International Journal of Technical Research and Applications e-ISSN: 2320-8163,
www.ijtra.com Volume 3, Issue4 (July-August 2015), PP. 376-379

A STUDY ON THYROID DISEASE USING DATAMINING ALGORITHM

Dr. G. Rasitha Banu MCA., M.Phil.,Ph.D[1] , M.Baviya MCA[2] ., Dr.Murtaza Ali[3]
[1]Assistant Professor, Department of HIT&HIM, MPhil, Jazan University, Jazan KSA
[2]Department of Computer Science, Mother Teresa Women's University Chennai, India,
[3]HOD, Department of HIT&HIM, Jazan University,Jazan, KSA.
[1]rashidabanu76@gmail.com
[2]baviyabalaji@gmail.com
[3]muAli@jazan.edu.sa

***Abstract*— Data mining has been used intensively and extensively by many organizations. In healthcare, data mining is becoming increasingly popular, if not increasingly essential. Data mining applications can greatly benefit all parties involved in the healthcare industry. For example, data mining can help healthcare insurers detect fraud and abuse, healthcare organizations make customer relationship management decisions, physicians identify effective treatments and best practices, and patients receive better and more affordable healthcare services. The huge amounts of data generated by healthcare transactions are too complex and voluminous to be processed and analyzed by traditional methods. Data mining provides the methodology and technology to transform these mounds of data into useful information for decision making. There are two main methods of Data mining: Clustering and Classification. In many cases the concept of classification is confused by means of clustering, but there is difference between these two methods. According to the perspective of Machine learning clustering method is unsupervised learning and tries to group sets of objects having relationship between them, whereas classification method is supervised and assigning objects to sets of predefined classes. in proposed system are classified and cluster of the thyroid disease in data mining.**

***Index Terms*— DSCN algorithm, Hierarchical multiple classifier, TSH, Hypothyroidism, Hyperthyroidism, clustering, classification.**

I. INTRODUCTION

Hypothyroidism is a relatively common problem worldwide often with insidious onset and is relatively asymptomatic. Respiratory manifestations are seldom the major complaints in hypothyroidism. Lung volumes are usually normal, but few studies have shown findings suggestive of restrictive pattern of impairment. This has been attributed to decreased in both expiratory and inspiratory muscle strength, alveolar hypoventilation due to depression of hypoxic and hypercapnic ventilatory drives and decreased in maximal breathing and diffusing capacity in patients with hypothyroidism.Weight gain independently of physical activity is frequently associated with hypothyroidism. Most of the restrictive defects seen among these patients has been thought to be a consequence of obesity and microatelectesis. Overt thyroid dysfunction is well recognized to affect weight, but the influence of minor perturbations of thyroid function remains unclear.

Hypothyroidism can have numerous effects on the respiratory system. Fatigue and dyspnoea on exertion are frequent symptoms. But in the absence of primary respiratory disease, the dimunition of the respiratory function in the hypothyroid patients is not significant in most cases.

Nevertheless, it does affect the respiratory system including respiratory muscle weakness, alveolar hypoventilation due to decreased hypoxic and hypercapnic ventilatory drives, upper airway obstruction, central and obstructive sleep apnoea and even pleural effusion. Lung volumes are usually normal or mildly reduced, but maximal breathing capacity and diffusing capacity are usually reduced. Technologies for generating high-density arrays of cDNAs and

oligonucleotides are developing rapidly, and changing the landscape of biological and biomedical research. They enable, for the first time, a global, simultaneous view on the transcription levels of many thousands of genes, when the cell undergoes specific processes and in certain conditions. For several organisms, the sequences of all genes are available, and thus, transcript levels of the complete gene collection can already be monitored today. The potential of such technologies is tremendous: Monitoring gene expression levels in different developmental stages, tissue types, clinical conditions and different organisms can help understanding gene function and gene networks, assist in the diagnosis of disease conditions and reveal the effects of medical treatments. Undoubtedly, other applications will emerge in coming years.

II. METHODOLOGY

Neural networks are the branch of artificial intelligence. Their models are inspired by the neural systems of human brain. And have been applied in many research fields such as biology, psychology, statistics, mathematics, medical science, computer sciences, and also, a variety of business areas like finance, management and decision making, marketing and production. Here we use DBSCAN and Hierarchical multiple classifier are their discrimination power, discovery the complex and nonlinear relationship among the attributes, and prediction of the cases are the most important advantages. Nevertheless, they can be over-fitted for training data, and time consuming because of computational requirements. The selection of an appropriate training algorithm, transfer functions, initial values of network weights and also, the number of parameters and hidden layers to define the network size determine the performance of DBscan. But compared to logistic regression analysis, neural network models are more flexible.

In this study, we use a type of neural networks namely feed-forwards network with Hierarchical multiple classifier algorithm to model the relations among attributes in our clinical dataset. A feed forward back propagation neural network is a supervised network. That is, it uses training and testing data to build a model. The data involves a set of input attributes with their corresponding output. The network uses the training data to "learn" how to predict the known output, and the testing data is used for validation. The aim is to predict the output for any given inputs in such a way that the distance between the observed and predicted outputs becomes minimized. This algorithm repeatedly examines all the training data to update its weights. These weights are adjusted during training and the process is only in the forward direction through the network without any feedback loops.

A. DBSCAN

The **DBSCAN (Density-based spatial clustering of applications with noise**) algorithm is based on the concepts of density reachability and density-connectivity. These concepts depend on two input parameters: epsilon (eps) and minimum number of points (minPts). Epsilon is the distance around an object that defines its eps-neighborhood. The DBSCAN algorithm works as follows. Initially, all objects in the data set are assumed to be unassigned. DBSCAN then chooses an arbitrary unassigned object p from the data set. If DBSCAN nds (Non-Deliverable Swap) p is a core object, it nds all the density-connected objects based on eps and minPts. It assigns all these objects to a new cluster. If DBSCAN nds p is not a core object, then p is considered to be noise and DBSCAN moves onto the next unassigned object. Once every object is assigned, the algorithm steps.

B. HIERARCHICAL MULTIPLE CLASSIFIER

Hierarchical multiple classifier classification scheme, which preserves the strength of the multiple-classifier approach and also manages to reduce some of the problems faced by other multiple-classifier algorithms. In our scheme, the

system resource requirements are reduced and so is the training time. From the results on pap-smear data, it can be seen that our approach produces better performance than other multiple-classifier algorithms and better than a classifier produced by human experts.

C. COMPARISON WITH EXISTING SYSTEM AND PROPOSED SYSTEM

In existing system people want to know the thyroid level means consider a doctor to find the thyroid level suppose people are unable to consider doctor means the thyroid level is increased and there are no mechanism to people find the thyroid level. Also no effective method is in the current world. To overcome this problem need a effective classification and clustering method.

D. DISADVANTAGES

- Difficult to find the thyroid level
- Large Time Consuming process.

E. PROPOSED SYSTEM

Using **DBSCAN** Algorithm for clustering. The DBSCAN algorithm works as follows. DBSCAN then chooses an arbitrary unassigned object p from the data set. Also classify the dataset using **Hierarchical multiple classifier** classification scheme, which preserves the strength of the multiple-classifier approach and also manages to reduce some of the problems faced by other multiple-classifier algorithms. In this way the data are classified in efficient way provide accurate information.

F. ADVANTAGES

- The prediction results are accurate.
- Short time to find the prediction results.
- Reduce Time and Cost.

III. RESULT

In proposed system user can identify the prediction result based on DBSCAN and Hierarchical multiple classifier used to classify and clustering the data set provide efficient result. If the user symptoms matches the thyroid symptoms means it display the result positive otherwise it shows negative also user can view the thyroid result

.

IV. CONCLUSION

DBSCAN algorithm is used for predicting the thyroid disease with the related symptoms. The DBSCAN algorithm works as follows. Initially, all objects in the data set are assumed to be unassigned. DBSCAN then chooses an arbitrary unassigned object p from the data set. Also classify the dataset using **Hierarchical multiple classifier** classification scheme, which preserves the strength of the multiple-classifier approach and also manages to reduce some of the problems faced by other multiple-classifier algorithms. In this way the data are classified in efficient way provide accurate information. The user can predict and test their health with the symptoms. The user can predict the thyroid disease with related symptoms. Before going to the hospital and check with the doctor. The user can predict the disease with relevant symptoms.

V. FUTURE ENHANCEMENT

This final section of the report outlines some features that could potentially be implemented in future releases. The current set of features implement is a minimum to what a consumer would expect. Every web application has its own merits and demerits. The project has covered almost all the requirements. Further requirements and improvements can easily be done since the coding is mainly structured or modular in nature. Changing the existing modules or adding new modules can append improvements.

Further enhancements can be made to the application, so that the web site functions very attractive and useful manner than the present one. We plan to develop algorithms to improve DBSCAN and Hierarchical multiple classifier support other probabilistic top-k queries in the future.

REFERENCES

1. G.K. Gupta. Introduction to Data Mining with Case Studies, 2006, Prentice Hall India.
2. Papageorgiou E, Kotsioni I & Linos A, (2005) "Data mining: a new technique in medical research", Hormones, Vol. 4, pp 189-191.
3. http://en.wikipedia.org/wiki/Thyroid
4. http://en.wikipedia.org/wiki/Thyroid_cancer
5. 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 42 296– 301.
6. Ginsberg J & Walfish PG. Post-partum transient thyrotoxicosis with painless thyroiditis. Lancet 1977 i 1125–1128.
7. Gerstein HC. How common is postpartum thyroiditis? Archives of Internal Medicine 1990 150 1397–1400.
8. Amino N, Tada H & Hidaka Y. Autoimmune thyroid disease and pregnancy. Journal of Endocrinological Investigation 1996 19 59–70.