Volume 3, No. 9, September 2012
Journal of Global Research in Computer Science
ISSN-2229-371X
RESEARCH PAPER
Available Online at www.jgrcs.info

ANALYSIS ON SUSPICIOUS THYROID RECOGNITION USING ASSOCIATION RULE MINING

K.Saravana Kumar[*1], Dr. R. Manicka Chezian[2]
[*1]Research Scholar NGM College Pollachi
Saravanakumarmca029@gmail.com
[2]Associate Professor NGM college Pollachi

Abstract: Thyroid cancer was the most common type of cancer in the country, overtaking gastric cancer for the first time in last year. This paper proposes to apply the association rule mining for suspected thyroid diseases. We apply the model of deception of set of thyroid dataset then applied apriori algorithm to generate the rules. .The rules generated are used to test the thyroid as deceptive or not. In particular we are interested in detecting thyroid about critical activities. After classification we must be able to differentiate the thyroid giving information about hyperthyroid, hypothyroid (Informative thyroid) and those acting as alerts (warnings) for the future critical activities.

INTRODUCTION

The thyroid is a gland in the neck. It produces thyroid hormones called thyroxine (T4) and triiodothyronine (T3). These hormones are very important and help control the body's metabolism (use of energy). The thyroid gland also produces calcitonin. This helps control the amounts of calcium and phosphate salts in the body. The level of calcitonin is raised when medullary thyroid cancer is present. Compared to breast, lung, prostate and bowel cancers thyroid cancer is much less common. In 2012, there were approximately 56,460 new cases (43,210 women, 13,250 men) diagnosed in the UK. (Compared to breast cancer 229,060: lung cancer 226,160: prostate cancer 241,740) [8] [10]. Data mining is powerful tool that enable to critical investigators who may lake extensive to training as data analyst to explore large database quickly and efficiently. Computers can process thousands of addition, installing and running software often costs less than hiring and training personality. Computers are also less prone to errors than human investigators. So this system is helps and supports the investigator. Thyroid diseases is also growing rapidly , creating needs for automated analysis.so to detect a critical of thyroid diseases should be applied to discover, identify pattern and Make identify pattern. We apply this model of deception to the set of thyroid dataset we used apriori algorithm to generate to generate a classified the categories the thyroid as deceptive or not.

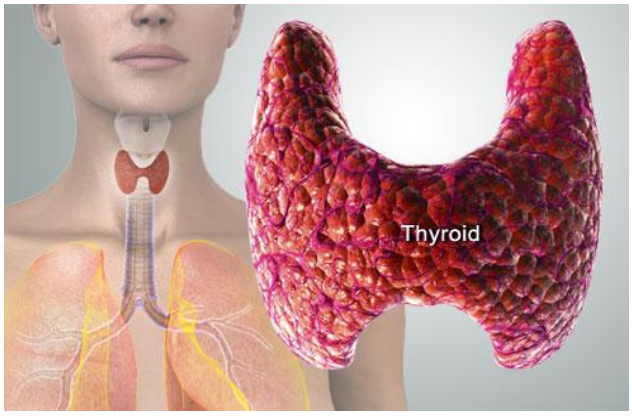


Figure: 1

PROBLEM STATEMENT AND RELATED WORK

Thus the problem is to find a system that identifies the deception in communication through thyroid. Even after classification of deceptive thyroid we must be able to differentiate the informative thyroid from the alert thyroid. We are given the example below h refer to informative thyroid as those giving details about you are affected hyperthyroid or hypothyroid [9].

Example of suspicious thyroid and normal thyroid

Suspicious thyroid	normal
Name : x	Name :y
Sub : test for thyroid	Sub : test for thyroid
Results	Results
Tsh = 0.4	Tsh = 0.4
T3 = 1.9 µg/l	T3 = 1.9 µg/l
T4 = 18 µg/l	T4 = 18 µg/l

Example of suspicious thyroid and informative thyroid

suspicious thyroid	informative thyroid
Name : x	Name :y
Sub : test for thyroid	Sub : test thyroid
Results	Results
Tsh = 0.4	Tsh = 0.4
T3 = 1.9 µg/l	T3 = 1.9 µg/l
T4 = 18 µg/l	T4 = 18 µg/l

Informative thyroid is provides information about he/she affected by hyperthyroid or hypothyroid .Informative thyroid is not a critical position. Suspicious thyroid is provides information about he/she affected by thyroid in critical position.

PROPOSED WORK

In this paper, we present an association rule mining algorithm (Apriori algorithm) to detect suspicious thyroid and the further classification into the alert and informative thyroid [6] [7]. In implementation, there are two parts: thyroid Preprocessing, Building the associative classifier and validation

Thyroid preprocessing:

Based on the theory of deception a deceptive thyroid based on the theory of deception a deceptive thyroid so, such words are set as keywords and extracted from the input dataset. Example for TSH, T3, T4 is the denoting keywords.

Thyroid function tests:

Test Normal ranges

TSH 0.3–3.0 μU/ml
T3 0.8–1.8 μg/l = 80–180 ng/dl

T4 46–120 μg/l = 4.6–12.0 μg/dl

The output after the preprocessing is in the table format in which the attributes are given as the table headers and the records are given in column. The class attribute is to detect either informative, alert or normal thyroid.

Problem statement related work:

Table: 1

T2	T3	T4	Results
High	Low	Low	Informative
Low	High	High	Informative
Normal	Normal	High	Suspicious
Normal	High	Normal	Suspicious
Normal	Normal	Normal	Normal
Normal	Normal	Low	Suspicious
Normal	low	Normal	Suspicious
High	Normal	Normal	Informative
Low	Normal	Normal	Informative

Building the associative classifier:

Thyroid Classification is the process of finding a set of models (or functions) that describes and distinguish data classes and concepts, for the purpose of being able to use the model to predict class of objects whose label is unknown [7].

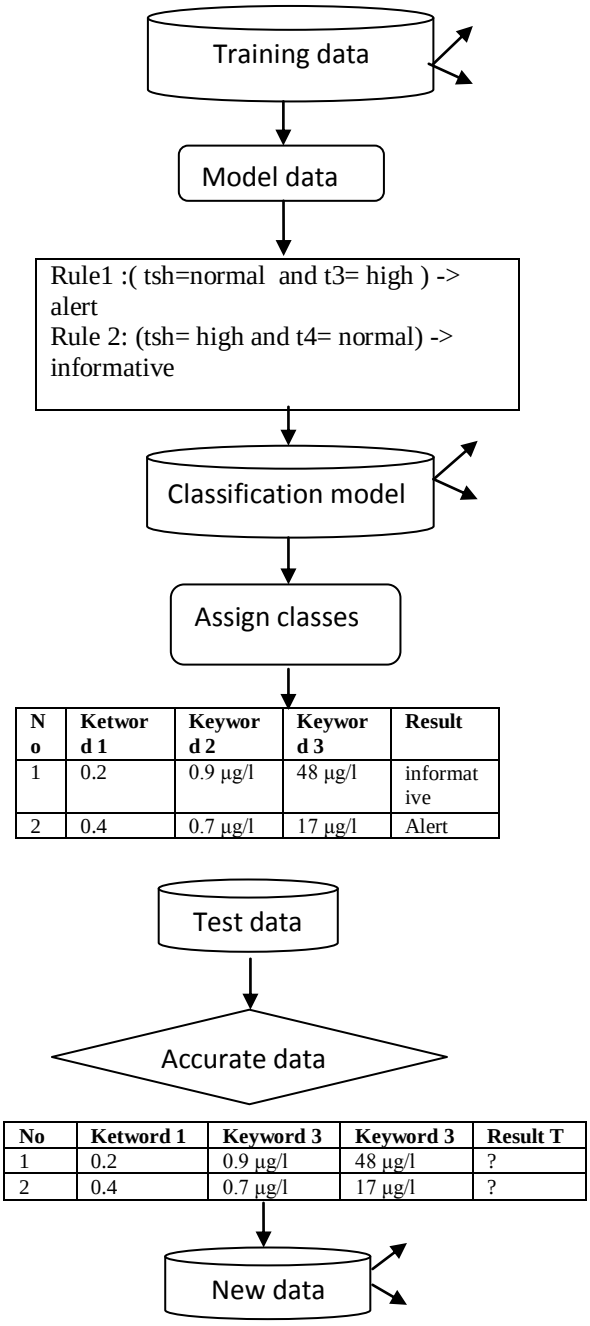
The training data contains two transactions of class Alert thyroid that have keyword TSH/T3/T4 in them, one transaction of TSH=high, T3,T4=normal. Apriori algorithm we obtain a model containing two second step the model just built is tested using test data containing two transactions. If accuracy is measured as a percentage of messages correctly classified, If accuracy is not satisfactory then one or several steps of the classifier need to be modified.

APRIORI ALGORITHM FOR SUSPICIOUS THYROID DETECTION

Association Rule mining searches for interesting association or correlation relationships among items in a given large data set [2]. The Apriori algorithm is used for mining frequent item sets in transactional databases to find frequent sets of words in

Table: 2

No	tsh	T3	T4	Resillt
1	high	normal	normal	informative
2.	normal	High	normal	Alert
3	normal	normal	Low	Alert
4	normal	normal	normal	normal



The Thyroids of the training set [1]. Given the frequent sets of words and topical category assigned to the transaction from which they were extracted association rules are deduced with constraints on the antecedent and consequent of the rules such that the antecedent always contains words while the consequent is exclusively a topical category.

This paper attempt to the algorithm for suspicious thyroid detection as informative and alert. The algorithm produces a list of itemsets. For example Itemsets for the suspected informative Thyroid could include {TSH= high, T3= Normal, T4= Normal} and alert thyroid could include {TSH= normal, T3= normal, T4= high} The support shows how many cases have the Itemset values {Tense =high, T3, T4=high thyroid =Suspicious thyroids} and the confidence shows the likelihood of Thyroid Suspicious or Deceptive for a case having TSH= high and T4= low suppose an thyroid contain the item or keyword. {TSH, T3, T4}. This rule's confidence is the percentage of transactions containing {TSH=high, T3=normal,} that also contain {T4=normal}.

The support for the rule is the number of transactions that contain {TSH=low, T3=normal} and T4=normal}.An association rule can have many items in its antecedent (left hand side) and many items in its consequent (right hand side). The rule {TSH=high, T3=Normal, T4= normal}-> {Result=informative} has antecedent { TSH=high, T3=Normal, T4= normal }and consequent {Class=Informative}.This Item sets are then used to generate Association Rules and one such rule is [3][4]

TSH=High, T4=Normal, T4 =Normal -> informative thyroid
TSH=Normal, T4=Low, T4 =Normal -> alert thyroid
TSH=Normal, T4=Normal, T4 =Normal -> informative .

EXPERIMENTAL RESULTS

The application of data mining to the task of suspect thyroid detection is done; experiments were carried out on a small thyroid corpus. A mixture containing 1000 informative thyroid, 1000 alert thyroid and 1000 normal thyroid. The system was trained with the training dataset and the default support and confidence threshold were used [11].

Tsh = high t3= normal t4=normal {suspicious informative thyroid}
Tsh = normal t3= high t4=normal {suspicious thyroid}
Tsh = normal t3= normal t4= high {suspicious thyroid}

Tsh = normal t3= normal t4=normal { normal }
The frequent itemset {Tsh=normal T3 = High, T4=normal} and the resulting association rule is if Tsh=normal T3 = High, T4=normal then thyroid =suspicious (alert).this is suspicious thyroid alert of alert type that is it will lead to any consequences in future.

CONCLUSION

Association rule mining has a wide range of applicability such as market basket analysis, suspicious e-mail detection, suspicious thyroid detection, library management and many areas. We can find it that a simple apriori algorithm can provide better classification results for suspicious thyroid detection one major advantages of association rule based classifier is that it does not accept that terms are independent and its training is relatively fast. Furthermore, the rules are human reasonable and easy to be maintained or pruned by human being. In this paper, a method of applying Association rule mining for suspected thyroid detection is presented using keyword extraction and considering key attribute called TSH, T3, T4. The proposed work will be helpful for identifying the misleading thyroid and also assist the detectives to get the information in time to take effective actions to reduce the critical actions.

REFERENCES

[1]. R.Agrawal, R. J. Bayardo, and R. Srikant. Athena "Mining-based interactive management of text databases," In Proc. 7th Int. Conf. Extending Database Technology (EDBT'00), pages 365-379, Konstanz, Germany, 2000.

[2]. R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," In Proc. 1993 ACM-SIGMOD Int. Conf: Management ofData, pages 207- 216, Washington, D.C., May 1993.

[3]. R.Agrawal and R.Srikant, "Fast algorithms for mining association rules," In Proc. 20th Int. Conf. Very Large Data Bases (VLDB'94), pages 487-499, Santiago, Chile, 1994.

[4]. G. Boone. "Concept features in re:agent, anintelligent email agent," In Proc. 2nd Int. Conf. Autonomous Agents (Agents'98), pages 141-148, New York, 1998.

[5]. S. Chakrabarti, B. E. Dom, R. Agrawal, and P. Raghavan, "Using taxonomy, discriminants, and signatures for navigating in text databases," In Proc. 23rd Int. Conf. Very Large Data Bases, pages 446-455, Athens, GR, 1997.

[6]. Liu,W. Hsu, and Y. Ma, "Integrating classification and association rule mining," In ACMInt. Conf on Knowledge Discovery and Data Mining (SIGKDD'98), pages 80-86, New York City, NY, August 1998.

[7]. K. Wang, S. Zhou, and S. C. Liew, "Building hierarchical classifiers using class proximity," In Proc. 25th Int. Conf. Very Large Data Bases (VLDB'99), pages 363-374, Edinburgh, UK, 1999

[8]. Sasieni PD, Shelton J, Ormiston-Smith N, Thomson CS, Silcocks PB What is the lifetime risk of developing cancer?: the effect of adjusting for multiple primaries. Br J Cancer, 2011. 105(3): p. 460-5

[9]. Takats IK, Peter F, Rimanoczi E, et al. The blood spot thyrotropin method is not adequate to screen for hypothyroidism in the elderly living in abundant-iodine intake areas: comparison to sensitive thyrotropin measurements. Thyroid 2000;10:79–85.

[10]. Gilbert ES, Tarone R, Bouville A, Ron E. Thyroid cancer rates and 131I doses from Nevada atmospheric nuclear bomb tests. J Nat Cancer Inst 90:1654 –1660; 1998.

[11]. Kerber RA, Till JE, Simon SL, Lyon LL, Thomas DC, Preston-Martin S, Ralison ML, Lloyd RD, Stevens W. A cohort study of thyroid disease in relation to fallout from nuclear weapons testing. JAMA 270:2076 –2082; 1993.

Short Bio Data for the Author

K.Saravana Kumar received his B.Sc Physics from Chikkanna Government Arts and Science College, Tirupur, India. He completed his Master of Computer Applications in Gopi Arts and Science College, Gopi, India. Presently he is pursuing his Ph.D Computer Science in the Research Department of Computer Science, NGM College, Pollachi. He has published papers in International journal and National Conferences.

Dr.R.Manickachezian received his M.Sc Applied Science from PSG College of Technology, Coimbatore, India in 1987. He completed his M.S. degree in Software Systems from Birla Institute of Technology and Science, Pilani, Rajasthan, India and Ph.D degree in Computer Science from School of Computer Science and Engineering, Bharathiar University, Coimbatore. He has 25 years of Teaching experience and 17 years of Research Experience. He served as a Faculty of Maths and Computer Applications at P.S.G College of Technology, Coimbatore from 1987 to 1989. Presently, he is working as an Associate Professor of Computer Science in NGM College (Autonomous), Pollachi, India. He has published 55 papers in various International Journals and Conferences. He is a recipient of many awards like Desha Mithra Award and Best paper Award . He is a member of various Professional Bodies like Computer Society of India and Indian Science Congress Association. His research focuses on Network Databases, Data Mining, Distributed Computing, Data Compression, Mobile Computing and Real Time Systems