Decision Support for Diagnosing Thyroid Diseases using Machine Learning

W.K.D. Jayamini[*], K.G.H.D. Weerasinghe[**]

[*] Department of Statistics and Computer Science, Faculty of Science, University of Kelaniya, Sri Lanka
[**] Department of Computer Systems Engineering, Faculty of Computing and Technology, University of Kelaniya, Sri Lanka

Abstract- This research was carried out to compare the applicability of machine learning techniques in a way to support decision making in Thyroid disease diagnosis. In this research work, two types of machine learning algorithms; Multilayer perceptron Neural Network with Back-propagation and Random Forest were used, and a comparison of their performance was done to identify the most suitable machine learning algorithm which can be used to assist the diagnostic process of Thyroid diseases. Each of the models developed under the two machine learning techniques were trained and tested using two thyroid datasets from the UCI machine learning repository. The performance of the models was evaluated using the performance measures, Accuracy (%), Mean Absolute Error, Root Mean Squared Error, True Positive rate, False Positive rate, Precision and Recall.

Index Terms- Decision Support, Neural Network, Random Forest, Thyroid Disease Diagnosis

I. INTRODUCTION

In the medication process, one of the major components is medical diagnosis. Disease diagnosis is a very complex it requires more knowledge and expertise. The main objective in healthcare is to identify a disease at its early stage with higher accuracy [1]. Among many diseases, thyroid disease diagnosis is very critical because regulating the body functions and controlling the metabolism of the body are done by thyroid hormones which are secreted from thyroid gland.

Therefore, any dysfunctionality in the thyroid gland will cause adverse effects to the patient. The main thyroid hormones produced by the thyroid gland are Triiodothyronine (T3) and Thyroxin (T4). Too little thyroid hormone production (hypothyroidism) and too much thyroid hormone production (hyperthyroidism) are the general causes for thyroid malfunctioning. Therefore, it is very important to diagnose these diseases correctly and accurately [2].

Decisions on thyroid disease diagnosis are made after investigating laboratory tests which include the amount of thyroxin hormone, triiodothyronine hormone and thyroid stimulating hormone (TSH). In addition to that, physicians take thyroid scan of the patients or results of thyroid uptake test which measures the gland iodine collection ability. Hence, it will consume more time for diagnosing the thyroid function abnormalities. Therefore, to save the precious time of the patient and even the physician, a computer aided diagnosis system will be more useful which can provide a guidance to the physicians in the diagnosis process [3].

One reason for medication errors is inadequate decisions in primary care and hospital settings. For improving the effectiveness and efficiency on these decisions, computer-based decision support can be defined as an assistance for the health professionals in making clinical decisions. In developing computer-based decision support systems, many techniques can be used. Machine learning which comes under artificial intelligence can be used as one approach to develop medical decision support systems.

This research work has focused on diagnosis of thyroid diseases using machine learning techniques; Multilayer perceptron Neural Networks (MLP) and Decision Trees. Neural networks with back-propagation algorithm and an ensemble classifier called random forest which consists of many decision trees were utilized in this work. After developing the models, those were trained with two different thyroid data sets and tested by applying 10-fold cross validation. Then the models were evaluated using the performance measures; Accuracy (%), Mean Absolute Error, Root Mean Squared Error, True Positive rate, False Positive rate, Precision and Recall. At the end, each of these models were compared using the performance evaluation measures.

II. LITERATURE REVIEW

Machine learning is a branch of Artificial Intelligence. It focuses on algorithms and methods which provide a possible way for a computer to learn via existing examples by capturing characteristics of interest based on underlying probability distribution [4]. Machine Learning aims at providing computational methods for accumulating, changing, and updating knowledge in intelligent systems, and in particular learning mechanisms that will help to generate knowledge from examples or data. In cases where there is absence of algorithmic solutions, lack of formal models or the poor definition of knowledge about the application domain, machine learning plays an important role [5]. The interest of developing automated assistance for disease diagnosis and decision making exists in both theoretical and practical contexts.

Following the improvement of classification algorithms with the high performing computational units, it has been started to use machine learning algorithms in diagnosing various kinds of diseases with massive datasets available today. Among many classification problems, one important application is diagnosis

of thyroid gland disorders [6]. However, according to Hannan et al., [7] heart disease diagnosis is more complex than others which needs high level of proficiency. They have developed an expert system for heart disease diagnosis using support vector machine and feed-forward back-propagation technique. At present, there are huge number of successful applications which use neural networks. James E. Avenir et al. [4] explored the ways to assist health workers in diagnosing and prescribing treatment to patients with tuberculosis using machine learning techniques. They used the decision tree algorithm called ID3 algorithm to build the system.

Diagnosis of thyroid gland disorders by assessing the data of thyroid in clinical applications comes out as an important classification problem [8]. For these kinds of classification problems researchers have conducted many studies by constructing different types of models. Among them, multi-layer or probabilistic artificial neural networks, hybrid systems, support vector machines, genetic algorithm, meta-heuristic methods are some of the common models. Razia and Narasinga Rao [9] have explored the neural network models which describe the aspects related to thyroid gland functionality, autoimmune condition and various conditions of thyroid disease. Makas and Yumusak [3] have designed multi-layer neural network (MLNN), learning vector quantization (LVQ) network, probabilistic neural network (PNN), general regression neural network (GRNN), Elman neural network (ENN), focused time delay neural network (FTDNN) and cascade feed-forward neural network (CFNN) to diagnose the thyroid disease. They were able to increase the accuracy of these networks by only creating suitable training data subsets and using proper network topologies and network parameters. Finally, the MLNN had been retrained one more time by using particle swarm optimization (PSO), artificial bee colony (ABC) and migrating birds optimization (MBO) algorithms, which are recent popular metaheuristic algorithms based on swarm intelligence. In developing the MLNN, they used logarithmic sigmoid function as the activation function and scaled conjugate gradient (SCG) algorithm as the training algorithm.

The study by Li et al. [10] present an effective and efficient computer aided diagnosis (CAD) system based on principal component analysis (PCA) and extreme learning machine (ELM) to provide an assistance for diagnosing thyroid disease. They conclude that the proposed method PCA-ELM can be considered as a new powerful tool for diagnosing thyroid diseases with a high accuracy and effectiveness. Turanoglu-Bekar et al., [11] have done a comparison on commonly used decision tree algorithms to classify the thyroid disease types and also have compared the performance using the six performance metrics (ACC (%), MAE, PRE, REC, FME, and Kappa Statistic). For diagnosis of thyroid diseases, the study [2] presented Linguistic Hedges Neural-Fuzzy Classifier with Selected Features (LHNFCSF). They have used classification accuracy and k-fold cross-validation to estimate the performance evaluation of their system.

A review on utilization of the distinctive neural system models for distinguishing the dysfunctionalities of thyroid has been discussed in [9]. Additionally, some parameter estimation methods, execution of the distinctive neural system models have been explored. In this survey, the authors have shown an elaborated work that has been done earlier using artificial neural networks with various kinds of algorithms in diagnosis of thyroid disease. Saiti et al. [12] have done a study on comparing thyroid disease diagnosis with two kinds of classifiers; support vector machines and probabilistic neural network while Aziz [13] has presented a study on thyroid disease diagnosis by using neural networks with second order training algorithm and has used genetic algorithm to find the optimum network structure with high classification accuracy. Prasad et al. [14] have used artificial intelligence techniques, rough data sets theory and machine learning algorithms to design a hybrid architecture. In Usage of rough data set theory has eliminated problem of dealing with uncertain data and missing attribute values in data [14].

III. DATA AND METHODOLOGY

For developing the models, a suitable dataset is required. In this research 2 datasets have been used in developing the models. These 2 datasets have some similar features as well as some variant features. One of those datasets is having 215 instances with 6 dimensions and the other set is having 1621 instances with 9 dimensions. Both the datasets were obtained from UCI machine learning repository. The first dataset consists of 5 laboratory test results related with thyroid diagnosis and the thyroid class to which each patient belongs. In this dataset there are three different classes to which a patient can be correlated with after the investigation of those laboratory tests. The other dataset consists of 9 attributes. This dataset includes some patient details like Age, Gender and Pregnancy of the patient with 5 laboratory test results and the thyroid class that each patient belongs. Data was cleaned and normalized before use and WEKA was used to develop the machine learning models. The models were of two types which use two different machine learning algorithms. The two algorithms are multilayer perceptron neural network with back propagations and random forest classifier.

The Neural Network models were designed by changing the attributes; number of hidden layers, learning rate of the network and momentum. On the other hand, Random Forest models were developed by varying the number of trees in the forest and number of features per tree. With different values for each attribute, several machine learning models were developed under both machine learning techniques.

When developing the models, training data was fed into the neural networks designed with different structures. The designing of the networks was done by passing values for the relevant fields like hidden layers, learning rate, momentum etc. When the network model was built while learning from the data set provided, the information after execution was displayed including the details about the structure of the network and the data set provided. When the inputs are fed into the input layer of the neural network, the outputs from each node in the hidden layers are calculated.

While these models are developing, each of them was tested using k-fold cross-validation. So, in this methodology, the original dataset is partitioned into k number of subsamples

randomly. Of the k subsamples, one subsample is retained as the validation data for testing the model, and the remaining k-1 subsamples are used for training. Then the cross-validation process is repeatedly done for k times which the number of folds, with each of the k subsamples used exactly once as the validation data. After that, it takes the average of the k results from the folds or otherwise it will combine all the results to make a single answer. Using all the data from the sample dataset for both training and validation and each instance used for validation only onetime is a benefit gain from this method.

Then all these models were evaluated and compared using the measures Accuracy (%), Mean Absolute Error, Root Mean Squared Error, True Positive rate, False Positive rate, Precision and Recall.

IV. RESULTS AND DISCUSSION

Discussion over the Results of Multi-layer Perceptron Neural Network with Back Propagation

When consider all the neural networks developed in this study using the Dataset 1, in Figure 1 it has shown a maximum accuracy of 96.7442% and a minimum accuracy of 69.7674%. As per these results neural networks with 3 or 5 neurons in a single hidden layer is performing better than others regardless of the learning rates. This will be due to the less complexity in the neural network. With the highest accuracy of 96.7442% and the lowest error values; 0.0289 for mean absolute error and 0.1274 for root mean squared error, neural network with 5 neurons in single hidden layer performs well. The neural network with 3 neurons in a single layer with learning rate 0.2 is also giving the same accuracy 96.7442%, but with somewhat higher error values when compared to error values of 5-hidden node structure. However, Temurtas et al. (2009) have reported that by selecting a hidden layer and Logarithmic sigmoid activation function for hidden layer and 6 neurons in the hidden layer, can reach the classification accuracy for thyroid disease to 98.6%.

By giving minimum accuracy and maximum error values, the neural networks with '5,10,5' and '5,10,10,5' structures perform less as shown in the graph in Figure 1 which would be due to the high complexity within the networks. This graph shows that for a single hidden unit the accuracy is somewhat low while again accuracy drops when the hidden units are going to be increased having the peak point around 3 and 5 hidden units.
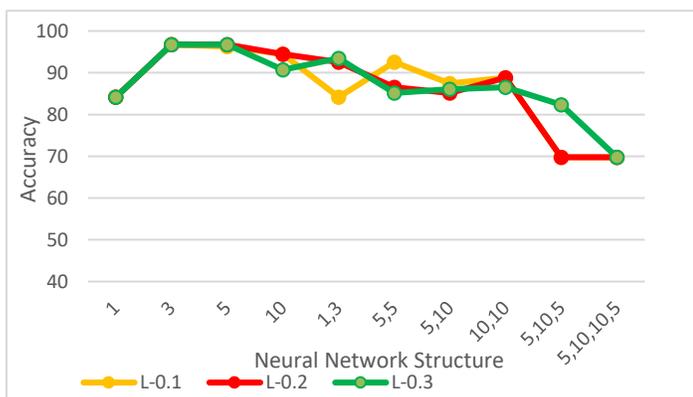

Figure 1: Accuracy of different neural networks for various learning rate for Dataset 1

For Dataset 2, as shown in Figure 2, 95.8904% is the highest accuracy and it is for the network with '5, 10' structure and learning rate 0.1. In Dataset 2, there are eight features as input values. So, to learn the patterns in it, it may need somewhat complex structure than a simple structure with a single hidden layer with few nodes. However, with a learning rate of 0.3, neural network with only 3 nodes in a single hidden layer gives an accuracy level at 95.3033%. Anyway, when the network becomes much more complex, regardless of the learning rate, the accuracy level drastically drops to about 42% as it shows in the Figure 2. It was clear that this drop is caused with the higher values in mean absolute error, root mean squared error, FP rate and lower values in TP rate, precision and recall compared to others.

Figure 2: Accuracy of different neural networks for various learning rates for Dataset 2

Discussion over the Results of Random Forest

Random Forest models give the highest accuracy level at 96.2791% for the Dataset 1 as shown in Figure 3. But Turanoglu-Bekar et al. (2016) have reported only 65.00 % of accuracy for Random Forest algorithm in their study while the highest accuracy rate with the value of 75% has been reported for NBTree. The results showed a significant thing that the accuracy remains unchanged for random forest models with 3 and 4 features per tree. This is even much clearer with the graph in Figure 3. Anyhow, for the random forest models with 5-featured trees, there are slight changes in the accuracy with the change in the number of trees in the forest. From the Figure 3, it is clear that accuracy drops when the number of trees in the forest gets increased. That is accuracy has become 94.8837% for 100-tree forest. This may be due to the reason that at a certain point performance of the tree stops from growing. Also, when the number of trees increases, to get the output there need an average over many trees which causes a changing in the output to another direction. Though, increasing the number of maximum features at each node can generally increase the performance of a Random Forest, as each node of the tree has a higher number of features available to consider, this would not be always the same.
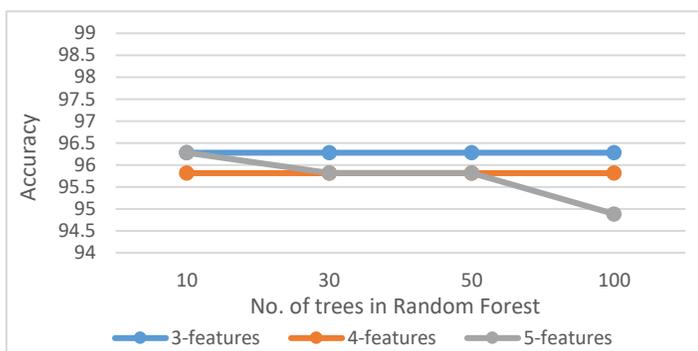
Figure 3: Accuracy of random forests with different number of trees and features for Dataset 1

For the Dataset 2, random forest models give little bit different results which is not as in the case for Dataset 1. As shown in the Figure 4, all the three groups are having the mean accuracy of 98.4915% and there is no significant gap between the lowest and the highest accuracy values. In 3-feature group, random forest with 100 trees gives the highest accuracy but its root MSE value 0.986 is the highest and TP rate 0.01 is the lowest within the group. For 5-feature group, random forest with 10 trees gives the highest accuracy but with the highest root MSE value 0.085 and mean absolute error which is almost near to the highest, 0.013 within the group. In the 8-feature group, when consider all the measures, random forest with 100 trees shows the best performance compared to others in that group as well as for models in the other groups.
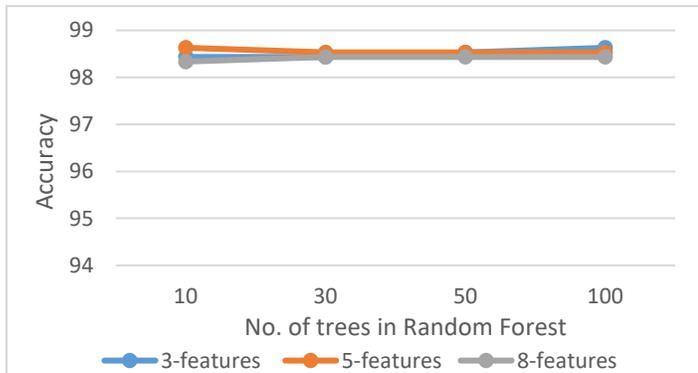

Figure 4: Accuracy of random forests with different number of trees and features for Dataset 2.

Comparison Between MLP-NN and Random Forest
For Dataset 1, Multilayer Perceptron Neural Network perform better than the Random Forest. On the other hand, for Dataset 2, Random Forest performs better. The main reason for this interchange may be the difference in the two datasets. Dataset 1 is having only 5 features while Dataset 2 is having 8 features. So, the patterns in those features are different and also learning patterns of different models are not the same. MLP-NN learn by adjusting weights for the branches in the network while Random Forest randomly create trees with best fit approach. Random Forest predicts results as an average of the outputs of all the trees in the forest, while in MLP-NN predict using the network built after updating the weights in a way that back-propagated error is

minimized. Both types of the models have precision closer to 1 which means that there is perfect precision.

In the study by (Makas and Yumusak (2013), they have reported the accuracy values between 98% and 99% for all the neural network models developed. This study shows that Scaled Conjugate Gradient algorithm gives higher accuracy rates. Experimental results from the study by Li et al. (2012) show the mean accuracy of 97.73% and the maximum accuracy of 98.1% for the PCA_ELM model developed. However, the classification accuracies obtained in the study done by Saiti et al. (2009) for the first dataset are at the maximum level and it is 100% for both the models while the accuracy for the second dataset being 97.9% and 99.7% for the model PNN with GA feature selection and SVM with GA feature selection, respectively. Although (Aziz, 2009) has reported that the proposed method achieved the highest accuracy rate which is in the range 96% and 98% when comparing the related previous studies except that which using genetic algorithm, in this research it has got the same accuracy results.

V. CONCLUSION

Thyroid hormones control all the metabolic functions of the human body. Therefore, thyroid is essential throughout the whole life and any dysfunctionality with it must not be underestimated. As a result, to provide an assistance for diagnosis of thyroid disorders, machine learning techniques are widely used today.

In diagnosing thyroid disorders, providing decision support using machine learning algorithms was the main goal in this research. This research focused on a comparative study on thyroid disease diagnosis by using two kinds of classifiers: Multilayer Perceptron Neural Network with Back-propagation and Random Forest algorithms. For training the models two datasets were used. Dataset 1 having 5 features and Dataset 2 having 8 features related for the thyroid disease diagnosis. The developed models were tested using 10-fold cross validation while models were mainly compared using Accuracy and many other measures like Mean Absolute Error, Root Mean Squared Error, True Positive rate, False Positive rate, Precision and Recall were considered for further evaluation of performance of the models.

Multilayer Perceptron Neural Network with Back-propagation algorithm performed well for Dataset 1 with an accuracy of 96.7442% and Random Forest algorithm performed well for the Dataset 2 with a mean accuracy level of 98.4915%. So, it can be concluded that Multilayer Perceptron Neural Network with Back-propagation and Random Forest perform well in diagnosing thyroid disorders and those machine learning algorithms and techniques can be used in medical decision support systems to diagnose thyroid diseases.

REFERENCES

[1] Prerana, P. Sehgal, and K. Taneja, "Predictive data mining for diagnosis of thyroid disease using neural network," Int. J. Res. Manag. Sci. Technol., vol. 3, no. 2, pp. 75–80, 2015, [Online]. Available: http://papersim.com/wp-content/uploads/Neural_Network_Thyroid_Disease_2015.pdf.

[2] A. T. Azar, A. E. Hassanien, and T. H. Kim, "Expert system based on neural-fuzzy rules for thyroid diseases diagnosis," Commun. Comput. Inf.

Sci., vol. 353 CCIS, pp. 94–105, 2012, doi: 10.1007/978-3-642-35521-9_13.

[3] H. Makas and N. Yumusak, "A comprehensive study on thyroid diagnosis by neural networks and swarm intelligence," in 2013 International Conference on Electronics, Computer and Computation (ICECCO), Nov. 2013, pp. 180–183, doi: 10.1109/ICECCO.2013.6718258.

[4] A. James E. Avenir, J. D. Villota, and M. F. Rosas, "Machine Learning: A Means of Diagnosing and Prescribing Treatments for Tuberculosis Patients," Int. J. Comput. Theory Eng., vol. 7, no. 3, pp. 201–206, 2015, doi: 10.7763/ijcte.2015.v7.957.

[5] G. D. Magoulas and A. Prentza, "Machine Learning in Medical Applications," in Machine Learning and Its Applications, Springer, Berlin, Heidelberg, 2001, pp. 300–307.

[6] I. Al-muwaffaq and Z. Bozkus, "MLTDD : Use of Machine Learning Techniques for Diagnosis of Thyroid Gland Disorder," in Computer Science & Information Technology (CS & IT), Apr. 2016, pp. 67–73, doi: 10.5121/csit.2016.60507.

[7] S. A. Hannan, V. D. Bhagile, R. R. Manza, and R. J. Ramteke, "Diagnosis and Medical Prescription of Heart Disease Using Support Vector Machine and Feedforward Backpropagation Technique," Int. J. Comput. Sci. Eng., vol. 02, no. 06, pp. 2150–2159, 2010.

[8] K. Yilmaz, "A fast intelligent diagnosis system for Thyroid diseases based on extreme learning machine," Anadolu Univ. J. Sci. Technol. Appl. Sci. Eng., vol. 15, no. 1, p. 41, May 2015, doi: 10.18038/btd-a.89202.

[9] S. Razia and M. R. Narasinga Rao, "Machine Learning Techniques for Thyroid Disease Diagnosis - A Review," Indian J. Sci. Technol., vol. 9, no. 28, Jul. 2016, doi: 10.17485/ijst/2016/v9i28/93705.

[10] L.-N. Li, J.-H. Ouyang, H.-L. Chen, and D.-Y. Liu, "A Computer Aided Diagnosis System for Thyroid Disease Using Extreme Learning Machine,"

J. Med. Syst., vol. 36, no. 5, pp. 3327–3337, Oct. 2012, doi: 10.1007/s10916-012-9825-3.

[11] E. Turanoglu-Bekar, G. Ulutagay, and S. Kantarcı-Savas, "Classification of Thyroid Disease by Using Data Mining Models: A Comparison of Decision Tree Algorithms," Oxford J. Intell. Decis. Data Sci., vol. 2016, no. 2, pp. 13–28, 2016, doi: 10.5899/2016/ojids-00002.

[12] F. Saiti, A. A. Naini, M. Aliyari Shoorehdeli, and M. Teshnehlab, "Thyroid Disease Diagnosis Based on Genetic Algorithms Using PNN and SVM," in 2009 3rd International Conference on Bioinformatics and Biomedical Engineering, Jun. 2009, pp. 1–4, doi: 10.1109/ICBBE.2009.5163689.

[13] S. B. Aziz, "Thyroid Disease Diagnosis using Genetic Algorithm and Neural Network," J. Al-Qadisiyah Comput. Sci. Math., pp. 1–13, 2009, [Online]. Available: https://www.iasj.net/iasj?func=article&aId=49972.

[14] V. Prasad, T. S. Rao, and M. S. P. Babu, "Thyroid disease diagnosis via hybrid architecture composing rough data sets theory and machine learning algorithms," Soft Comput., vol. 20, no. 3, pp. 1179–1189, Mar. 2016, doi: 10.1007/s00500-014-1581-5.

AUTHORS

First Author – W.K.D Jayamini, Department of Statistics and Computer Science, Faculty of Science, University of Kelaniya, Sri Lanka. darshawk@gmail.com

Second Author – K.G.H.D Weerasinghe, Department of Software Systems Engineering, Faculty of Computing and Technology, University of Kelaniya, Sri Lanka. hesiri@kln.ac.lk