learning Python packages such as SciPy (Scientific Python), Matplotlib (plotting library), Scikit-learn, etc depends on to a reasonable extent.
- Pandas – Pandas (Python Data Analysis Library) takes data (like a CSV or TSV file, or a SQL database) and creates a Python object with rows and columns called data frame that looks very similar to table and Numpy treat these as array.
- Scikit-learn – The library is built upon the SciPy (Scientific Python) and Extensions or modules for SciPy are conventionally named SciKits. Scikit-learn provide a range of supervised and unsupervised learning algorithms via a consistent interface in Python.
- Matplotlib – Matplotlib is a plotting library and an open source alternative to MATLAB like plotting.
- Seaborn – Seaborn is a data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics
- Tensorflow – Tensorflow is a foundation library created and released by Google for fast numerical computing that can be used to create Deep Learning models directly or by using wrapper.

70% of datasets are trained and 30% are tested

Deep Neural Network (DNN) classifier is taken with 10, 20, and 10 layers of neurons. (Note: A Multilayer Perceptron Classifier in Tensorflow)

Weights are randomly selected by the DNN Classifier

Sigmoid Activation Function is taken for the classifier model

Two classes i.e. 0 and 1 are the output

Batch Size is 20

For the experiment, 100 iterations is taken

3.3 Input

Total dataset = 534,

Dataset Trained = 70% Data (i.e. 334)

Dataset Tested = 30% Data (i.e. 160)

3.4 Output

The below given output is the actual output which was obtained by running the Python Code (Code and Dataset is not provided due to confidentiality):

```
Python 3.6.6 |Anaconda, Inc.| (default, Jan 28 2019, 11:27:44) [MSC v.1900 64 bit (AMD64)]
```

```
Type "copyright", "credits" or "license" for more information.
```

```
IPython 7.0.1 -- An enhanced Interactive Python.
```

```
In [1]: runfile('C:/Users/Downloads/RinkuTest.py', wdir='C:/Users/Downloads')
```

```
C:\Users\Anaconda3\lib\site-packages\scipy\stats\stats.py:1713: FutureWarning: Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.
```

```
return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```

```
C:\Users\Anaconda3\lib\site-packages\statsmodels\nonparametric\kde.py:488: RuntimeWarning: invalid value encountered in true_divide
```

```
binned = fast_linbin(X, a, b, gridsize) / (delta * nobs)
```

```
C:\Users\Anaconda3\lib\site-packages\statsmodels\nonparametric\kdetools.py:34: RuntimeWarning: invalid value encountered in double_scalars
```

```
FAC1 = 2*(np.pi*bw/RANGE)**2
```

```
C:\Users\Anaconda3\lib\site-packages\numpy\core\fromnumeric.py:83: RuntimeWarning: invalid value encountered in reduce
```

```
return ufunc.reduce(obj, axis, dtype, out, **passkwargs)
```

```
INFO:tensorflow:Using default config.
```

```
WARNING:tensorflow:Using temporary folder as model directory: C:\Users\AppData\Local\Temp\tmp36k5tvjq
```

```
INFO:tensorflow:Using config: {'_model_dir': 'C:\\\\Users\\\\AppData\\\\Local\\\\Temp\\\\tmp36k5tvjq', '_tf_random_seed': None, '_save_summary_steps': 100, '_save_checkpoints_steps': None, '_save_checkpoints_secs': 600, '_session_config': allow_soft_placement: true}
```

```
graph_options {
```

```
rewrite_options {
```

```
meta_optimizer_iterations: ONE
```

```
}
```

```
}
```

```
, '_keep_checkpoint_max': 5, '_keep_checkpoint_every_n_hours': 10000, '_log_step_count_steps': 100, '_train_distribute': None, '_device_fn': None, '_protocol': None, '_eval_distribute': None, '_experimental_distribute': None, '_service': None, '_cluster_spec': <tensorflow.python.training.server_lib.ClusterSpec object at 0x00000272DF594CF8>, '_task_type': 'worker', '_task_id': 0, '_global_id_in_cluster': 0, '_master': '', '_evaluation_master': '', '_is_chief': True, '_num_ps_replicas': 0, '_num_worker_replicas': 1}
```

```
WARNING:tensorflow:From C:\Users\Anaconda3\lib\site-packages\tensorflow\python\estimator\inputs\queues\feeding_queue_runner.py:62: QueueRunner.__init__ (from tensorflow.python.training.queue_runner_impl) is deprecated and will be removed in a future version.
```

Instructions for updating:

To construct input pipelines, use the `tf.data` module.

```
WARNING:tensorflow:From C:\Users\Anaconda3\lib\site-packages\tensorflow\python\estimator\inputs\queues\feeding_functions.py:500: add_queue_runner (from tensorflow.python.training.queue_runner_impl) is deprecated and will be removed in a future version.
```

Instructions for updating:

To construct input pipelines, use the `tf.data` module.

INFO:tensorflow:Calling model_fn.

INFO:tensorflow:Done calling model_fn.

INFO:tensorflow>Create CheckpointSaverHook.

INFO:tensorflow:Graph was finalized.

INFO:tensorflow:Running local_init_op.

INFO:tensorflow:Done running local_init_op.

WARNING:tensorflow:From C:\Users\Anaconda3\lib\site-packages\tensorflow\python\training\monitored_session.py:804: start_queue_runners (from tensorflow.python.training.queue_runner_impl) is deprecated and will be removed in a future version.

Instructions for updating:

To construct input pipelines, use the `tf.data` module.

INFO:tensorflow:Saving checkpoints for 0 into C:\Users\AppData\Local\Temp\tmp36k5tvjq\model.ckpt.

INFO:tensorflow:loss = 15.023012, step = 1

INFO:tensorflow:Saving checkpoints for 19 into C:\Users\AppData\Local\Temp\tmp36k5tvjq\model.ckpt.

INFO:tensorflow:Loss for final step: 7.7784033.

INFO:tensorflow:Calling model_fn.

INFO:tensorflow:Done calling model_fn.

INFO:tensorflow:Graph was finalized.

INFO:tensorflow:Restoring parameters from C:\Users\AppData\Local\Temp\tmp36k5tvjq\model.ckpt-19

INFO:tensorflow:Running local_init_op.

INFO:tensorflow:Done running local_init_op.

<function confusion_matrix at 0x00000272DB5C7C80>

[[74 6]

[11 69]]

Precision 0.925

Sensitivity 0.870588235

Specificity 0.8625

Accuracy 0.89375

4. Discussion

With the accuracy result of 89%, is quiet motivating to see the designed model good enough for classification of SCD. And also the Precision, Sensitivity, Specificity for the model are found to be acceptable.

5. CONCLUSION

Use of the proposed classifier based on FFNN with BP will result in more accurate and reliable diagnosis of SCD. Using our system, doctors can screen out SCD with confidence. Moreover, our technique can also be applied by the doctors in order to confirm their diagnostic decision.

REFERENCES

- [1] J. Makani, S. F. Ofori-Acquah, O. Nnodu, A. Wonkam, and K. Ohene-Frempong, 2013, Sickle Cell Disease: New Opportunities and Challenges in Africa, The Scientific World Journal, Volume 2013, Article ID 193252
- [2] Multi-Layer Neural Networks with Sigmoid Function—Deep Learning for Rookies <https://towardsdatascience.com/multi-layer-neural-networks-with-sigmoid-function-deep-learning-for-rookies-2-bf464f09eb7f>
- [3] <https://www.nigeriagalleria.com/Community-Health/Sickle-Cell.html>
- [4] <https://www.mayoclinic.org/diseases-conditions/sickle-cell-anemia/symptoms-causes/syc-20355876>
- [5] Antwi-Boasiako C, Ekem I, Abdul-Rahman M, Sey F, Doku A, Dzudzor B, Dankwah GB, Hagar Otu K, Ahenkorah J, Aryee R, Hematological parameters in Ghanaian sickle cell disease patients, 31 October 2018 Volume 2018:9 Pages 203—209, <https://doi.org/10.2147/JBM.S169872>
- [6] Sanjeev Shyam Rao, Jagdish Prasad Goyal, S.V. Raghunath, Vijay B. Shah, Hematological profile of sickle cell disease from South Gujarat, India, Hematology Reports [22 May 2012, 4(2):e8], <https://doi.org/10.4081/hr.2012.e8>
- [7] Ademola Samson Adewoyin, Management of Sickle Cell Disease: A Review for Physician Education in Nigeria (Sub-Saharan Africa), Volume 2015, Article ID 791498, 21 pages, <http://dx.doi.org/10.1155/2015/791498>
- [8] Sickle Cell Disease, https://www.hopkinsmedicine.org/healthlibrary/conditions/hematology_and_blood_disorders/sickle_cell_disease_85,P00101

- [9] Shodhganga, Chapter-4, Artificial Neural Networks, http://shodhganga.inflibnet.ac.in/bitstream/10603/48/6/chaper%204_c%20b%20bangal.pdf
- [10] <https://medium.com/@rajachak127/a-funny-thing-about-the-perceptron-21c145c639a>
- [11] <https://towardsdatascience.com/from-fiction-to-reality-a-beginners-guide-to-artificial-neural-networks-d0411777571b>
- [12] <https://www.analyticsindiamag.com/most-common-activation-functions-in-neural-networks-and-rationale-behind-it/>
- [13] Meenakshi Verma, Medical Diagnosis using Back Propagation Algorithm in ANN, International Journal of Science, Engineering and Technology Research (IJSETR), Volume 3, Issue 1, January 2014, <http://ijsetr.org/wp-content/uploads/2014/01/IJSETR-VOL-3-ISSUE-1-94-99.pdf>
- [14] Feed Forward Neural Network, <https://www.learnopencv.com/understanding-feedforward-neural-networks/>
- [15] Sigmoid Function, https://en.wikipedia.org/wiki/Sigmoid_function
- [16] Feed Forward Neural Network architecture, <https://cs.stanford.edu/people/eroberts/courses/soco/projects/neural-networks/Architecture/feedforward.html>
- [17] Qeethara Kadhim Al-Shayea, Artificial Neural Networks in Medical Diagnosis, 2011, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011, ISSN (Online): 1694-0814
- [18] RNN(Recurrent neural network) https://en.wikipedia.org/wiki/Recurrent_neural_network
- [19] Mohammed Khalaf, Abir Jaafar Hussain, Robert Keight, Dhiya Al-Jumeily, Paul Fergus, Russell Keenan, Posco Tso, 2016, Machine learning approaches to the application of disease modifying therapy for sickle cell using classification models
- [20] RFM - Random Forest Model <https://www.scirp.org/journal/PaperInformation.aspx?paperID=81016>
- [21] SVM (Support Vector Model) <https://www.kdnuggets.com/2016/07/support-vector-machines-simple-explanation.html>
- [22] Farzbod Amirhossien, Faridhossieni Alireza, Javan Kazem, Sharifi Mohammadbagher, A Comparison of ANN and HSPF Models for Runoff Simulation in Balkhichai River Watershed, Iran, American Journal of Climate Change, 2015, 4, 203-216, <http://dx.doi.org/10.4236/ajcc.2015.43016>
- [23] Vahid Nourani, Mehdi Komasi, A Multivariate ANN-Wavelet Approach for Rainfall-Runoff Modeling, Water Resources Management · November 2009, DOI: 10.1007/s11269-009-9414-5
- [24] Akinsegun Akinbami, Adedoyin Dosunmu, Adewumi Adediran, Olajumoke Oshinaike, Phillip Adebola, and Olanrewaju Arogundade, Haematological values in homozygous sickle cell disease in steady state and haemoglobin phenotypes AA controls in Lagos, Nigeria, 1 August 2012, <https://doi.org/10.1186/1756-0500-5-396>
- [25] C. E. Omoti, Haematological Values In Sickle Cell Anaemia In Steady State And During Vaso-Oclusive Crisis In Benin City, Nigeria, Vol. 4, No. 2; 2005: 62 – 67, <http://www.bioline.org.br/pdf?am05016>
- [26] Okocha Emmanuelchide, Odenigbo Charle, Okonkwo Uchenna, Hematological parameters in association with outcomes in sickle cell anemia patients, Indian Journal of Medical Sciences [01 Sep 2011, 65(9):393-398], <https://doi.org/10.4103/0019-5359.108955>