

# A Study of Thyroid Prediction using Big Data Analysis

P. Praveena<sup>1</sup>, Sylvia Mary<sup>2</sup>

<sup>1</sup>Student, Department of Computer Science, Women's Christian College, Chennai, India

<sup>2</sup>Assistant Professor & HoD., Department of Computer Application, Women's Christian College, Chennai, India

**Abstract:** Thyroid is a world's common disease. One of the unpredictable disease in the medical field. The thyroid gland produces the thyroid hormone which helps us to control many process in our body like growth, energy balance, heat temperature, heart beat rate etc. The thyroid hormone fails to work due to some illness like excess of heat, depression, stress, etc and it leads us to suffer from this disorder. If we are able to predict the thyroid disorder at an earlier stage we can prevent from thyroid cancer. Because prevention is better than curing the disorder particularly as thyroid cancer is one of the treatable diseases. Big Data analytics is the best solution for prediction. Data Analytics plays an important role in health care especially in prediction. This paper presents methods of prediction of the different types of thyroid disease at an earlier stage.

**Keywords:** Thyroid disease, hypothyroidism, hyperthyroidism, Big data, Naïve Bayesian Algorithm.

## 1. Introduction

The most difficult task is to diagnose disease at an earlier stage with high accuracy in the medical field. The Big Data analysis is a technique, which are used in health care organizations for predicting, decision making, diagnosing and to give a better treatment for the patients. Many disease can be predicted using big data analytics for eg., lung cancer, breast cancer, brain tumor diabetics, thyroid disease etc. Big Data analytics have the potential to reduce the cost and energy of treatment, predicting the outbreak of the disease, preventing at an earlier stage and improve the quality of life. It poses the new challenges to today's treatment methods. This paper is about thyroid prediction using big data Analytics Thyroid disease is a calm and upcoming dangerous disease. The thyroid gland looks like a butterfly shape situated in the front of the neck. The thyroid gland helps us to maintain our metabolism rate in our body. Thyroid diseases are many types and they are hypothyroidism, hyperthyroidism, thyroid cancer, grave disease, Hashimoto's, goiter etc. The most two common disorder of this thyroid disease are hyperthyroidism and hypothyroidism. This leads to upset the normal balance of chemical reactions in our body. Thyroid disease leads to cardiac effects because it can make unpredictable pulse and harm to heart muscle. The most noticeably horrible of these symptoms indications are cardiovascular impacts

## 2. Thyroid disease

Thyroid disease leads to different infections in our body like coronary illness, corpulence, bareness and in some cases it prompts to tumor. Mostly women's are suffering from this disorder than men. Newly conceived youngsters, pregnant ladies and teenagers are mostly suffering from this disease. Nowadays 2 out of Ten people suffer from this disorder at age of 12 to 60.

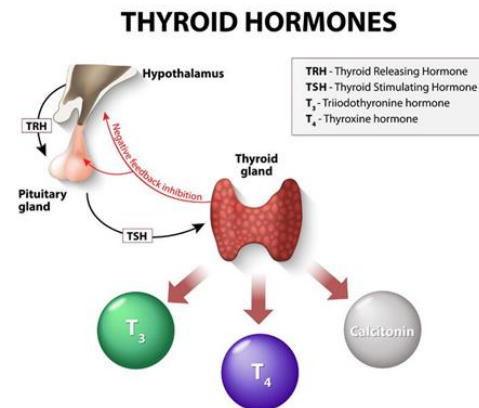


Fig. 1. Thyroid hormones

The common symptoms of thyroid disease are swelling of hands and legs, a small bulge in front of neck, abnormal body temperature, itching skin, unexpected weight gain or loss, memory loss, dry skin etc. These awful symptoms leads to cardiac disease. Thyroid disease can be classified into two 1.hypothyroidism and 2.hyperthyroidism.

### A. Hyperthyroidism

Hyperthyroidism occurs when the thyroid gland produces enormous amount of thyroid hormones in our body. It is also called as overactive thyroid. Hyperthyroidism is mainly caused by auto immune disorder grave disease. The common symptoms of Hyperthyroidism are.

- Restlessness
- Shaking
- Anxiety
- Abnormal menstruation
- Feeling tired

- Weight loss
- Muscle weakness
- Increased sweating
- Delayed puberty
- Bulging Eyes(in Grave's disease)

**B. Hypothyroidism**

Hypothyroidism is opposite of hyperthyroidism. It occurs when the thyroid gland produces low level of thyroid hormone in our body. It is called as non-active thyroid. It is mainly caused by a damage in the thyroid gland or inflammation. The common symptoms of Hypothyroidism are.

- Dry skin
- Weight gain
- Joint pains
- Abnormal menstruation
- Depression
- Memory loss
- Slow heart rate
- Back pain
- Sensitivity to cold
- Constipation

**3. Existing system**

In existing system, diagnosing a disease is too difficult, because it needs lot of knowledge and experience. Many data mining techniques are used to predict the thyroid disease and their level. Currently no effective method is available for predicting the disease. To overcome this problem an effective classification method is needed.

*Disadvantages*

- Difficult to find the accurate level
- Large Time Consuming process

**4. Proposed system**

The Proposed thyroid prediction method utilizes the Big Data Analysis and information mining strategies. This type of analysis can find and concentrate on veil information related to the thyroid infirmity from a relate thyroid datasets. It can give complex hearing for thyroid spot. It can help us to deprecate the treatment cost. Here, Naïve Bayesian algorithm is used for classification. Naïve Bayes is a challenging algorithm in big data analysis. It needs only a small amount of training datasets. Bayes is a probabilistic technique for constructing a classifier. In this way the data are classified in efficient way to provide accurate results.

*Advantages*

- The results are accurate
- Short time to find the results
- Reduced Time and Cost

**5. Big data analysis**

Big data analysis is a form of advanced analytics which

involves complex applications with elements such as predictive models, statistical algorithms etc.

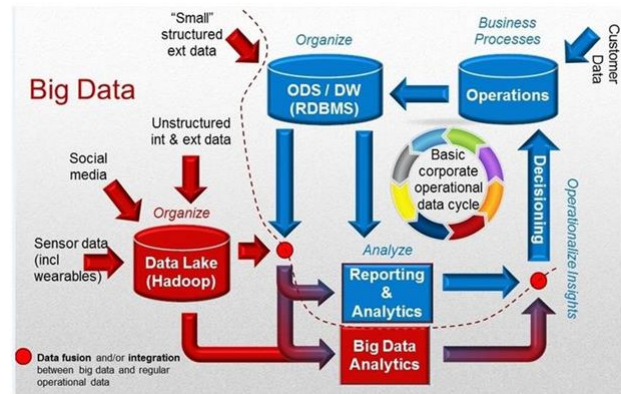


Fig. 2. Big data architecture

Big data is derived by specialized analytics system and software as well as high powered computing system. It offers various business benefits including new revenue opportunities more effective marketing better customer service, improved operational effective and competitive advantages. Big data analytics plays an important role in prediction especially in healthcare organization. It can give efficient result while comparing to other methods of prediction. To store and pre process the data's HADOOP software is used.

HADOOP is an open source distributed processing framework that manages the pre-processing and storage for big data applications running in clustered-systems. It is a centre of a growing ecosystem of big data technologies. They are primarily used for advanced analytics including predictive analysis, data mining and machine learning application. It can handle various forms of structured and unstructured data and gives more flexible to the user.

**6. Naïve bayesian algorithm**

It was proposed by Thomas Bayes. Naive Bayesian Algorithm is a statistical method for classification and it is an supervised algorithm. It can solve the problems both the categorical and continuous valued attributes. It is very simple and easy to implement. It Need less training data and is used for both binary and multi- I class classification problem. It has high accuracy and speed when applied to large amount of data. When the database is too large, some algorithm might not scale up well. It is a conditional probability model, represented by a vector

$$x = (x_1, \dots, x_n)$$

$$p(C_k/x, \dots, x_n)$$

The conditional probability can be decomposed as..,

$$p(C_k/x) = \frac{p(C_k)p(x|C_k)}{p(x)}$$

Here  $n$  is the number of features,  $C$  is a class variable, for each of  $k$  possible outcomes.

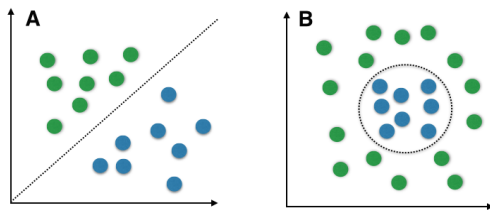


Fig. 3. Naive Bayesian classifier

Linear (A) vs. non-linear problem(B). Random samples for two different classes are shown as coloured spheres, and the dotted lines indicate the class boundaries that classifiers try to approximate by computing the decision boundaries. A non-linear problem (B) would be a case where linear classifiers, such as naive Bayes, would not be suitable since the classes are not linearly separable. In such a scenario, non-linear classifiers (e.g., instance-based nearest neighbour classifiers) should be preferred. There are three types of Naive Bayesian classifier

- Gaussian
- Multinomial
- Bernoulli

Naïve Bayesian algorithm is an effective technique for classifying data in real time and also in predicting future. Bayes theorem is one of the challenges in analysing Big Data is of course its volume of data. There are four applications they are as follows.

1) *Real time prediction*

It is a eager algorithm and it could be used for making prediction in real time.

2) *Multi class prediction*

This algorithm can predict the probability of multiple classes of target variable

3) *Text classification/sentiment analysis*

This algorithm mostly used in text classification. It has more success result as compared to other algorithm. It is mostly used in spam filtering.

4) *Recommendation system*

This system uses machine learning and data mining techniques to filter hidden information from the datasets.

### 7. Conclusion

This study paper analyzed the thyroid prediction using big data analysis. The methodology gives more efficient result in both predictive modelling and retrieving strategies. In this paper we utilized a naïve bayes hypothesis to predict patients with thyroid disease, we study the effectiveness and performance

analysis of our proposed system with an experiment set, consisting of scalability and quality. The future scope of our research is to conduct big data analysis by using different methodology to calculate the similarities between the patient's symptoms among hazard computation devices for more advanced predictive types. To extend our proposed framework and to predict other clinical risk like breast cancer, lung cancer, brain tumour, diabetics, skin diseases etc.

### References

- [1] G.K. Gupta. Introduction to Data Mining with Case Studies, 2006, Prentice Hall India.
- [2] Gerstein HC. How common is postpartum thyroiditis? Archives of Internal Medicine 1990 150 1397–1400.
- [3] <http://en.wikipedia.org/wiki/Thyroid>
- [4] [http://en.wikipedia.org/wiki/Thyroid\\_cancer](http://en.wikipedia.org/wiki/Thyroid_cancer)
- [5] Amino N, Tada H & Hidaka Y. Autoimmune thyroid disease and pregnancy. Journal of Endocrinological Investigation 1996, vol. 19, pp. 59–70.
- [6] Ginsberg J & Walfish PG. Post-partum transient thyrotoxicosis with painless thyroiditis. Lancet 1977, pp. 1125–1128.
- [7] Papageorgiou E, Kotsioni I & Linos A, (2005) “Data mining: a new technique in medical research”, Hormones, Vol. 4, pp. 189-191.
- [8] Amino N, Miyai K, Onishi T, Hashimoto T, Arai K, Ishibashi K et al. Transient hypothyroidism after delivery in autoimmune thyroiditis. Journal of Clinical Endocrinology and Metabolism 1976, vol. 42, pp. 296-301.
- [9] Kakulapati, Vijayalakshmi. (2018). Intelligent Thyroid prediction system using Big data. International journal of computer science and engineering. 6. 348-353.
- [10] Banu, Gulmohamed & Phil, M & Baviya Mca, M & Ali, Murtaza. (2015). A Study on Thyroid Disease Using Data mining Algorithm. International Journal of Technical Research and Applications, vol. 3, pp. 2320-8163.
- [11] Lewiński A, Sewerynek E, Karbownik M: Aging processes and the thyroid gland. In Aging and Age-Related Diseases: The Basics. Edited by: Karasek M. New York: Nova Science Publishers, Inc; 2006:131–172.
- [12] Faggiano A, Del Prete M, Marciello F, Marotta V, Ramundo V, Colao A: Thyroid diseases in elderly. Minerva Endocrinol 2011, 36: 211–231. PubMedGoogle Scholar International Journal of Computer Sciences and Engineering Vol.6(1), Jan 2018,
- [13] Papaleontiou M, Haymart MR: Approach to and treatment of thyroid disorders in the Elderly. Med Clin North Am 2012, 96: 297–310.
- [14] Bahn, R., Burch, H, and Cooper, D, “Hyperthyroidism and Other Causes of Thyrotoxicosis: Management Guidelines of the American Thyroid Association and American Association of Clinical Endocrinologists. Endocrine Practice,” vol. 17, no. 3 May/June 2011.
- [15] Braverman, L, Cooper D. Werner & Ingbar's the Thyroid, 10th Edition. WLL/Wolters Kluwer; 2012.
- [16] Rishitha Banu, “Predicting thyroid disease using data mining Technique” International Journal of Modern Trends in Engineering and Research on 11 October 2016, pg: 666-670.
- [17] Senthil Kumar, “Classification of Multi-dimensional Thyroid Dataset Using Data Mining Techniques: Comparison Study” Advances in Natural and Applied Sciences, 9(6) Special 2015, pp. 24-28.
- [18] Rasitha Banu, and Baviya “A study on Thyroid disease using Data Mining Technique,” IJTRA Journal, August 2015.
- [19] Banu, “Predicting Thyroid Disease using Linear Discriminant Analysis (LDA) Data Mining Technique”, Communications on Applied Electronics (CAE), vol. 4, no. 12, January 2016.