

Automated Prediction of Thyroid Disease using ANN

Vikram V Hegde¹, Deepamala N²P.G. Student, Department of Computer Science and Engineering, RV College of Engineering and Technology,
Bangalore, Karnataka, India¹Assistant Professor, Department of Computer Science and Engineering, RV College of Engineering and Technology,
Bangalore, Karnataka, India²

ABSTRACT: All the living creatures are subjected to health disorder in one or the other ways. Thyroid disease is a major cause of concern in medical diagnosis and the prediction, onset of which is a difficult proposition in medical research. It is possible to have accurate prediction of disease by understanding the symptoms and the pathological reports, by looking at the generated report patterns and models. With the help of machine learning algorithms, computerised prediction system can be developed using artificial neural networks (ANN). We collected the thyroid samples from Lakshmi laboratory, chikmagalore-KA, after cleaning and pre-processing, classified them to hypo, hyper-thyroid, normal. The dataset is divided in to 80:20 ratio for training and testing, applied to ANN. The set was implemented in ANN toolbox of matlab, which resulted average accuracy of 85% during model training phase. The model was tested with different classes of inputs, resulting 82% accuracy.

KEYWORDS: Neural networks, Thyroid Disease, Medical Diagnosis, matlab ANN toolbox.

I. INTRODUCTION

The thyroid is a gland located at the lower front of the neck that produces the hormones that help regulate many body processes, including growth, energy balance, body temperature, and heart rate [1]. Thyroid function involves the interaction of many hormones, including triiodothyronine (T3) and thyroxine (T4). Both hormones exist in two forms in the blood and this research can help diagnose hyperthyroidism (when the thyroid gland produces too much thyroid hormone) and hypothyroidism [2]. A thyroid stimulating hormone (TSH), produced by the pituitary, is common blood test to check the functionality of thyroid gland. Production of too little hormones causes hypothyroidism, on the other hand more hormone production causing hyperthyroidism.

Machine learning is a subfield of computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence. Machine learning explores the study and construction of algorithms that can learn from and make predictions on data. Machine learning systems automatically learn programs from data. This is often a very attractive alternative to manually constructing them.

In machine learning science, artificial neural networks (ANNs) are a family of models inspired by biological neural networks (the central nervous systems of animals, in particular the brain) and are used to estimate or approximate functions that can depend on a large number of inputs and are generally unknown.

Artificial neural networks are generally presented as systems of interconnected "neurons" which exchange messages between each other. The connections have numeric weights that can be tuned based on experience, making neural nets adaptive to inputs and capable of learning. Artificial Neural Network is mathematical model or computational model that is inspired by the structure and functional aspects of biological neural networks.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Modern neural networks are non-linear statistical data modelling tools. They are usually used to model complex relationships between inputs and outputs or to find patterns in data. The architecture of neuron is shown in below figure.

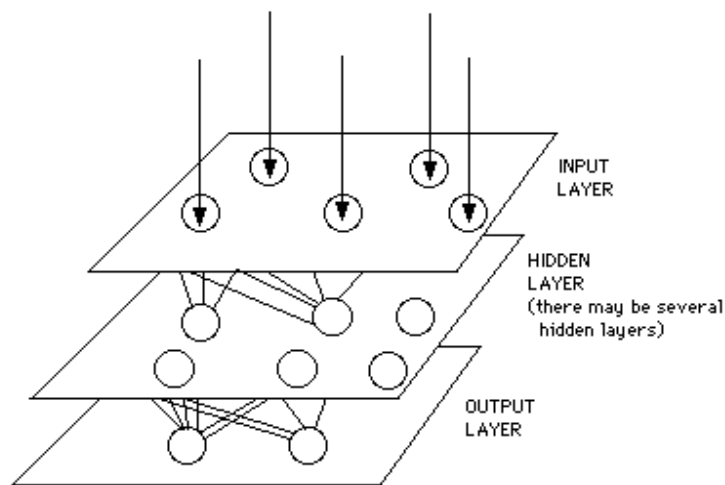


Fig 1: The architecture of Neuron

Artificial neural networks provide a powerful tool to help doctors to analyse, model and make sense of complex clinical data across a broad range of medical applications. Perhaps the greatest advantage of ANNs is their ability to be used as an arbitrary function approximation mechanism that 'learns' from observed data. With the correct implementation, ANNs can be used naturally in online learning and large data set applications. Their simple implementation and the existence of mostly local dependencies exhibited in the structure allows for fast, parallel implementations in hardware.

II. RELATED WORK

Many theoretical works have been proposed for thyroid disease diagnosis. S.B.Patel worked to predict the diagnosis of heart disease using classification mining techniques [3]. Jacquelin Margret et al proposed a decision tree splitting rule for this disease diagnosis [4]. FarhadSoleimanianGharehchopogh et al performed a case study in diagnosis of thyroid disease using artificial neural network [4]. Savelonas et al in 2005 proposed variable background active contour model for detection of thyroid nodules in USG images [5, 6].

Chang et al adopted seven feature extraction technique viz., co-occurrence matrix, grey level run-length matrix, laws textures energy measures, wavelet and Fourier features based on local Fourier coefficients[7]. Senol et al proposed a hybrid structure in which Neural Networks and Fuzzy logic are combined to diagnose the thyroid disease [8]. Rouhani et al compared several ANN models for diagnosing the thyroid disease [9].Isa et al has experimented for several activation functions such as sigmoid, Hyperbolic tangent, Neuronal, Logarithmic and Sine activation function for the MLP Neural Network and determined the most suitable function to classify the thyroid disease as Hypothyroid and Hyperthyroid[10].

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

III. METHODOLOGY

Previously collected data consisting of more than 400 human subjects (Hypo Thyroid, Hyper Thyroid, normal) is used to develop machine learning algorithms to predict the diseases. elabport is used to collect the data and the disease is marked (normal, HIV, Dengue, Thyroid) with the help of an expert clinician. Artificial Neural Networks are used to apply the machine learning.

Data set for the study is based on the self-developed centralized clinical product –elabport, which is been used in various parts of Karnataka state since year 2013. It is the medical reporting software which is used to save and maintain the pathological reports online. The type of disease based on its parameters is classified in table.

Type	T3	T4	TSH
Hypo	Increase	Increase	Reduction
Hyper	Reduction	Reduction	Increase

Table 1: Classification of thyroid disease

The data set includes various pathological reports and the respective patient information like name, age, gender etc. The data is collected from elabport with the help of an expert clinician and stored in excel like spread sheet.

The considerations made are tri_iodo, thyroxine and thyrotropin tests which are normally done for the thyroid investigation. The considered normal ranges are (80-180 ng/dl), (0.5-6 uU/ml) and (4.6-12 ug/dl) respectively. For the implementation of the datamining, the dataset is collected mainly from Sri Lakshmi Clinical laboratory, located in birur of Chikmagaloredst, Karnataka. The centre is serving the patients from past two decades nearly and it is using elabport software from two years.

Since the units of the testing are different, it was a challenge to bring all the test values to a single unit form, form applying to machine learning. So all outliers got taken off and the missing data and noisy data like rows with texts written in them are removed and hence data cleaning is done. The dataset which are generated after the data pre-processing will then have all same units, mean to say are (ng/dl), (uU/ml) and (ug/dl) for are tri_iodo, thyroxine and thyrotropin respectively.

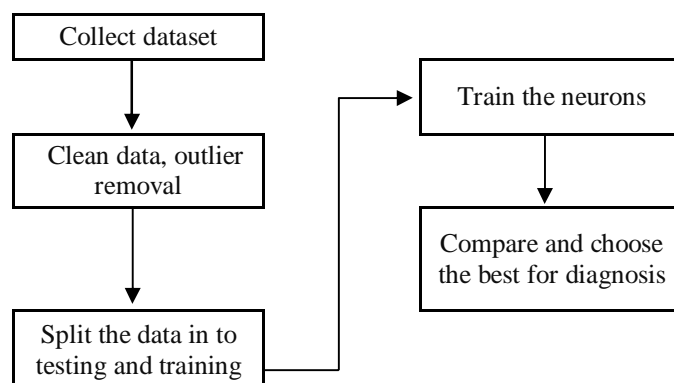


Fig 2: Methodology in Disease prediction

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

After the generalization, dataset has been validated and the result of each individual row is analyzed. To make the dataset suitable for the machine learning, one column named result has been added to dataset. The result will be named 0 or 1 depending on the test result. 0 is given formal normal observation and 1 for abnormal, either low or high compared to the normal ranges. The number of laboratory samples are determined based on the type of disease.

Type	samples	Class
Normal	130	1
Hyper	100	2
Hypo	120	3
Total	350	

Table 2: Laboratory samples

Artificial neural network (ANN) is used to develop the machine learning model for the disease prediction. The data which is randomised and the data set is divided into 80:20 ratio, where 80% is used for training the model while remaining 20% is used for testing. The ANN model was created and is implemented in the ANN tool of matlab.

IV. RESULT

With the training: testing proportion, the dataset is processed with the ANN model and various types of inputs in randomized order were given to the testing model along with three classes of normal, hypo and hyper valued data and the variation in analysed to get a better accuracy. The training model with all kind of varied random input produced the average of 85% accuracy. Then the testing part of the data is subjected to test on the basis of trained samples, which provided accuracy of 82% on an average.

The results of the analysis and the accuracy is presented in terms of confusion metrics. Results predicted is compared with the past data available to make sure that the results obtained are accurate.

V. CONCLUSION

It is seen that neural network can be used successfully for the diagnosis of thyroid disease. For both clinical and statistical classifications, thyroid disease detection is an important yet difficult task. Due to extremely unbalanced groups in the thyroid diagnosis and a large number of interrelated data attributes and the patient true group membership causes the poor performance of the traditional model based statistical methods. Artificial neural network is a flexible modelling technique for complex function mapping, show promise in the thyroid disease diagnosis.

Diagnosis of thyroid using artificial neural network is robust with respect to sampling variations. The cross validation technique provides decision makers with a method for examine predictive validity and hence the usefulness of the classification method. A useful model is the one which is robust across different samples or time periods. Neural networks are both robust and accurate for the task of diagnosing thyroid dysfunction. For medical diagnosis, employing a classifier with high robustness and reliability in different sampling situations is a very critical issue.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Special Issue 10, May 2016

REFERENCES

- [1] Rastogi, AsthaBhalla, Monika, A Study of Neural Network in Diagnosis of Thyroid,in International Journal of Computer Technology and Electronics Engineering (IJCTEE) Vol 4, Issue 3, June 2014.
- [2] Zhang, Guoqiang Peter Berardi, Victor L, An investigation of neural networks in thyroid function diagnosis In Health Care Management Science,vol.1,issue1,pages29-37
- [3] S.B Patel, P. K Yadav, Dr. D. P.Shukla,"Predict the Diagnosis of Heart Disease Patients Using Classification Mining Techniques", (IOSR-JAVS), e-ISSN: 2319-2380, p-ISSN: 2319-2372. Volume 4, Issue 2 (Jul. - Aug. 2013), Pg.no 61-64.
- [4] J.Margret ,B. Lakshmi pathi and S.A.Kumar , "Diagnosis of Thyroid Disorders using Decision Tree Splitting Rules",International Journal of Computer Applications (0975 – 8887) Volume 44– No.8, April 2012
- [5] D.E. Maroulis, M.A. Savelonas and S.A. Karkanis, "Computer –Aided Thyroid Nodule Detection in Ultrasound Images" 18th IEEE, International Symposium on computer based medical systems (CBMS'05), 2005.
- [6] M. Savelonas, D. Maroulis, D. Iakovidis and S. Karkanis, "Variable Background Active Contour Model for Automatic Detection of Thyroid Nodules in Ultrasound Images ", 2005.
- [7] Chuan-Yu Chang, Ming-Fang Tsai and Shao-Jer Chen," Classification of the Thyroid Nodules Using Support Vector Machines" International joint conference on Neural Networks 2008, pp 3093- 3098.
- [8] Canon SENOL,Tulay YILDIRIM, "Thyroid and Breast Cancer Disease Diagnosis using Fuzzy-Neural Network ", unpublished
- [9] ModjtabaRouhani,KamranMansouri, "Comparison of several ANN architectures on the thyroid diseases grades diagnosis " , 2009 International Association of computer science and information technology-spring conference, 2009, pp 526-528.
- [10] I.S.Isa,Z.Saad,S.Omar,M.K.Osman,K.A.Ahma , "Suitable MLP Network Activation Functions for Breast Cancer and Thyroid Disease Detection" ,Second international conference on computational intelligence, modelling and simulation,2010 IEEE, pp 39–44.