OP 16 - Design and development of a machine learning-based predictive modelling tool to accurately predict thalassemia carrier state using full blood count indices and haemoglobin variants

C. A. D. M. N. C. Kolambage[1*], H. W. W. Goonasekara[2], R. Hewapathirana[2]

[1]Post Graduate Institute of Medicine, University of Colombo, Sri Lanka

[2]Department of Anatomy, Faculty of Medicine, University of Colombo, Sri Lanka

Introduction: This study attempted to use artificial intelligence (AI) techniques to develop a cost-effective and time-saving tool to accurately predict the alpha thalassemia carrier state using a simple blood test, without the need to perform genetic testing.

Methods: Ethical clearance was obtained from the Ethics Review Committee of PGIM, University of Colombo (ERC/PGIM/2020/068). Two AI algorithms, random forest and an Artificial Neural Network (ANN) were used to train three models using complete blood count and high-performance liquid chromatography data from a database of 288 cases from the Human Genetics Unit (HGU) of the Faculty of Medicine, Colombo. Performances of the models were assessed using two metrics: accuracy and F1 score, which is a measure of sensitivity and positive predictive value.

Results: Dataset included predominantly alpha thalassemia carriers (n = 149), but also individuals with normal red cell indices (n = 55), iron deficiency anaemia (n = 21) and beta thalassemia carriers (n = 17). Model 1 was trained to differentiate between alpha thalassemia carriers and normal individuals, and random forest outperformed the ANN with an Accuracy of 93.5% and an F1 score of 87.5%. Model 2 differentiated between alpha thalassemia silent carriers and alpha thalassemia traits, and random forest again showed superior performance with an accuracy of 88.8% and an F1 score of 86.3%. Model 3 differentiated between thalassemia carriers, iron deficiency, and normal individuals; but did not show satisfactory performance.

Conclusions: Model 1 and 2 showed satisfactory performance and can be combined into a diagnostic tool to identify alpha thalassemia carrier states. The tool could be used in situations where there are constraints to genetic testing, after validating prospectively with a larger dataset.

Keywords: alpha-thalassemia, thalassemia screening, artificial intelligence, predictive modelling, random forest

*Email: cmiduna@gmail.com