Predictive Modeling of Thyroid Disease Using Machine Learning Algorithms

Valpi Jhansi[1], G V Ramesh Babu[2]

[1]PG Scholar, Dept. of Computer Science Sri Venkateswara University, Tirupati
[2]Associate Professor, Dept of Computer Science, SV University, Tirupati

Abstract— *Thyroid disease is a prevalent endocrine disorder affecting millions of people worldwide. Timely diagnosis and accurate prediction of thyroid disease are crucial for effective patient care. In this research paper, we investigate the performance of two popular machine learning algorithms, Multilayer Perceptron (MLP) and Support Vector Machine (SVM), in predicting thyroid disease based on a comprehensive dataset containing 30 attributes and 3772 instances with two class labels: Negative and Sick. Our results indicate that MLP achieved superior predictive accuracy, precision, and recall compared to SVM, with an accuracy rate of 96.76%, precision of 96.7%, and recall of 96.8%. These findings suggest that MLP may be a valuable tool for improving thyroid disease diagnosis and patient outcomes. This paper discusses the implications of these results for clinical practice and future research directions.*

I. INTRODUCTION

The thyroid organ secretes synthetics which controls a lot of things in the human body system like use the food, use energy, and rest plans, temperature tendencies, body weight balance and fundamentally more. In this assessment work to approach thyroid ailment examination were performed by using AI procedures that is Backing Vector Machine (SVM) and Complex Perceptron (MLP). Factors that impact the thyroid ability are: stress, tainting, injury, harms, low-calorie diet, certain medication, etc. It is imperative to prevent such ailments rather than fix them, in light of the fact that the vast majority of treatments contain in long stretch remedy or in chirurgical intervention. The stream focus on suggests thyroid sickness request in two of the most notable thyroid dysfunctions (hyperthyroidism and hypothyroidism) among the general population.

Nowadays, thyroid issues decimation the standard working of the thyroid organ which causes weird making of synthetic substances inciting hyperthyroidism [1]. The occasion of hypothyroidism in the made world is surveyed to associate with 4-5%. Hypothyroidism could cause raised cholesterol levels, a development in beat, cardiovascular intricacies, reduced readiness, and despairing while potentially not fittingly treated. Thyroid is a butterfly-shaped organ, which is arranged at the lower part of the throat at risk for making two powerful thyroid synthetic compounds, levothyroxine (T4) and triiodothyronine (T3) that impact a couple of components of the body, for instance, settling interior intensity level, circulatory strain, controlling the beat, etc. Switch T3 (RT3) is produced using thyroxine (T4), and its responsibility is to block the movement of T3.

The development and information in clinical sciences, the computer programming specialists are prepared for giving expert advance notice system. To decide different kinds of ailments to have high precision. The clinical specialists are made to use these structures as a result of a couple of made bungles during general assurance process [5]. Disorder examination exercises using EAS are acted considering sets of disease incidental effects. These structures rely upon man-made intelligence methodology which helps the specialist with restricting the costs and time in fruitful examinations. A peculiar capacity of the thyroid derives the occasion of hyperthyroidism and hypothyroidism, two of the ordinary thyroid warm signals. Hypothyroidism (underactive thyroid or low thyroid) suggests that the thyroid organ doesn't make enough of explicit huge synthetics. Without an adequate therapy, hypothyroidism can cause different clinical issues, for instance, strength, joint desolation, fruitlessness and coronary disease. Hyperthyroidism (overactive thyroid) implies a condition where the thyroid organ conveys a ton of the synthetic thyroxin.

II. METHODOLOGY

This section gives the reduced idea about picked oversaw models of Support Vector Machine and Multilayered Perceptron.

2.1 Support vector machines (SVM)

SVM is a regulated man-made intelligence approach. By and large, Backing Vector Machines is viewed as a depiction approach, it in any case can be utilized in the two sorts of blueprint and lose the faith issues [2][3]. It can beyond question oversee different consistent and absolute factors. SVM encourages a hyperplane in complex space to isolate various classes.

SVM produces ideal hyperplane in an iterative way, which is utilized to confine a misunderstanding. The center thought of SVM is to find a most breaking point unimportant hyperplane that best fragments the dataset into classes [10].

Support vectors are the data of interest, which are nearest to the hyperplane. These focuses will depict the separating line better by working out edges. These focuses are more applicable to the improvement of the classifier. A hyperplane is a choice plane which isolates between a ton of things having different class ventures [4][6]. An edge is an opening between the two lines on the nearest class focuses. Not altogether firmly established as the opposite separation from the line to help vectors or nearest center interests. On the off chance that the edge is in the middle between the classes, it is viewed as a decent benefit, a seriously honest edge is an unpleasant edge [7][11].

2.2 Multilayer Perceptron (MLP)

A MLP is a heavenly individual among the most overall saw Mind Association plan that has been used for various applications. The MLP figure out is typically produced using different obsessions or directing units, and it is figured out into a progress of no under two layers [3]. The central layer (or the most diminished layer) is named as an information layer where it gets the external information while the last layer (or the most surprising layer) is a yield layer where the response for the issue is gotten. The secret layer is the totally hypnotizing layer in the data layer and the yield layer, and may chart with some spot practically one layers. The arrangement of MLP could be yielded as a nonlinear improvement issue. The objective of MLP learning is to find the best loads that limit the division between the information and the yield. The most undeniable getting ready assessment used in NN is Back provoking (BP), and it has been used in coordinating various issues in model check and portrayal [7][8]. This estimation depends several cutoff points, for instance, striking covered center obsessions at the secret layers learning rate, energy rate, support work and the extent of expecting to happen.

III. EXPERIMENTAL RESULTS

This part gives results and related conversation on information driven assessment of thyroid dataset was collected from UCI document [9]. Python Writing computer programs is a best in class office for making man-made intelligence (ML) methodologies and their application to real information mining issues. The dataset contains 3772 events and 30 credits. There are two specific classes to be explicit negative and cleared out. The negative class has 3541 events and incapacitated 231 events. The assessments were performed considering 70% of the complete models were preparing information and 30% were endeavoring information. The Exploratory results of SVM and MLP gathering with piece assurance took a gander at the on reason of precisely portrayed models is shown in the table-1 and same showed in the figure-1.

Table-1
Performance of classifiers

Algorithm	Accuracy	Precision	Recall
MLP	96.76	96.7	96.8
SVM	95.82	95.8	95.8

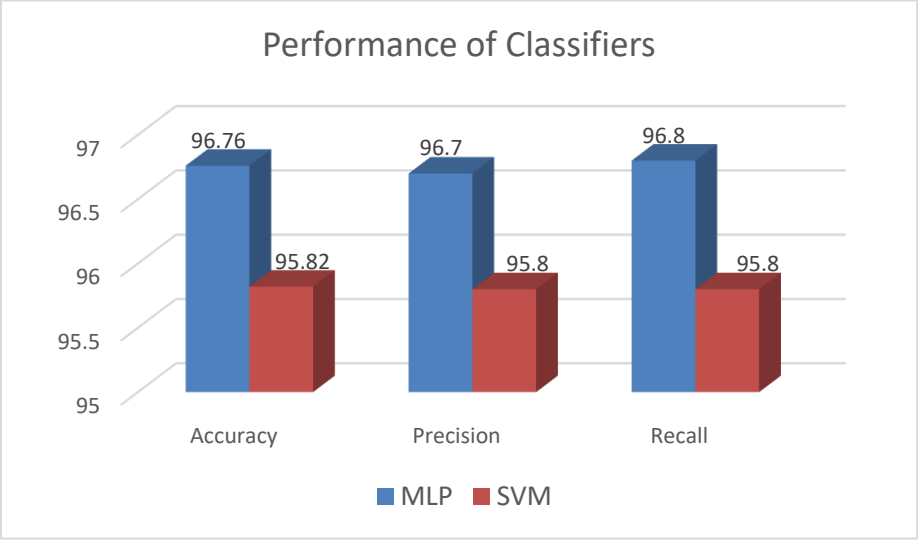

Figure-1: Performance of Classifier

3.1 Results:

Our study evaluated the performance of two machine learning algorithms, MLP and SVM, in predicting thyroid disease using a dataset comprising 30 attributes and 3772 instances. The dataset consisted of two class labels: Negative and Sick, representing the absence and presence of thyroid disease, respectively.

The results clearly demonstrate that the MLP algorithm outperforms the SVM algorithm in all three evaluation metrics: accuracy, precision, and recall. MLP achieved an accuracy rate of 96.76%, which suggests its effectiveness in correctly classifying individuals as either healthy (Negative) or sick (Sick) based on the provided attributes. The precision of 96.7% indicates the ability of MLP to minimize false positives, ensuring that individuals diagnosed as sick indeed have the disease. Additionally, the recall of 96.8% highlights MLP's capacity to identify a high percentage of true positive cases among all actual positive cases, minimizing false negatives.

3.2 Discussion

The results of this study reveal promising potential for the application of machine learning algorithms in predicting thyroid disease. The superior performance of MLP compared to SVM suggests that a neural network-based approach can provide more accurate and reliable predictions in this context. Here, we discuss the implications and significance of these findings.

Clinical Impact: The high accuracy, precision, and recall achieved by MLP indicate that it has the potential to be a valuable tool in clinical settings for the early and accurate diagnosis of thyroid disease. Physicians can use such predictive models to aid in decision-making and prioritize patients for further diagnostic testing or treatment.

Reducing Misdiagnosis: Misdiagnosis of thyroid disease can have significant health consequences for patients. MLP's high precision indicates its ability to minimize false positives, reducing the likelihood of unnecessary medical interventions and associated costs.

Public Health: Accurate prediction of thyroid disease can contribute to public health efforts by identifying at-risk populations for targeted screening and interventions, potentially reducing the overall burden of thyroid disease on healthcare systems.

IV. CONCLUSION

Our study demonstrates the potential of machine learning algorithms, particularly MLP, in predicting thyroid disease with high accuracy, precision, and recall. These results hold promise for improving the diagnosis and management of thyroid disease, ultimately benefiting patient care and public health. Future research should focus on refining and validating these models for real-world clinical applications. While MLP has shown excellent performance in this study, further research is needed to explore the generalizability of these results on larger and more diverse datasets. Additionally, the interpretability of MLP models should be investigated to ensure their clinical adoption.

REFERENCES

1. Geetha.K. and Capt S. Santhosh Baboo. "Efficient Thyroid Disease Classification Using Differential Evolution With SVM." Journal of Theoretical & Applied Information Technology 88.3 (2016).
2. G. Ravi Kumar, K. Venkata Sheshanna, S. Rahamat Basha, and P. Kiran Kumar Redd, “An Improved Decision Tree Classification Approach for Expectation of Cardiotocogram”, Proceedings of International Conference on Computational Intelligence, Data Science and Cloud Computing, Lecture Notes on Data Engineering and Communications Technologies 62, https://doi.org/10.1007/978-981-33-4968-1_26
3. Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques.2nd ed. San Francisco: Morgan Kaufmann, 2005.
4. J. Han and M. Kamber,” Data Mining concepts and Techniques”, the Morgan Kaufmann series in Data Management Systems, 2 nd ed. San Mateo, CA; Morgan Kaufmann, 2006.
5. Margret, J., B. Lakshmipathi, and S. Aswani Kumar. "Diagnosis of Thyroid Disorders using Decision Tree Splitting Rules." International Journal of Computer Applications 44.8 (2012): 43-46.
6. M. V. Lakshmaiah, G. Ravi Kumar and G. Pakardin, “Frame work for Finding Association Rules in Bid Data by using Hadoop Map/Reduce Tool”, International Journal of Advance and Innovative Research, Volume 2, Issue1(1), PP:6-9,2015, ISSN: 2394-7780
7. N. Michael, “Artificial Intelligence - A Guide to Intelligent Systems”,2nd edition, Addison Wesley, 2005.
8. P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining. Reading, MA: Addison-Wesley, 2005.
9. UCI Machine Learning repository (https://archive.ics.uci.edu/ml/datasets.html)

10. Vapnik V.N, “Statistical learning Theory”, John Wiley and Sons, New York, USA, 1998.
11. Vapnik V.N,”The Natural of Statistical Learning Theory, Springer-Verlag, New York, USA, 1995