A SURVEY OF THYROID DETECTION USING DATAMINING TECHNIQUES

A.Vinitha

Assistant Professor, Department of Computer Science,
A.V.P College of Arts and Science,
Tirupur, Tamilnadu, India.

ABSTRACT

The health is the major issue in today's world. Most of them are suffering from improper secretion of hormones. The improper hormone secretion arises due to lack of proper healthy food. Among the one is the thyroid deficiency. The thyroid problem occurs due to improper secretion of hormones in our body. The women are the most which are suffering from the thyroid problem. In this paper we will we will review some classification techniques which helps detect the thyroid problem.

KEYWORDS: *T3, T4 hormone, classification techniques*

1. INTRODUCTION

The classification is the one of the data mining technique which helps to classify the data items according to the predefined class labels. For example if the bank is offering loan means the bank manager has to classify the customers according to the predefined class labels like low risk medium risk and high risk. The bank manager should know the risk of providing loan to the customers. There are different types of classification techniques are there. The thyroid gland is located at the neck. It exactly looks like butterfly. It helps to maintain the development and growth of the human body. It helps in metabolism process. The metabolism process is nothing but whatever we eat and drink is converted into energy by the hormones secreted by the thyroid gland. This is the main work of hormones of thyroid gland. If the hormones are not secreted properly the thyroid deficiency occurs and our intake food is not properly converted into energy.

2. LITERATURE SURVEY

The author states that (Umar Sidiq, 2019) thyroid deficiency occurs due to the deficiency of iodine. The author uses the Anaconda 3 5.2 which helps to find the accuracy in the classification algorithms. The author compared the k- nearest neighbour, support vector machine, naïve bayes, decision tree. The author compares the accuracy in above stated algorithms and tells the decision tree offers the high accuracy in detecting the thyroid problem. The accuracy is about 98.89 percentage. The

author concluded that the decision tree can be combined with any other algorithm to get higher accuracy. The author has downloaded the data set from one recognized lab from Kashmir.

The author (Bibi Amina Begum, 2019) compares different data mining techniques and compares the accuracy of the algorithm. Each algorithm gives different levels of accuracy. Hence any one classification algorithm can be used to detect the thyroid problem.

The author (Aswathi A K, 2018) states the earlier identification of thyroid disease will help to avoid the replacement of thyroid gland and severe health issues. The author uses the classification technique in that he used the support vector machine for classification. The naïve bayes is used to find the missing values in the data. After detecting the thyroid disease the author focused on giving some medical advice. The swarm optimization technique is also used to find the thyroid deficiency.

The author states (Ankita Tyagi, 2018) different types of classification algorithms like support vector machine, k nearest neighbour, decision trees are used to predict the thyroid deficiency. The author also uses machine learning algorithm called as artificial neural network which helps to find the thyroid problem. The author compared all the algorithms in finding the accuracy of detecting thyroid problem. The author finally suggested that the parameters used for detecting the thyroid problem should be minimized

EPRA International Journal of Research and Development (IJRD)

because the patient should undergo more test and it also incurs more expenses and time consuming. The conclusion of the paper is to detect the thyroid problem with minimum parameters.

The author (Muhammad Anjum Qureshi, 2017) says that the thyroid disease can be predicted by using the online mode. The decision tree is used to predict the thyroid deficiency. Feature rejection, online ensemble, offline ensemble are the techniques used in detecting the thyroid problem. The accuracy of using decision tree is about 97.43 percentage. The feature ensemble helps to reduce the computational complexity.

The author (Qiao pan, 2016) tells that the thyroid deficiency is the major health issue in today's environment. The author uses the random forest to detect the thyroid problem. The accuracy of using random forest method is 95.63%.

The author (K.Sindhya, 2020) has stated that the predictive classification algorithm should take less time to predict he disease. The author compares with random forest, naïve bayes and J48 algorithm. Among these techniques the J48 produced the low time and highest accuracy at 99%.

3. RELATED WORK

There are different types of classification techniques which help to predict the thyroid problem. There are two types of disease in thyroid gland. They are hypothyroid and hyper thyroid. The hypothalamus and pituitary gland which creates the thyroid stimulating hormone which tells the thyroid gland to produce the more hormone if it is low. If the thyroid secretion is high then it will make the thyroid gland to produce fewer hormones. Thus thyroid gland should maintain the metabolism of the body. The hormone produced by thyroid gland is T3 and T4.
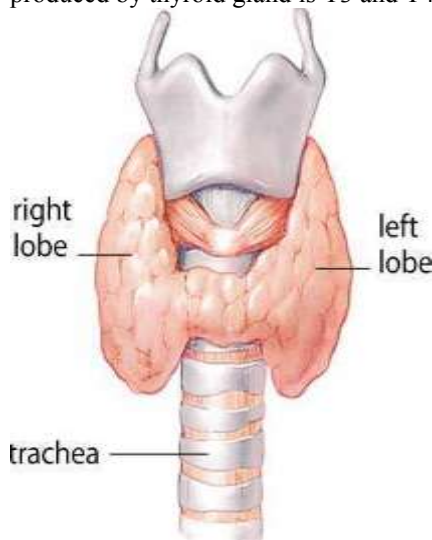

Fig 1: Thyroid gland

4. CLASSIFICATION TECHNIQUES

There are different techniques in classification. We shall discuss few techniques which helps to detect the thyroid problem.

Naïve Bayes

It is based on bayes theorem. The features of naïve bayes it should be independent of another. It is one of the classification algorithms. No pair of features is dependent on other (P Deepika, 2016).

Decision tree

It is one of the supervised learning techniques. It can solve both classification and regression problems. It uses tree representation to solve the problem of classification. The leaf node represents the class labels and internal nodes represent the attributes. The attributes are selected by using two measures called as information gain and Gini index. (K.Thenmozhi. P. M., 2014)

Artificial Neural networks

It is an information processing model. It exactly looks like human brain. It has many layers which help to give maximum accuracy. The neural network is the supervised learning technique. In supervised learning there will be one input and one output. The algorithm should learn by example. It mainly focuses on maximum accuracy. It helps to solve the classification and regression problems. The output of one node is connected with input of another node. Each node has weights which help to increase the accuracy.

Support Vector Machines

It is also a supervised learning technique. It represents the attributes as points in the hyper plane (Vinitha, 2015). It helps to separate the data items by developing the hyper plane in to two items. The support vector machine can perform well both in linear and non-linear. In support vector machine we plot each point in the hyper plane and finally the plane that separates the two groups is identified. (P.Deepika S. D., 2019)

5. DISCUSSIONS

The different types of classification algorithms are used to predict the thyroid problem. Hence each algorithm generated different levels of accuracy.

S.No	Methodology	Accuracy
1	Decision Tree	98.89%
2	Random forest method	95.63%
3	J48	99%
4	Decision Tree	97.43%

These are the different levels of accuracy generated by different types of classification algorithm (P.Deepika D. S., 2018). Among this the J48 is the algorithm which gives maximum accuracy and less time to detect the thyroid deficiency. The above stated accuracy is shown as pie chart. (K.Thenmozhi. P. , 2014)
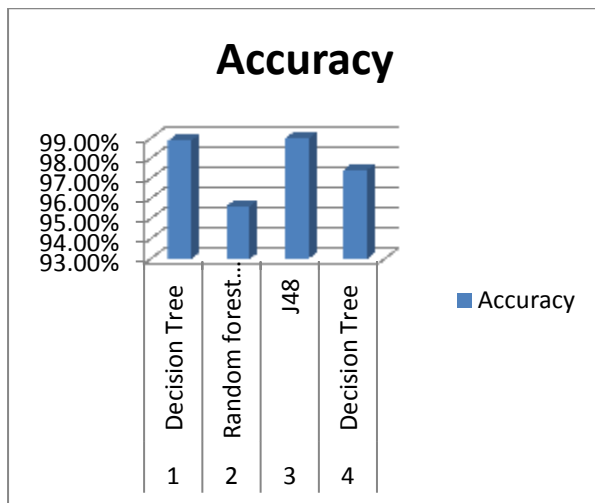
6. CONCLUSION

The thyroid problem is nowadays found in many people. Early identification of thyroid problem will help to recover from the thyroid. The reviewed survey will help to detect the thyroid problem early. Many classification algorithms are discussed. Among that the J48 is the algorithm which provides the maximum accuracy and it is also less time consuming algorithm. So in future the J48 can be combined with any other advanced methodology to give maximum accuracy and less time consuming.

7. REFERENCES

Bibliography

1. A Kiruthika, P. D. (2018). Predicting Ailment of thyroid using Classification and Recital Indicators. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 4.
2. A vinitha, V. J. (2013). Classification Algorithms in Intrusion Detection System: A Survey. International Journal of Computer Technology and Applications, 4.
3. Ankita Tyagi, R. M. (2018). Interactive Thyroid Disease Prediction System Using Machine Learning Technique. 5th International Conference on Parallel, Distributed and Grid Computing , 5.
4. Aswathi A K, A. A. (2018). An Intelligent System for Thyroid Disease Classification and Diagnosis. 2nd International Conference on Inventive Communication and Computational Technologies, 4.
5. Bibi Amina Begum, D. P. (2019). Prediction of thyroid Disease Using Data Mining Techniques. 5th International Conference on Advanced Computing and Communication Systems, 4.
6. K.Sindhya, M. (2020). Effective Prediction of Hypothyroid Using Various Data Mining Techniques. EPRA International Journal of Research and Development , 7.
7. K.Thenmozhi., P. (2014). Heart Disease Prediction Using Classification with different Decison Tree Techniques. International Journal of Engineering Reseacrh and General Science, Vol-2, Issue-6.
8. K.Thenmozhi., P. M. (2014). Different Data Mining Techniques involved in Heart Disease Prediction. International Journal of Scientific Research, Vol-3.
9. Muhammad Anjum Qureshi, K. E. (2017). Expert Advice Ensemble for Thyroid Disease Diagnosis. IEEE , 4.
10. P Deepika, S. S. (2016). Anticipating Heart Disease using C4. 5 Classification Augmented with Feature Selection. International Journal for Research & Development in Technology, Vol-6.
11. P.Deepika, D. S. (2018). A Noval Classification and Prediction Algorithm for Heart disease Identification. International Journal of Engineering Science Invention, Vol-7, Issue-2.
12. P.Deepika, S. D. (2019). An Effective Archetype Design of Heart Disease Anticipation using Optimization Techniques. EPRA International Journal of Research and Development (IJRD), vol-4, Issue-11.
13. Qiao pan, Y. Z. (2016). Improved Ensemble Classifcation Method of Thyroid Disease Based on Random Forest. 8th Internatioanl Conference on Information Technology in Medicine and Education, 5.
14. Umar Sidiq, D. M. (2019). Diagnosis of various Thyroid Ailments using Data Mining Classification Techniques. International Journal of Scientific Research in Computer Science, Engineering and Information Technology , 6.
15. Vinitha, V. J. (2015). Intrusion Detection System Using Multiclass Batch Algorithm. International Journal of Advanced Research in Computer and Communication Engineering, Vol-4, Issue-2.