Revolutionizing Sickle Cell Treatment: Harnessing AI for Personalized Hydroxyurea Response Prediction

Dere Abdulhameed Abiola
University of Ilorin

Introduction

Sickle Cell Anemia (SCA) is one of the most prevalent hereditary diseases worldwide and a major source of considerable morbidity and mortality in Africa [1].

In Sickle cell patients, there is a non-conservative missense mutation where the amino acid Glutamate (hydrophilic) in the 6th position of the beta (β) globin is substituted with a Valine (hydrophobic). The Red Blood Cells (RBCs) then take the shape of a crescent/sickle [2].

Hydroxyurea is effective in enhancing sickle cell patient's hematological characteristics by stimulating the formation of red cells with higher hemoglobin F (HbF) content and less inclination to polymerize, hence lowering sickling phenomena with a corresponding increase in hematocrit [3]

Given that SCD is such a complex disorder with several systemic manifestations, precision medicine may have significant advantages for people with the condition. It appears likely that various patients will require various therapies [4]. Since hydroxyurea is one of the most promising therapies for SCD and also an antineoplastic drug, meaning that it has potential side effect, same prescription for all patients might be a catastrophic move, hence, the need for an early onset prediction and knowledge of pattern.

Therefore, we used state-of-the-art AI tools to (i) classify the patients into responders and non-responders and (ii) to predict the fetal hemoglobin content one-month ahead.

Data Collection

This study made use of the hospital record of 364 sickle cell disease (SCD) patients between the ages of 1 and 16 in a hospital facility who have been on hydroxyurea therapy for a minimum of 8 months.

The study leveraged a diverse dataset comprising key clinical indicators such as Mean Corpuscular Hemoglobin (MCH), Packed Cell Volume (PCV), Mean Corpuscular Hemoglobin Concentration (MCHC), Age, Sex, Platelet Count, and Hemoglobin Count.

These factors, intricately related to sickle cell disease severity and treatment response, were processed and harmonized.

Materials and Methods

- We underscored the use of Artificial Neural Network (ANN)
- The dataset was split into training (80%, n=292) and validation subsets (20%, n=72).
- We utilized neural network models including Tensorflow and Keras.
- Disparity in the range of the dataset was handled by normalization and scaling using Pandas.
- K-fold cross-validation approach was used to evaluate the performance of the performing models.
- The neural network used had two (2) hidden layers.
- 64 input neurons, 32 neurons for the first hidden layer, 16 for the second hidden layer and then a single neuron at the output layer (64-32-16-1).
- Adam Optimizer was used for the iterative update.
- ReLu activation function was used on all the layers with the exception of the output layer.
- The loss was evaluated using mean squared error.

Network Architectures

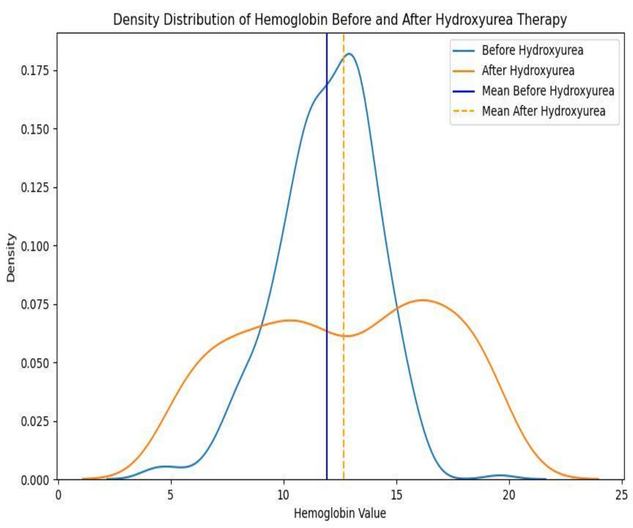
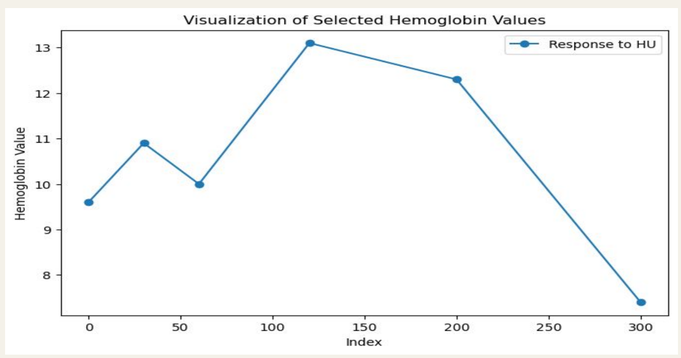
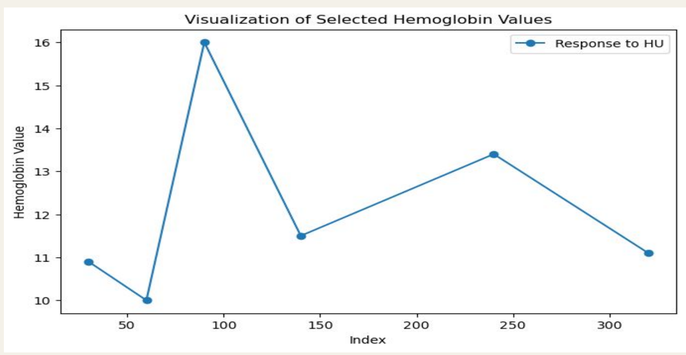


Figure 1: *Density distribution of SCA patients %HbF before and after undergoing hydroxyurea therapy*

Statistical Analysis





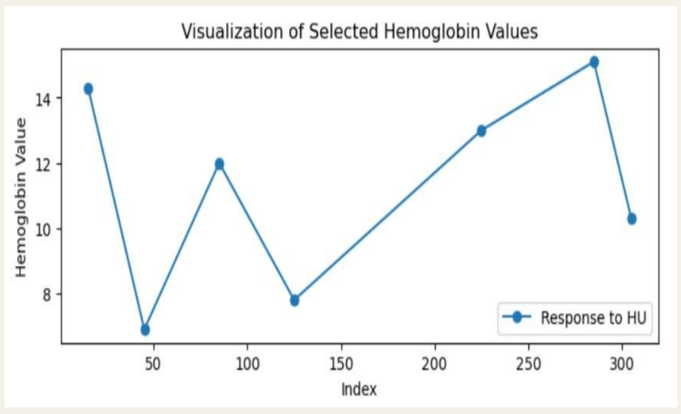


Figure 2: Plots showing the varying response of different patient's Fetal Hemoglobin to Hydroxyurea

Results

Aim 1: Classify responders and non-responders

- With a backpropagation learning technique, we created a shallow two-stage fully connected, feed-forward artificial neural network architecture for training and testing.
- Patients whose HbF experienced a 50% increase or more compared to their initial state prior to HU therapy were classified as responders and anyone whose values did not meet that threshold were classified as non-responders.
- ANN was able to identify responders with 80% accuracy after using n-fold cross validation.

Aim 2: Predicting HbF percentage one month ahead

- We combined a recurrent neural network (RNN) with Long short-term memory (LSTM) for the time series prediction.
- This was carried out as a series of regression problems with continuous output values at each stage, equally to the 8 months that was considered for the study.
- We monitored the HbF progression on a monthly basis for each patient as well as predicted values for corresponding months.
- Pearson's Correlation Co-efficient (PCC) was used as a performance evaluation metric to determine the similarity between the predicted value and the expected value.
- Out of the 5 patients that were evaluated, 1 of them had values between 0.5-1 indicating some form of stronger correlation, two of them had values between 0.30 and 0.49 while the remaining 2 had less than 0.29

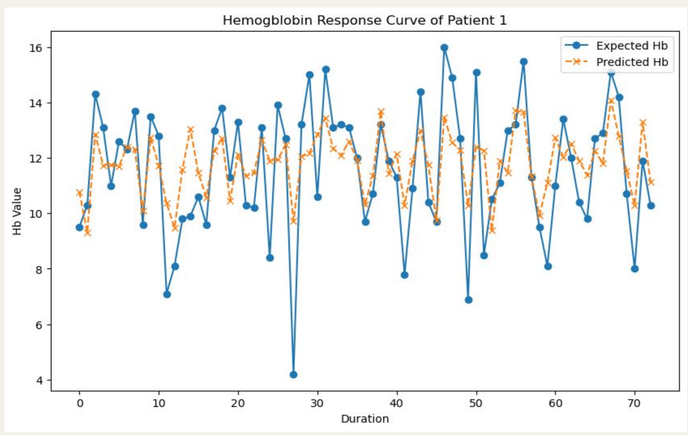
Limitations

- Insufficient sample dataset as ANNs require large amount of dataset to work with.
- Training complex ANNs are computationally expensive, requiring powerful hardware (GPUs/TPUs) and time.

Next Steps

- Gather a more robust dataset of the SCD patients particularly that of the fetal hemoglobin quantification.
- Enhance the study scope to include the observation of hydroxyurea in steady state versus crisis state.

Discussion



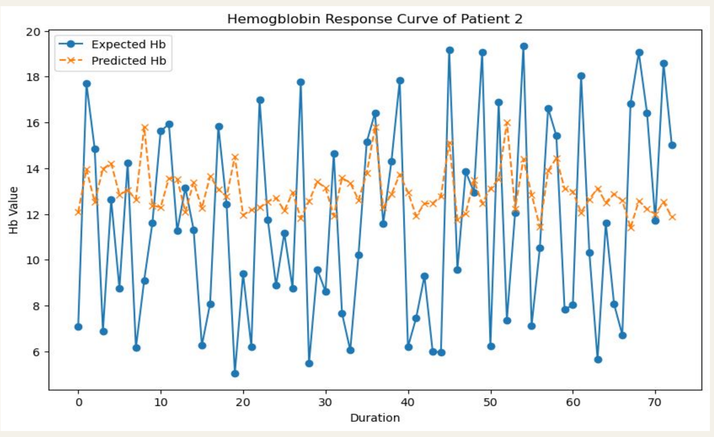


Figure 3: Pearson Correlation Co-efficient evaluation of the LSTM neural network showing the expected and predicted HbF level.

Conclusion

Neural networks have the capacity of discovering patterns and intricate correlations between parameters that are difficult for a human to notice on their own. The study's parameter values were collected both prior to and following the patients receival of hydroxyurea medication.

Our study also demonstrated that with the use of state-of-the-art AI tool, we predicted the response of patients to hydroxyurea with 80% accuracy. This will improve personalized medicine by optimizing the dosage of drug to be given to patient to be able to achieve maximum response while also preventing side effects of anti-cancer drugs.

This study is the first in many aiming to explore predictive medicine in SCD patients. With this, it is expected that hydroxyurea will be used optimally for them.

ANNs may be trained to distinguish between various patient classes accurately, regardless of how the output classes are organized. This is useful to the medical practice at large.

This study demonstrates that ANNs can take advantage of medical data correlations that doctors are unable to recognize on their own. This strategy and others like it will improve the use of precision medicine and have a favorable impact on how interventions are delivered and medical decisions are made.

References

[1] Mbunga, B. K., Mapatano, M. A., Strand, T. A., Gjengedal, E. L. F., Akilimali, P. Z., & Engebretsen, I. M. S. (2021). Prevalence of anemia, iron-deficiency anemia, and associated factors among children aged 1–5 years in the rural, malaria-endemic setting of Popokabaka, Democratic Republic of Congo: A cross-sectional study. Nutrients, 13(3), 1010

[2] Odigwe, B. E., Eyitayo, J. S., Odigwe, C. I., & Valafar, H. (2019). Modelling of Sickle Cell Anemia Patients Response to Hydroxyurea using Artificial Neural Networks. arXiv preprint arXiv:1911.10978.

[3] Ofakunrin, A. O. D., Oguche, S., Adekola, K., Okpe, E. S., Afolaranmi, T. O., Diaku-Akinwumi, I. N., et al. (2020). Effectiveness and Safety of Hydroxyurea in the Treatment of Sickle Cell Anaemia Children in Jos, North Central Nigeria. J. Trop.Pediatr. 66 (3), 290–298. https://doi.org/10.1093/tropej/fmz070

[4] El Hoss, S., El Nemer, W., & Rees, D. C. (2022). Precision Medicine and Sickle Cell Disease. HemaSphere, 6(9), e762. https://doi.org/10.1097/HS9.0000000000000762

Contact Information

Tel: +2348062688490
Email: daabiola3@gmail.com