Thyroid Disease Detection

Detailed Project Report

Project by: Ahmad Taquee, Priyanka Gupta, Diksha Sharma

Contents

Abstract

Thyroid gland plays a major role in maintaining the metabolism of human body. Data mining in health care industry provides a systematic use of the medical data. Thyroid diseases are most common today. Early changes in the thyroid gland will not affect the proper working of the gland. By the early identification of thyroid disorders, better treatment can be provided in the early stage thus can avoid thyroid replacement therapy and thyroid removal up to an extent.

1. Introduction

The thyroid gland is one of the important organs of the human body. It produces the thyroid hormone which is vital to control the body's metabolism. Thyroxine and triiodothyronine are two active thyroid hormones that have important effects on protein production, body temperature regulation, energy production, and energy regulation of the human body. Therefore, if the thyroid gland is diseased, the metabolism and regulation of the human body will lose the necessary control and that may be life threatening in severe cases.

Thyroid disease has now become the second largest disease in the endocrine field [1], which can lead to death when the disease is severe. The total number of patients with thyroid disorders worldwide is more than 300 million, of whom the number of females is about 6~10 times of that of male patients, and the number of females over 40 years is about 10%~20%. In China, there are more than 40 million people with primary hypothyroidism and more than 10 million people with primary hyperthyroidism. The treatment rate for hyperthyroidism in China is less than 5%.

In practical clinical practice, many approaches can be used to diagnose thyroid diseases, such as clinical evaluation, blood examination, thyroid hormone (TSH) detection, imaging examination, and tissue biopsy. The comprehensive application of various detection methods has been very common in clinical diagnosis, such as the combined use of TSH detection data and blood examination data.

2. Business Problem Statement

Thyroid disease is a common cause of medical diagnosis and prediction, with an onset that is difficult to forecast in medical research. The thyroid gland is one of our body's most vital organs. Thyroid hormone releases are responsible for metabolic regulation. Hyperthyroidism and hypothyroidism are one of the two common diseases of the thyroid that releases thyroid hormones in regulating the rate of body's metabolism. The main goal is to predict the estimated risk on a patient's chance of obtaining thyroid disease or not.

3. Related Work

In the CAD study of thyroid disease, a large amount of literature focuses on the diagnostic research employing TSH data due to the open dataset of TSH in the UCI machine learning repository. In this dataset, thyroid diagnosis is considered as a classification problem with three classes of normal, hyperthyroidism, and hypothyroidism. If we consider the diagnosis of thyroid disease as a classification problem, we can introduce a powerful machine learning technology to discover the complex relationship of biomedicine to build CAD system. With the development of machine learning technology, the recognition accuracy of this problem has been improved gradually. The diagnosis accuracy was 78.14% when employing probabilistic neural network method in 1997 [2]. The accuracy achieved 88.3% when multilayer perception was used in 2004 [3]. The accuracy rate of wavelet-based support vector machine (SVM) recognition was 91.86% in 2011 [4], and the SVM method with particle swarm optimization achieved the accuracy of 97.49% in 2012 [5]. And extreme learning machine method (ELM) has achieved the accuracy rate of 97.73% [6].

4. Proposed Method

In the proposed solution here, we used Random Forest machine learning model to classify the different thyroid disease type. Here, we are performing Data preprocessing step, in which feature engineering, feature selection, feature scaling steps are performed and then we are doing model building and performance testing.

5. Data

Data Requirement completely depend on our problem. For training and testing the model, we are using Thyroid disease dataset which is available on UCI portal. From user we are taking following input

Names of attributes are:

age, sex, on thyroxine, query on thyroxine, on antithyroid medication, sick, pregnant, thyroid surgery, I131 treatment, query hypothyroid, query hyperthyroid, lithium, goitre, tumor, hypopituitary, psych, TSH, T3, TT4, FTI.

6. Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

7. <u>Software and Tools</u>

- PyCharm is used as IDE.

- Matplotlib and Seaborn are used for visualization of the plots.

- Heroku is used for deployment of the model.

- Mongo dB is used to retrieve, insert, delete, and update the database.

- Front end development is done using html, CSS, bootstrap

- Flask is used for backend development and for API development.

- GitHub is used as version control system.

8. <u>Conclusion</u>

This project proposed a method for the classification of thyroid diseases using Random Forest machine learning.

9. <u>Timeline - Duration of Project</u>

Data collection: 1 Week

Data Cleaning and EDA: 3 Week

Model building and Evaluation: 2 Week

Deployment: 1 Week

Documents Preparation: 1 Week

References

- K. Chandel, V. Kunwar, S. Sabitha, T. Choudhury, S. Mukherjee, "A comparative study on thyroid disease detection using K-nearest neighbor and Naive Bayes classification techniques", pp. 1-7, 2017.
- H. Kumar H S, "A Novel Approach of SVM based Classification on Thyroid Disease Stage Detection", pp. 836-841, 2020.
- S. Razia, M. R. N. Rao, "Machine Learning Techniques for Thyroid Disease Diagnosis - A Review", vol. 9(28), pp. 1-9, 2016.
- Z. Parry and R. Macnab, "Thyroid disease and thyroid surgery," *Anaesthesia & Intensive Care Medicine*, vol. 18, no. 10, pp. 488–495, 2017.View at: Publisher Site | Google Scholar
- G. Serpen, H. Jiang, and L. Allred, "Performance analysis of probabilistic potential function neural network classifier," in Proceedings of Artificial Neural Networks in Engineering Conference, vol. 7, pp. 471–476, St. Louis, MO, USA, 1997.View at: Google Scholar
- Ozyilmaz and T. Yildirim, "Diagnosis of thyroid disease using artificial neural network methods," in Proceedings of International Conference on Neural Information Processing, pp. 2033–2036, Singapore, 2002.View at: Google Scholar
- E. Dogantekin, A. Dogantekin, and D. Avci, "An expert system based on generalized discriminant analysis and wavelet support vector machine for diagnosis of thyroid diseases," Expert Systems with Applications, vol. 38, no. 1, pp. 146–150, 2011.View at: Publisher Site | Google Scholar
- H.-L. Chen, B. Yang, G. Wang, J. Liu, Y.-D. Chen, and D.-Y. Liu, "A three-stage expert system based on support vector machines for thyroid disease diagnosis," Journal of Medical Systems, vol. 36, no. 3, pp. 1953–1963, 2011.View at: Publisher Site | Google Scholar