Thyroid Development Of A Deep Learning Model To Predict Thyroid Diseases Based On Medical Characteristics

Chethana T
M.Tech Student
Dept of CSE, SIET, Tumkur
chethanac428@gmail.com

Dr. Girish L
Associate Professor
Dept of AI&DS, SIET, Tumkur
girishltumkur@gmail.com

Dr. Basavesha D
Associate Professor
Dept of CSE, SIET, Tumkur
basavesha@gmail.com

Abstract

The thyroid gland secretes hormones that control metabolism. It is a vascular organ essential to human physiology. The two main conditions that influence it are hyperthyroidism and hypothyroidism, which cause aberrant hormone levels to be released, upsetting metabolic equilibrium. Although they are frequently noisy and hazy, blood tests for thyroid function are essential for diagnosis. In order to improve these tests for precise analytics and make the prediction of illness risk possible, data purification techniques are used. To predict a patient's risk of thyroid disease, machine learning techniques such as logistic regression, decision trees, and support vector machines, K-nearest neighbors, and artificial neural networks are used. A online application makes it easier for users to enter data, which helps with illness prediction.The principal diagnostic modality for determining the likelihood of malignancy in thyroid nodules and directing the process of fine-needle aspiration is ultrasound (US).However, unneeded FNA and operations are widespread because of operator-dependency and moderate to large interobserver heterogeneity in picture interpretation. Artificial intelligence (AI)-based computer-aided diagnostic (CAD) systems have been introduced as a solution to this problem. By offering a uniform and correct interpretation of US features, these systems hope to cut down on pointless FNA procedures. This paper examines future directions for tailored and optimal nodule care while providing an overview of the state-of-the-art AI-based CAD systems for thyroid nodules.

I. Introduction

Computational biology's advancement in the healthcare industry enables the utilization of stored patient data for disease prediction. Various prediction algorithms aid in early disease diagnosis. While medical information systems contain vast datasets, just a small number intelligent systems can effectively analyze diseases. As time goes on, machine learning algorithms have become pivotal in resolving complex and nonlinear issues when developing models. In disease prediction, models select features from different datasets to classify healthy patients accurately. Failure to do so may result in misclassification and unnecessary treatment for healthy individuals.Detecting thyroid disease proactively is crucial for timely treatment and saving lives and healthcare costs.Machine techniques and methods for deep learning are being used to anticipate thyroid disorders in theirearly phases and recognize types such as hyperthyroidism and hypothyroidism, thanks to developments in data processing and computation. Healthcare has undergone a revolution because to technologies like data mining, big data, parallel computing, image and video processing, and early disease detection. These technologies have also made it possible to predict viral outbreaks, find new drugs, manage healthcare data, and propose individualized therapy. Healthcare professionals aim to identify diseases early to provide prompt treatment and achieve faster recovery with lower costs. Thyroid disease, affecting millions worldwide, underscores the importance of proactive detection. [1]

According to the American Thyroid Association, 20 million Americans have a thyroid condition of some kind disease.Current research primarily focuses on binary classification problems, distinguishing between thyroid patients and healthy subjects, with only a handful addressing multiclass detection, typically categorizing subjects into normal, hypothyroidism, and hyperthyroidism. Feature selection remains largely understudied or neglected in thyroid disease research, with an overemphasis on optimizing machine learning and deep learning models. Despite claims of high accuracy, these approaches are often tested on small sample sizes, typically under 1000, and lack validation. There's a need for classification based on patient status, including treatment condition, health status, and general health issues, to effectively predict thyroid disorders as well as administer proactive treatment. Additionally, Performance comparisons between There are insufficient machine learning and deep learning models. This research attempts to address these gaps and provides the following contributions.Machine learning algorithms offer effective solutions to challenging

problems . Classification, a data extraction technique within machine learning, is widelyutilized to forecast and recognize a range of illnesses, including thyroid disease, because of its high performance and efficiency . While the integration of computer learning and artificial intelligence in medicine has a long history [, there's a growing recognition of the potential for machine learning-driven healthcare solutions.Analysts predict that in the near future, machine learning will be used more and more in the healthcare industry.[2]

Hyperthyroidism is characterized by an overactive thyroid gland, leading to excessive release of thyroid hormones . Symptoms include dry skin, sensitivity to temperature changes, hair thinning, fast heartbeat, elevated blood pressure, and weight loss, excessive sweating, enlargement of the neck, nervousness, shortened menstrual cycles, irregular bowel movements, and trembling hands. [3]

Hypothyroidism, conversely, occurs when the gland responsible for thyroid is underactive.Hyperthyroidism occurs when the thyroid gland produces an excessive amount of thyroid hormones, typically stemming from elevated levels of these hormones.Hyperthyroidism symptoms include dry skin, heightened sensitivity to temperature changes, hair thinning, fast heartbeat and weight loss, hypertension, excessive sweating, enlargement of the neck, nervousness, shortened menstrual cycles, irregular bowel movements, and trembling hands.[4]

On the other hand, hypothyroidism appears when the thyroid gland is unable to create enough thyroid hormones. In medical terminology, the prefix "hypo" denotes inadequacy or deficiency. The main causes of hypothyroidism include thyroid gland damage and inflammation. Weight gain, a slowed heartbeat, increased temperature sensitivity, neck swelling, dry skin, hand numbness, hair loss, gastrointestinal issues and high menstrual flow are a couple of the potential symptoms. These symptoms can gradually worsen if they are not treated.[5]

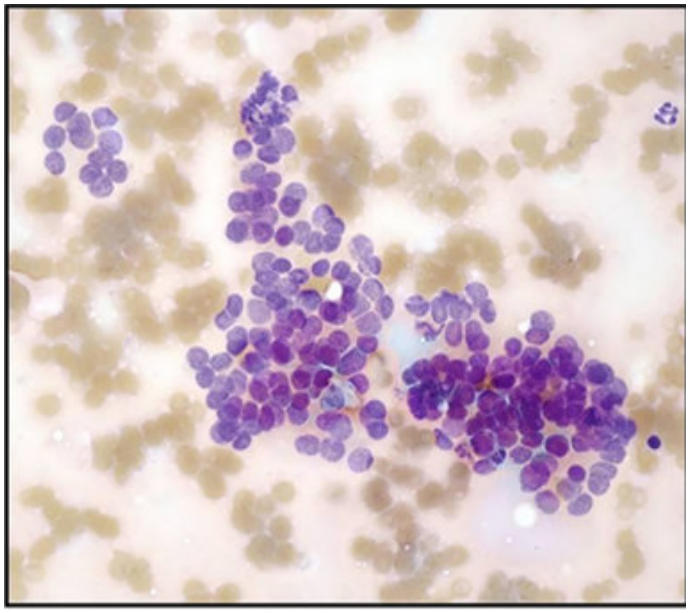

Fig. 1. Microscopic image of thyroid cell

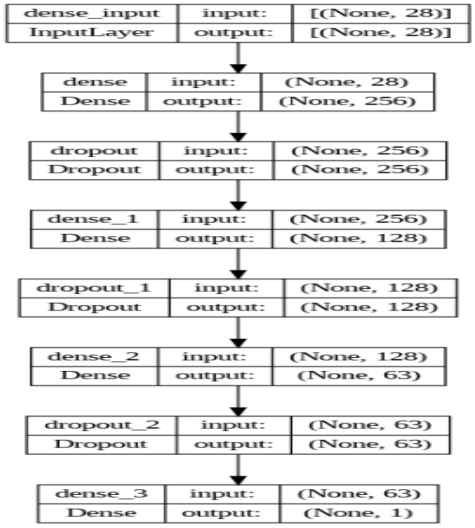

Fig. 2. Proposed Model Working Flow

II. Literature Survey

Technological developments in computers and data processing have resulted in the application of machine learning and deep learning techniques in a number of research that attempt to predict thyroid illness. It is crucial to diagnose and categorize thyroid issues, such as cancer, hyperthyroidism, or hypothyroidism, as soon as possible to enable prompt treatment and recovery.

A literature survey was conducted utilizing peer-reviewed databases like Scopus and Google Scholar , focusing on the past five years to gather recent insights. The search employed keywords like "Thyroid disease," "Thyroid cancer," "machine learning," and "deep learning" to pinpoint relevant articles. Despite yielding a substantial number of results, further refinement was performed to narrow down the selection. Initial screening identified over 100 pertinent articles, from which 25 were meticulously chosen based on their close relevance to the study. Machine learning and deep learning methods have been employed for both thyroid detection of thyroid disorders and malignancy, each requiring distinct approaches, thus warranting separate discussions. [6]

The aforementioned investigation identified ultrasonic features related with malignant thyroid nodules using the logistic regression (LR) model in conjunction with the least absolute shrinkage and selection operator (LASSO). Subsequently, logistic lasso regression (LLR) and random forest (RF) were used in conjunction with a grading system

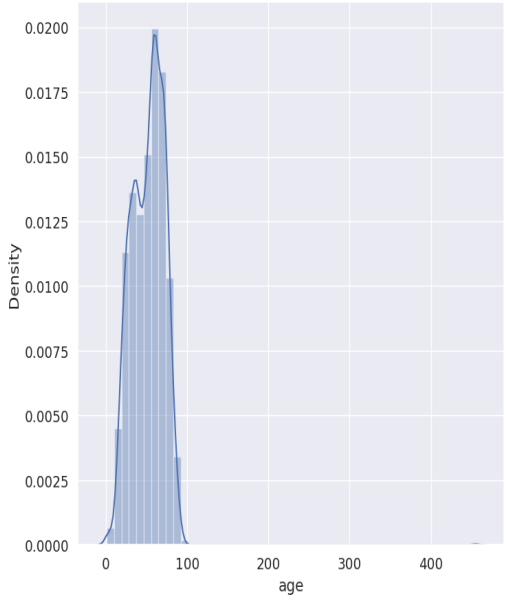

Fig. 3. Distribution of ages in a population

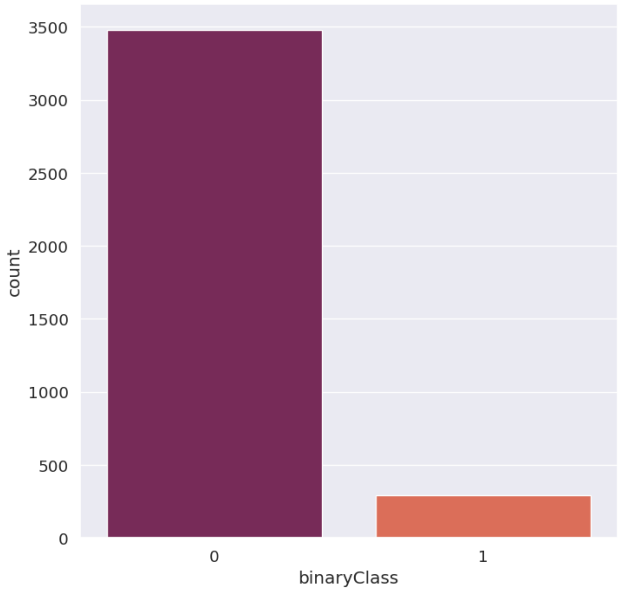

Fig. 4. Binary Class Classification

to diagnose malignant nodules, resulting in an accuracy rate of 82%. [7]

In a different study, machine learning was utilized to predict if confirmed malignant thyroid nodules included the BRAF mutation. The study examined 86 radiomic characteristics taken from 96 thyroid nodule ultrasound images using three distinct machine learning models: logistic regression (LR), random forest (RF), and support vector machine (SVM). The data showed that the three models had a classification accuracy of 64.3%. Idarraga along with others. studied machine learning-based prediction of thyroid nodule malignancy using ultrasonic and fine-needle aspiration (FNA) data to prevent false-negative diagnosis in cases of early-stage thyroid cancer. Random forest (RF) outperformed gradient descent (GD) and decision trees (DT) approaches. Despite these efforts, Since the effectiveness of these procedures for the diagnosis of thyroid cancer is not perfect, there is still potential for development.Numerous classification techniques were used in this study, such as the k-Nearest Neighbor algorithm, Artificial Neural Network, Decision Tree, and Support Vector Machine. The UCI Repository provided the dataset that was utilized for classification and prediction. The resulting outputs were checked for accuracy. To identify the most accurate and effective strategy, the study looked at and evaluated the accuracy of the algorithms used.[8]

III. Methodology

We initially acquired an illness dataset from the UCI data repository, featuring multiple entries on thyroid disorders and various target categories. To ensure balanced representation, we used data balancing techniques. For the average class, which comprised 6771 samples, we randomly sampled 400 instances to achieve dataset equilibrium. The feature selection process was then conducted using various methods. Next, using an 80-20 train-test split, several machine learning and deep learning models were evaluated.[9]

For the average class, which comprised 6771 samples total, we randomly sampled just 400 instances in order to establish dataset equilibrium—much more than other target classes. Next, a number of techniques were used to carry out the feature choice process. After that, a variety of deep learning and machine learning models were tested utilizing an 80-20 train-test split.The dataset contains a great deal of information about thyroid-related illnesses and covers a wide range of target classes. We proceeded with the feature selection procedure using a range of feature selection algorithms after data balance. Finally, trials were conducted using an 80-20 train-test split, utilizing a range of machine learning and deep learning models.To anticipate thyroid disease, a blood report analysis is required.Thus, we will evaluate a Thyroid blood test dataset using a range of supervised machine learning classifier techniques. Taking into account the accuracy of multiple approaches, we will select the best-performing algorithm to generate the intended results. First, we will obtain the UCI repository's Thyroid dataset. The collection will consist of records with the labels hyper and hypo that relate to hyperthyroidism and hypothyroidism, respectively.The dataset must be examined for any discrepancies, such as null or redundant data, before training. To get rid of such data, data cleaning methods will be put in place. [10]

We contrasted our suggested method with previous research on disease prediction using numerical or categorical datasets in

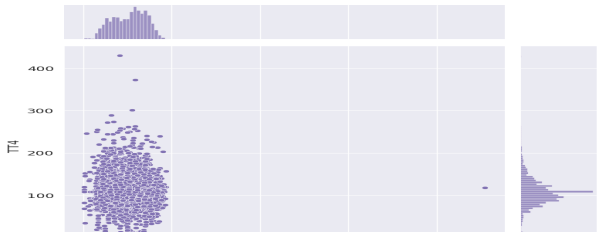

Fig. 5. Thyroid gland

order to highlight the importance of our approach. We used the techniques from a few chosen recent papers, assessing them according to F1 score and accuracy on our dataset. With these earlier studies, we didn't use feature selection; instead, we applied their methods straight to our dataset.

We put into practice the study's strategy, which predicted thyroid illness using RF. In a similar vein, we applied the study's method (which suggested a DNN for thyroid disease) and methodology (which employed DT for thyroid disease prediction). We also used the methodology from the study, which was successful on a dataset related to heart illness. This study merged three machine learning models—SVM, LR, and stochastic gradient descent classifier—into a hybrid model that used a CNN to extract features.

Our method outperformed all other research by a large margin, receiving a score of 0.99 for each evaluation parameter. Table 18 compares our methodology with previous research.

IV. Conclusion

The importance of detecting thyroid disease has increased recently, making the creation of effective computerized prediction models necessary. Previous research frequently places more of an emphasis on feature engineering and model optimization than on feature selection. Furthermore, short datasets are usually used in these research for model evaluation without appropriate validation.To address these shortcomings, this paper provides a technique that blends feature selection with machine learning and deep learning models. Numerous techniques for feature selection are used in conjunction with machine learning and deep learning models, such as FFS, BFS, BiDFE, and additional tree classifier-based features. The goal of the project idea "Thyroid Detection using Machine Learning" is to develop a clever and precise technique for identifying thyroid disease. The study improves the accuracy of thyroid illness prediction by training a dataset with the logistic regression technique. With this system, users enter their data using an online application, and the backend model handles the processing of the data. The outcome, which indicates whether the person has normal thyroid function, hyperthyroidism, or hypothyroidism, is quickly shown on the screen. Providing society with a dependable and efficient machine learning solution for applications is the primary objective related to disease diagnosis.

References

[1] J. Cai, J. Luo, S. Wang, and S. Yang. Feature machine learning selection: A fresh viewpoint. *Neurocomputing*, 300:70–79, 2018.

[2] Rajasekhar Chaganti, Furqan Rustam, Isabel De La Torre Díez, Juan Luis Vidal Mazón, Carmen Lili Rodríguez, and Imran Ashraf. Predicting thyroid illness using machine learning algorithms and selected features. *Cancers*, 14(16):3914, 2022.

[3] R. Chandan, M. S. Chethan, C. Vasan, and H. S. Devikarani. *International Journal of Science and Technology Applications in Engineering (IJEAST)*, 5(9), 2021.

[4] G Chaubey, D Bisen, S Arjaria, and V Yadav. utilizing machine learning techniques to predict thyroid illness. *Natl. Acad. Sci. Lett.*, 44:233–238, 2021.

[5] Sunila Godara. Prediction of thyroid disease using machine learning techniques. *International Journal of Electrical and Electronics Engineering (IJEE)*, 2018.

[6] L. Hong, M. Luo, R. Wang, P. Lu, W. Lu, and L. Lu. Big data in healthcare: Uses and difficulties. *Data and Information Management*, 2:175–197, 2018.

[7] I Ioniță and L Ioniță. Thyroid illness prediction with data mining methods. *BRAIN Broad Res. Artif. Intell. Neurosci.*, 7:115–124, 2016.

[8] P. Pavani and P. P. Sadu Naik. *International Journal of Innovative Research and Technology (IJIRT)*, 9(3), 2022.

[9] S. Solorio-Fernández, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad. An overview of techniques for unsupervised feature selection. *Artificial Intelligence Review*, 53:907–948, 2020.

[10] Emrullah Sonuç et al. Classification of thyroid diseases using machine learning methods. In *Journal of Physics: Conference Series*, volume 1963, page 012140. IOP Publishing, 2021.