THYROID DETECTION USING MACHINE LEARNING

Abstract

The thyroid gland has one of the most important functions in regulating metabolism. When the function of the thyroid gland is affected, it leads to inappropriate production of the thyroid hormone. Hypothyroidism and hyperthyroidism are two critical conditions caused by insufficient thyroid hormone production and excessive thyroid hormone production, respectively. The "Thyroid Detection Using Machine Learning" project is focussed on detecting and diagnosing thyroid disease.

The performance of three machine learning algorithms such as Random Forest, Logistic Regression, Support Vector Machine are compared to classify Thyroid disease into normal, hypothyroidism, or hyperthyroidism categories. The most significant features, which can be used to detect thyroid diseases more precisely are identified using forward feature selection, backward feature elimination, bidirectional feature elimination, and machine learning-based feature selection.

The dataset is taken from the Kaggle repository. The dataset contains 9172 sample observations and has 31 columns including 1 identifier, 1 class variable and 29 features.

Among the three algorithms, Random Forest is found to be best in terms of computational time and accuracy score, which make it significant for the proposed approach. The model will classify under three classes which are no thyroid, hyperthyroid and hypothyroid. An automated system can be very helpful to assist medical experts and even make automated disease predictions without any human mistakes. Patients can diagnose their condition without the assistance of a medical expert.

Submitted By:

    Nikhila Baby

    MAC23MCA-2044

Faculty Guide:

    Prof. Nisha Markose

    Associate Professor

    MCA Dept, MACE