



Machine Learning Algorithms to Improve Early Diagnosis of Thyroid Disease: A Sri Lankan-Based Study

W.M.U. Senani* and S. Thirukumaran

Department of Physical Science, Faculty of Applied Science, University of Vavuniya, Sri Lanka

Abstract: Thyroid diseases are a major health, well-being, and health care problem across the world. This study proposes a new thyroid diagnosis system using efficient feature selection techniques and machine learning methods for better early-stage diagnosis. The study has utilized data gathered from the Uva and Western Provinces, both known to exhibit specific patterns regarding thyroid diseases. We used the method 'SelectKBest' to filter the top five most influential factors out of 23 features about thyroid disease. These were further supplemented by three more factors after consultation with clinical experts to make sure that the model captures clinically relevant insights. The study assessed the performances of four machine learning models: the Support Vector Machine, Decision Tree, Random Forest, and Neural Network. Among all the rest, the best model is Random Forest with an accuracy of 86.20%, closely followed by Neural Network with an accuracy of 85.05%, then Decision Tree with 83.90%, and Support Vector Machine with 81.60% accuracy. Besides accuracy, we are including main metrics for evaluation; that is to say, precision and recall with F1-score, which are important when performing a diagnosis model evaluation, specifically within a medical context. It therefore follows from the results that the Random Forest model provides the most reliable performance in early detection of thyroid disease in Sri Lankan populations. The pictorial presentation of the visual analysis and figures will be included, together with thyroid-related images, for better understanding of the results. This study is a great contribution to healthcare in that it could give rise to higher accuracy in diagnosis and, consequently, effective treatment among those-again, notably women-who suffer from thyroid disorders in Sri Lanka.

Keywords: Early Diagnosis, Machine Learning, Random Forest, Sri Lanka, Thyroid Disease