Decision Support for Diagnosing Thyroid Diseases Using Machine Learning

W. K. D. Jayamini[a], H. D. Weerasinghe[b,1]

aDepartment of Statistics and Computer Science, University of Kelaniya, Kelaniya, Sri Lanka
bDepartment of Computer Systems Engineering, University of Kelaniya, Kelaniya, Sri Lanka

Diagnosis of thyroid disorders using two machine learning techniques was studied in this research. Multilayer Perceptron Neural Network with Back-propagation algorithm and Random Forest algorithm were the two algorithms used to build the models for classifying the thyroid diagnosis classes; Hyperthyroidism, Hypothyroidism, Normal. Models were developed with different structures by changing the relevant parameters and the outcomes of the developed models were compared with each other. For developing different neural networks, parameters such as the number of hidden layers, number of neurons in hidden layers and learning rates were changed. For developing different random forest models, parameters such as the number of features per tree and the number of trees in forest were changed. Those models were trained and tested using two different datasets of thyroid diagnosis (Dataset 1 and Dataset 2) which have different attributes that are related to diagnosing thyroid diseases. The models were tested using 10-fold cross-validation while the models were compared and evaluated using the measures Accuracy (%), Mean Absolute Error, Root Mean Squared Error, TP rate, FP rate, Precision and Recall. In diagnosing thyroid disease, both the algorithms performed well. Multilayer Perceptron Neural Network with Back-propagation algorithm performed well for Dataset 1 with an accuracy of 96.7442% and Random Forest algorithm performed well for the Dataset 2 with a mean accuracy level of 98.4915%.

Keywords: Decision Support; Thyroid Disease Diagnosis; Back-propagation; Random Forest

[1] Corresponding author: H. D. Weerasinghe; Tel.: +94-77-801-1033
 E-mail address: darshawk@gmail.com