JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

THYROID DISEASE CLASSIFICATION USINGDEEP LEARNING

[1]P.Snehalatha, [1]P.Sreeja, [1]P.Thanmayi, [1]P.Aravind Swami,
[2]J.Sofia, [3]Dr. Thayabba Khatoon
1 UG Student, 2 Assistant Professors, 3 Professors, and HOD

Department of Artificial Intelligence and Machine Learning, School of Engineering, Malla Reddy University, Maisammaguda, Dulapally, Hyderabad, Telangana, 500100

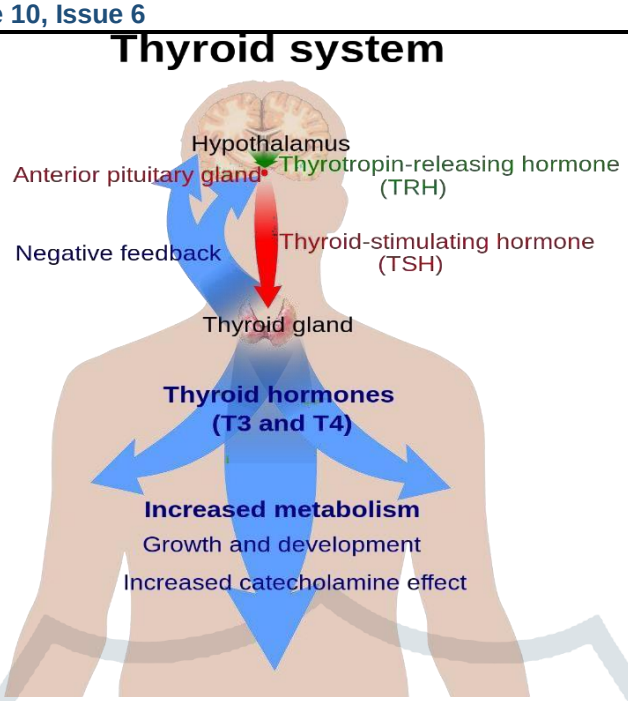
***Abstract*- Thyroid disease is a major cause of formation in medical diagnosis and in the prediction The secretions of thyroid hormones are culpable in controlling the metabolism. Hyperthyroidism and hypothyroidism are the two common diseases of the thyroid that releases thyroid hormones in regulating the rate of the body's metabolism. Data cleansing techniques were applied to make the data primitive enough for performing analytics to show the risk of patients obtaining thyroid. Machine learning plays a decisive role in the process of disease prediction and this paper handles the analysis and classification models that are being used in thyroid diseasebased on the information gathered from the dataset taken from the UCI machine learning repository. In this, we also proposed different machine-learning techniques and diagnoses for the prevention of the thyroid. Machine Learning Algorithms, K-NN, Decision Trees, and multi-layer perceptron were used to predict the estimated risk of a patient's chance of Obtaining thyroid disease.**

***Keywords*-** Deep learning, multilayer perceptron, thyroid disease

I. INTRODUCTION

Thyroid disease is a common cause of medical diagnosis and prediction, with an onset that is difficult to forecast in medical research. The thyroid gland is one of our body's most vital organs. Thyroid hormone releases are responsible for metabolic regulation. Hyperthyroidism and hypothyroidism are the two common diseases of the thyroid that releases thyroid hormones in regulating the rate of the body's metabolism. The main objective is to develop a system that can predict the type of thyroid disease that the patient is affected by. To predict thyroid disease with the usage of a minimum number of parameters. To predict all possible types of Thyroid diseases. As well as providing the ability to detect the disease before it forms into a more destructive anomaly. In the result, the patient will be classified to have either of the following Hyperthyroid, Hypothyroid, Sick, or Negative.

This is one of the common lifestyle diseases. The thyroid organ is a butterfly-molded organ that is present in the neck and underneath the mouth of the human body. It releases hormones that control metabolism like heart rate, body temperature, etc. It produces two main hormones T3 and T4. These hormones are responsible for various metabolic activities like body weight, heart rate, etc. These activities may get disturbed if the level of these hormones changes. So, the diagnosis of thyroid disease is important before its treatment. About 32 percent of the total Indian population suffers from thyroid disease. When the number of hormones exceeds the amount required by the human body, it causes hyperthyroidism. Hypothyroidism is the inverse of hyperthyroidism; it reduces body metabolism, causes drowsiness and pain in joint metabolism, and causes drowsiness and pain in joints.



II. RESEARCH AND RELATED WORK

Various researchers have used different pattern classifiers for developing lifestyle disease prediction systems. In this section, a brief study of the thyroid disease prediction system has been presented. I. D. Maysanjaya et.al. (2015) have used six different methods for the diagnosis of thyroid disease. After experimenting, the performance of multilayer perception was found highest as compared to that of the Indonesian J Elec Eng& Comp Sci, Vol. 17, No. 1, January 2020: 524 - 533 526 other five methods. Mohd. Reza et.al. (2017) have discussed the diagnosis of different types of thyroid disease using ANN by considering the age of an individual. The input to the thyroid prediction system is seven hormone tests including age and the output is the diagnosis of the thyroid. The various ANN structures used include MLP, PNN, GRNN, AND CFNN. Shivanee (2016) has proposed various data mining techniques like Bayes net, multilayer perception, RBF network, L4.5, CART, REP tree, and decision stump to develop classifiers for the diagnosis of hypothyroid disease. After performing the experiments, it is clear that tree and L4.5 techniques perform well as compared to others. Mazin Abdul Rasool Hameed et.al (2009) have proposed a method of classifying thyroid disease using multilayer feed-forward using the backpropagation learning rule. In this work, three inputs have been considered T3, T4, and TSH. Saeed Shariati and Mahdi Motanali Haghighi (2010) have used a fuzzy system to diagnose hepatitis and thyroid disease. The results of fuzzy neural networks with support vector machines and artificial neural networks were compared. Anupam Shukla et.al. (2009) in their work have trained the system using three ANN algorithms, backpropagation (BPA), the radial basis function (RBF), and the learning vector quantization (LVQ).

The objective is to develop a system that can predict the type of thyroid disease that the patient is affected by. To predict thyroid disease with the usage of the minimum number of parameters. To predict all possible types of Thyroid diseases. As well as providing the ability to detect the disease before it forms into a more destructive anomaly. In the result, the patient will be classified to have either of the following Hyperthyroid, Hypothyroid, Sick, or Negative.

Very well Health is used to clarify your diagnosis or track the progression of treatment. This site displays the analysis of the disease, but the disadvantage of these sites they don't predict the chances of getting thyroid disease. The main advantage of our model is we predict the type of thyroid disease they have.

III. PROPOSED SYSTEM

A multilayer perceptron is a feedforward artificial neural network that generates a series of outputs from a set of inputs (MLP). In an MLP, several layers of input nodes form a directed graph between the input and output layers. MLP employs backpropagation to train the network. MLP is a deep learning technique. A multilayer perceptron is a neural network that connects several layers 11 in a guided graph, meaning that the signal only travels in one direction between nodes. Each node, except for the input nodes, has a nonlinear activation function. An MLP uses backpropagation as a supervised learning process. MLP is a deep learning method that employs several layers of neurons. MLP is a commonly used system in supervised learning problems, computational biology, and parallel distributed processing analysis. Applications include speech recognition, image recognition, and automatic translation

First, we acquired the thyroid disease dataset from Kaggle (a famous data repository). The dataset consists of several thyroid-related disease records and many target classes. The samples for target classes are few, which is not enough to train models, so we select only those target classes whose samples are more than 250, as a result, after selecting the target classes for experiments, we performed the data balancing. It is followed by the feature selection process, which applies many feature selection techniques. Experiments are performed with an 80-20 train-test split using machine learning algorithms like decision tree, random forest, and k-nearest neighbors multi-layer perceptron, and finally, we take the result of maximum accuracy of prediction model posed Framework

The training and testing phases of the thyroid disease prediction system are clearly shown in Figure 3. As shown in Figure 3, the first step is to identify the typical parameters/risk factors which are responsible for thyroid disease in human beings. In the next step, a miscellaneous dataset of various patients of different categories is collected. In the conventional methods of thyroid diagnosis system, the majority of authors have used only three factors namely T3, T4

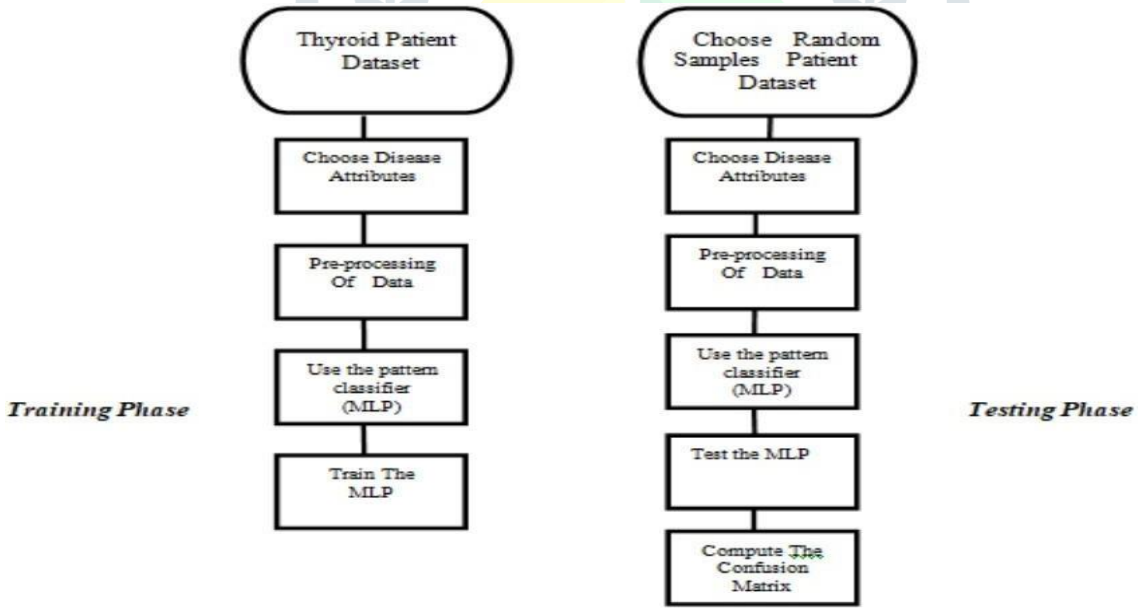


Figure 3. Framework for the proposed system

To classify a particular patient into any of the three classes a dataset of 120 samples has been obtained and preprocessed. To remove anomalies, and noise and to quantify Boolean values the data set is manually enriched. Once the dataset is prepared, a multilayer pattern classifier model is created

and trained with the dataset. The MLP pattern classifier model is stored for the testing phase. To check the accuracy of thyroid predictions a sample of randomly chosen patients is applied to the stored MLP prediction system.

MLP is one of the most common ANN which is widely used for different tasks like pattern classification, pattern recognition, etc. One of the most important features of MLP is that we can specify any number of output classes. The network architecture chosen for this problem is MLP having eleven input nodes and three output nodes. Each node present in the input layer is connected to every other node in the hidden layer through some weights. The value of the weighted input sum to a particular nodemay be large, therefore it is important to scale down the weighted sum by reducing it before producing the resulting output of that particular node. For this purpose, a function is applied to the weighted input. One of the best methods is backpropagation learning which works on the principle of gradient descent rule. The steps of the training and machine learning building algorithm are explained in the following section. The multilayer perceptron is trained with 11 nodes in the input layer of the network.

IV. RESULT AND ANALYSIS

The experiments were conducted on the real dataset of 120 instances collected from SKIMS Soura, Srinagar, India. The subjects were chosen carefully covering a wide range of populations including men, women, old, and youngsters. The values for eleven attributes were collected for all the 120 instances and some attributes have been quantified and factorized. The dataset's preprocessing has been done to remove ambiguities, anomalies, and errors. The pre-processed dataset is used to train the pattern classifier model using the back error propagation algorithm for the multi-layer perceptron. The data set has been split into train and test data instances. The machine learning model has been trained by varying the size of the training dataset and then tested on test data set to achieve the cross-validation

Figure 4 shows the best performance of MLP at epochs 9 on 30 instances with 7 attributes. In Figure 5, the gradient error graph has been shown on 30 instances and 7 attributes. The green cells in the confusion matrix represent correctly classified instances while as the red cells represent incorrect classification. The blue box represents the percentage of both correct as well as incorrect classificationclasses.

NO	Algorithms	Accuracy
1	Decision Tree	90.13
2	SVM	92.53
3	Random Forest	91.2
4	Naive Bayes	90.67
5	Logistic Regression	91.73
6	Linear Discriminant Analysis	83.2
7	KNeighbors Classifier	91.47
8	MLP	96.4

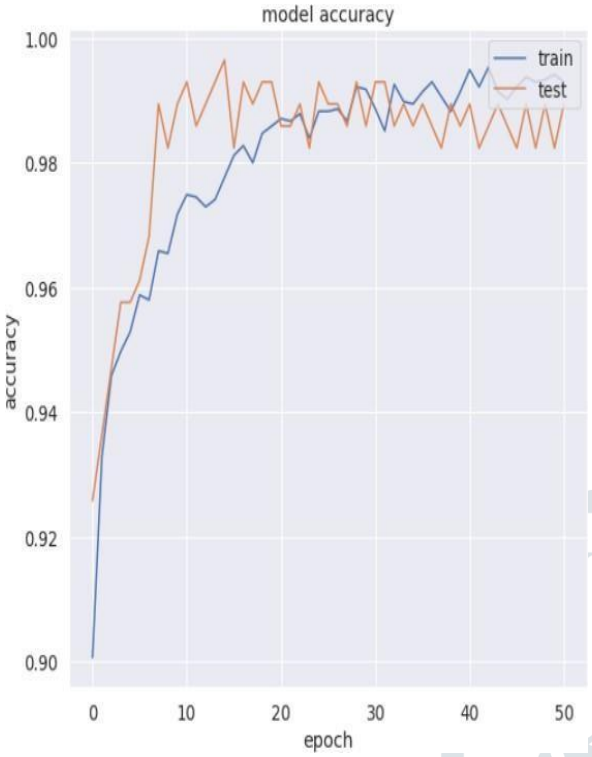


Fig 4

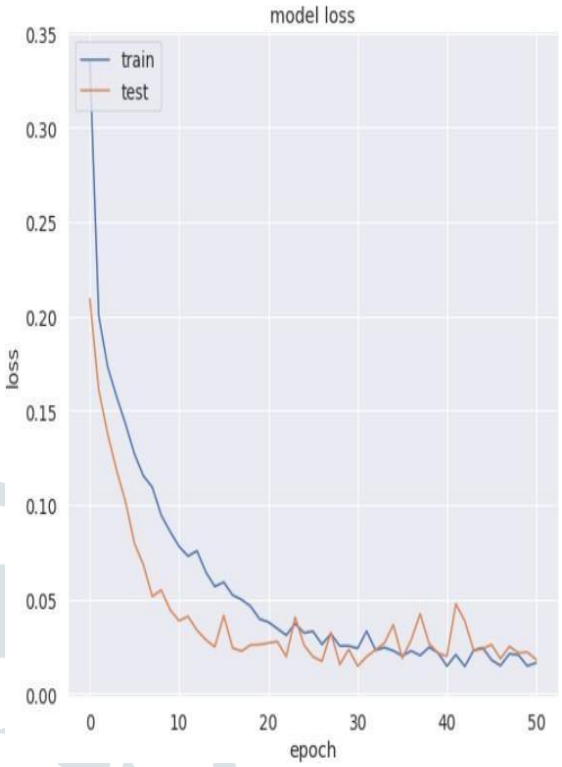


Fig 5

Table 2. Performance of the Intelligent Thyroid Prediction System with 11features (training, validation, testing, accuracy)

Iteration No.	Data Size	Performance	Training Accuracy	Validation Accuracy	Testing Accuracy	Overall Accuracy
1	30	0.031	85%	20%	40%	66.7%
2	60	9.87e-05	~100%	77%	~100%	96.7%
3	90	0.000158	~100%	~100%	~100%	~ 100%
4	120	3.10e-07	~100%	~100%	~100%	~100%

Table 3. Performance of theThyroid Prediction System with 7 Features (Accuracy)

Iteration No.	Data Size	Performance	Training Accuracy	Validation Accuracy	Testing Accuracy	Overall Accuracy
1	30	0.00209	75%	100%	20%	70%
2	60	0.0124	95.2%	100%	77.8%	93.3%s
3	90	0.000663	100%	100%	92.9%	98.9%
4	120	0.00852	~100%	94.4%	~100%	99.2%

V. CONCLUSION

The intelligent prediction and classification have been achieved by training the MLP model with a mixed dataset of various subjects collected from individuals living in different habitats. The proposed model was successfully tested with random samples including instances of hyperthyroid, hypothyroid, and normal individuals. The system exhibits excellent training and testing accuracy of almost 100% with 11 attributes and 99.2% with 7 attributes. The proposed thyroid prediction system exhibits better prediction

accuracy of Iteration No. Data Size Performance Training Accuracy Validation Accuracy Testing Accuracy Overall Accuracy is nearly 99.8% with 7 or 11 features compared to similar systems that used only 3 to 5 features. Moreover, a comparison has been carried out between the two experiments conducted with a different number of decision attributes. In future research, the ensembles using Random forests, Bagging and Boosting machine learning advanced techniques may be used to improve the accuracy. Furthermore, the proposed machine learning model may be extended to diagnose other types of lifestyle diseases like diabetes, blood pressure, and many more. Deep learning techniques like CNN may be used to further incorporate the intelligence in the proposed system.

VI. REFERENCES

[1] I. D. Maysanjaya, H. A. Nugroho, N. A. Setiawan, J. G. No, and K. Ugm, "A Comparison of Classification Methods on Diagnosis of Thyroid Diseases," 2015 Int. Semin. Intell. Technol. Its Appl., pp. 89–94, 2015.

[2] M. R. Obeid Avi, A. L. I. Rafiee, and O. Mahdiyar, "Diagnosing Thyroid Disease by Neural Networks," vol. 10, no. 2, pp. 509–524, 2017.

[3] P. Durga, V. S. Jeba Kumari, and D. Shanthi, "Diagnosis and Classification of Parkinson's Disease Using Data Mining Techniques," ISSN Online) Int. J. Adv. Res. Trends Eng. Technol., vol. 3, no. 14, pp. 2394–3777, 2016.

[4] A. Fabrics, "Artificial Neural Network System for," no. 2, pp. 518–528, 2009.

[5] S. Shariati and M. M. Haghighi, "Comparison of anfis neural network with several other anns and support vector machine for diagnosing hepatitis and thyroid diseases," 2010 Int. Conf. Comput. Inf. Syst. Ind. Manag. Appl. CISIM 2010, pp. 596–599, 2010.

[6] R. Tiwari, A. Shukla, and P. Kaur, "Diagnosis of Thyroid Decimal Neural Networks," no. March, pp. 6–7, 2009.

[7] N. Kumar and S. Khatri, "Implementing WEKA for medical data classification and early disease prediction," 2017 3rd Int. Conf. Comput. Intell. Commun. Technol., pp. 1–6, 2017.

[8] Y. Xing, J. Wang, Z. Zhao, and Yongcheng Gao, "Combination Data Mining Methods with New Medical Data to Predicting Outcome of Coronary Heart Disease," 2007 Int. Conf. Converg. Inf. Technol. (ICCIT 2007), pp. 868–872, 2007.

[9] T. Hsiang, C. P. Wei, and Vc. S. Tseng, "Feature selection for medical data mining: Comparisons of expert judgment and automatic approaches," Proc. - IEEE Symp. Comput. Med. Syst., vol. 2006, pp. 165–170, 2006.

[10] R. Dey and V. Bajpai, "Application of Artificial Neural Network (ANN) technique for Diagnosing Diabetes Mellitus," IEEE explorer, no. 155, pp. 8–11, 2008.

[11] Canan Senol and Tülay Yildirim, "Thyroid and breast cancer disease diagnosis using Fuzzy-neural networks," ELECO 2009 - 6th Int. Conf. Electr. Electron. Eng., pp. 390–393, 2009.

[12] S. Deshmukh and S. Shinde, "Diagnosis of Lung Cancer using Pruned Fuzzy Min-Max Neural Network," Int. Conf. Autom. Control Dyn. Optim. Tech. ICACDOT 2016, pp. 398–402, 2017.

[13] K. Viswanath and R. Gunasundari, "Design and analysis performance of kidney stone detection from ultrasound image by level set segmentation and ANN classification," Proc. 2014 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2014, pp. 407–414, 2014.

[14] S. Muthuselvan, K. S. Sundaram, and Prabasheela, "Prediction of breast cancer using classification rule mining techniques in blood test datasets," 2016 Int. Conf. Inf. Commun. Embed. Syst., no. Icices, pp. 1–5, 2016.

[15] V. J. Madhuri, M. R. Mohan, and R. Kaavya, "Stress Management Using Artificial Intelligence," 2013 Third Int. Conf. Adv. Comput. Commun., pp. 54–57, 2013.