

Thyroid Detection

Prof. Mohammad Sajid¹, Isha Patil², Mayuri Balapure³, Riya Bawane⁴, Sonali Gadekar⁵,

Samiksha Tembhurne⁶

¹Prof. Mohammad Sajid, Computer Science & Engineering & Nagpur Institute of Technology, Nagpur

²Isha Patil, Computer Science & Engineering & Nagpur Institute of Technology, Nagpur

³Mayuri Balapure, Computer Science & Engineering & Nagpur Institute of Technology, Nagpur

⁴Riya Bawane, Computer Science & Engineering & Nagpur Institute of Technology, Nagpur

⁵Sonali Gadekar, Computer Science & Engineering & Nagpur Institute of Technology, Nagpur

⁶Samiksha Tembhurne, Computer Science & Engineering & Nagpur Institute of Technology, Nagpur

Abstract

Thyroid Gland is a vascular gland and one of the most important organs of the human body. This gland secretes two hormones which help in controlling the metabolism of the body. The two types of thyroid disorders are Hyperthyroidism and Hypothyroidism. When this disorder occurs in the body they release certain types of hormones into the body which imbalances body's metabolism. Thyroid related blood test is used to detect this disease but it is often blurred and noise will be present. Data cleansing methods were used to make the data primitive enough for the analytics to show the risk of patient's getting this disease. Machine Learning plays a very deciding role in the disease prediction. Machine Learning algorithms SVM-Support Vector Machine, KNN-K-nearest neighbours, Random Forest, K-Means Algorithms are used to predict the patient's risk of getting thyroid disease.

Key Words: Hyperthyroidism, Hypothyroidism, SVM, KNN, K-Means.

1. INTRODUCTION

The evolution of computational biology is used in healthcare industry. It allows collection of stored patient data for the prediction of the disease. There are prediction algorithms which are available for the diagnosis of the disease at early stages. The medical information systems are rich of datasets but there are

only few intelligent systems which can easily analyze the disease. Over a period of time, the machine learning algorithms have started playing a crucial role in resolving the complex and non-linear problems in the developing model. The thyroid gland is an endocrine gland present in the human neck beneath the Adam's apple which helps in secretion of thyroid hormone that influences the rate of metabolism and protein synthesis. The thyroid hormones are useful in counting how briskly the heart beats and how fast we burn calories. The thyroid secretes two types of active hormones called levothyroxine (T4) and triiodothyronine (T3). These hormones help in regulating the body temperature. These also aid in energy-bearing and transmission in every part of the body and are decisive in protein management. Iodine is considered as the main building block of the thyroid gland. It is prostrated in few specific problems. Undersupply of these hormones can lead to hyperthyroidism. There are many originations related to hyperthyroidism and underactive thyroids. There are various kinds of medications like thyroid surgery is liable to ionizing radiation, continual tenderness of the thyroid, deficiency of iodine and lack of enzyme to make thyroid hormones.

2. PROBLEM STATEMENT

According to statistics, thyroid disorders are on the rise in India. Approximately 1 in 10 Indian adults suffer from thyroid problem. It has been estimated that around 42 million people suffer from thyroid disease. Predicting thyroid disorder by doctor is a tedious process which might lead to negative prediction, only experienced doctor can examine the

case properly. To assist doctors machine learning can help them in diagnosis of disease and reduces their burden.

3. OBJECTIVES

- The main objective is to develop a system which can predict the type of thyroid disease that patient is affected from.
- To predict thyroid disease with usage of minimum number of parameters.
- To predict all possible types of Thyroid diseases.

5. METHODOLOGY

Supervised learning is an information mining undertaking of inferring a function from named training information. The training information comprised of an arrangement of preparing illustrations. In managed adapting, every case is a couple comprising of an information input object (commonly a vector) and the desired output value (additionally called the supervisory flag). A supervised learning calculation investigates the training information and produces an indirect function, which can be utilized for mapping new illustrations. An ideal improvement will take into account the calculation to effectively decide the class names for unseen cases. This requires the taking in calculation to sum up from the training information to hidden circumstances in a “sensible” manner.

1. Performance Study of the proposed Algorithms:

K- Nearest Neighbor: When given a training tuple KNearest Neighbor simply stores it and waits until it is given a test tuple. Hence it is a “lazy learner” as it stores the training tuples or the “instances”, they are also known as “Instance- Based Learners” [21-22]. Thus, k is a positive integer and decides how many neighbors influence the classification. “Closeness” delineates as a distance metric such as “Euclidean Distance” or “Manhattan Distance”.

3. Random Forest Classifier: Random forest is a commonly-used machine learning algorithm trademarked by Leo Breiman and Adele Cutler, which combines the output of multiple decision trees to reach a single result. Its ease of use and flexibility have fueled its adoption, as it handles both classification and regression problems.

4. XGBoost: XGBoost stands for “Extreme Gradient Boosting” and it has become one of the most popular and widely used

Machine learning algorithms to handle large datasets and its ability to achieve state-of-the-art performance in many machine learning tasks such as classification and regression.

6. RESULT

The datasets for the thyroid disease have been possessed by the UCI machine learning repository. The work is dwelled with 2 different stages. The foremost phase comprised of the subset selection that is executed by adapting mutual information and prediction of the thyroid datasets done using ANN. Specifically in the interpretation of diseases neural networks are successfully enforced in the distinctive fields in the medical. The certainty of the analysis for the datasets of the thyroid diseases are assigned as the elected appearance by every feature selection algorithm. Indeed, the expansion of our unified representative is a constructive mechanism to predict thyroid disease based on the limited dataset that is available with us. The model can further be enhanced to any desired level by increasing the number of inputs and outputs and dynamic data can be generated as more data can be fed to it. In a nutshell, not only we have developed a prototype integrated frame work to diagnose the thyroid disease but also act as a decision maker for diagnosing the thyroid disease.

7. REFERENCES

- [1] Ankita Tyagi and Ritika Mehra. (2018). “Interactive Thyroid Disease Prediction System using Machine Learning Techniques” published on ResearchGate.
- [2] YongFeng Wang, (2020). “Comparison Study of Radiomics and Deep-Learning Based Methods for Thyroid Nodules Classification using Ultrasound Images” published on IEEEAccess.
- [3] Sunila Godara, (2018). “Prediction of Thyroid Disease Using Machine Learning Techniques” published on IJEE.
- [4] Hitesh Garg, (2013). “Segmentation of Thyroid Gland in Ultrasound image using Neural Network” published on IEEE.
- [5] L. Ozyilmaz and T. Yildirim, (2002). “Diagnosis of thyroid disease using artificial neural network methods,” in: Proceedings of ICONIP’02 9th international conference on neural information processing (Singapore: Orchid Country Club, pp. 2033–2036).
- [6] K. Polat, S. Sahan and S. Gunes, (2007) “A novel hybrid method based on artificial immune recognition system (AIRS) with fuzzy weighted preprocessing for thyroid disease diagnosis,” Expert Systems with Applications, (vol. 32, pp. 1141-1147).
- [7] F. Saiti, A. A. Naini, M. A. Shoorehdeli, and M. Teshnehlab, (2009) “Thyroid Disease Diagnosis Based on

Genetic Algorithms Using PNN and SVM,” in 3rd International Conference on Bioinformatics and Biomedical Engineering. ICBBE 2009.

[8] G. Zhang, L.V. Berardi,(2007) “An investigation of neural networks in thyroid function diagnosis,” Health Care Management Science,1998, (pp. 29-37.)

[9] V. Vapnik,(2012).Estimation of Dependences Based on Empirical Data, Springer, New York.

[10] Obermeyer Z,(2016). Emanuel EJ. Predicting the future— big data, machine learning, and clinical medicine; (375:12161219).

[11] Breiman L.(2001) Statistical Modeling: the two cultures.Stat Sci. ;16:199-231..

[12] Ehrenstein V, Nielsen H, Pedersen AB, Johnsen SP, Pedersen L. (2017) Clinical epidemiology in the era of big data: new opportunities, familiar challenges. Clin Epidemiol. ; 9:245-250

[13] S. Godara and R. Singh,(2016) "Evaluation of Predictive Machine Learning Techniques as Expert Systems in Medical Diagnosis", Indian Journal of Science and Technology, (Vol. 910). [14] Sunila, Rishipal Singh and Sanjeev Kumar.(2016) "A Novel Weighted Class based Clustering for Medical Diagnostic Interface." Indian Journal of Science and Technology (Vol 9).