A Comprehensive Approach for Thyroid Cancer Prediction Using Machine Learning Models

S. Santhoshini
Department of Electronics and Communication Engineering, Sir M. Visvesvaraya Institute of Technology, Bangalore, India | Visvesvaraya Technological University, Belagavi, India
santhoshini_ec@sirmvit.edu (corresponding author)

M. A. Goutham
Department of Electronics and Communication Engineering, Adichunchanagiri Institute of Technology, Chikkamagaluru, India | Sir M. Visvesvaraya Technological University, Belagavi, India
magoutham@gmail.com

Received: 6 June 2025 | Revised: 26 June 2025, 17 July 2025, 25 July 2025, and 2 August 2025 | Accepted: 3 August 2025

 DOI: https://doi.org/10.48084/etasr.12598

ABSTRACT

This study sought to predict the appearance of thyroid cancer by employing machine learning methods on an extensive collection of clinical and demographic variables. The Random Forest (RF) algorithm is the foundation of the prediction model, which combines diverse data sources to enhance its predictive accuracy. The preprocessing steps involved handling missing values, normalizing data, and selecting relevant features, ensuring high-quality inputs for the model. The RF model demonstrated high recall, precision, and accuracy in the prediction of thyroid cancer, validated through rigorous cross-validation techniques. The results highlight the potential of machine learning to improve early and timely detection and management of thyroid cancer, thereby leading to better patient outcomes. A user-friendly Flask-based frontend was developed to make real-time risk predictions accessible to healthcare professionals.

Keywords-thyroid cancer; machine learning; random forest; data preprocessing; real-time predictions

I. INTRODUCTION

Thyroid cancer represents the leading endocrine cancer worldwide, with increasing rates that constitute a major public health challenge. This growth pattern is commonly associated with advances in diagnostic methods and increased use of medical imaging, which has led to the discovery of smaller, often symptom-free tumors. However, evidence also supports an actual increase in larger and more invasive cancer forms. In terms of worldwide thyroid cancer statistics, according to [1], approximately 821,214 new diagnoses occurred worldwide in 2022, corresponding to an age-standardized incidence rate of 9.10 cases per 100,000 people. This study identified a consistent upward pattern between 1990 and 2019, showing that the global age-standardized incidence rate increased from 2.01 to 2.83 per 100,000 population over this 29-year period. In [2], predictive analysis suggested that the incidence of thyroid cancer will increase from approximately 233,847 cases in 2019 to 305,078 cases by 2030, marking a projected increase of approximately 30.46% during this period. The diverse types of thyroid cancers and limitations in current diagnostic methods pose significant challenges that can greatly affect patient outcomes. The thyroid gland, situated in the neck, plays a critical role in regulating metabolism through hormone production. Early identification and precise diagnosis lead to successful treatment and better results for patients. Traditional diagnostic methods include physical examinations, Fine-Needle Aspiration (FNA) biopsies, and imaging techniques such as ultrasound. Blood tests, ultrasound imaging, and invasive methods, such as biopsies, are used to detect thyroid cancer. FNA offers a minimally invasive alternative to more involved procedures such as incisional or excisional biopsies [3], and it is a quick way to collect cell samples, helping doctors confirm diagnoses or guide treatment decisions. FNA cytology is an important diagnostic technique in the evaluation of thyroid nodules, praised for its minimally invasive nature and cost-effectiveness. However, its diagnostic utility is hindered by several inherent limitations. A notable challenge is the rate of inadequate sampling, which ranges from 5-15% of specimens, necessitating repeat procedures. In addition, a significant proportion of samples, approximately 15-30%, yield indeterminate cytological results, leaving clinicians with an unclear diagnosis. Compounding these issues, FNA often struggles to definitively differentiate between benign and malignant follicular adenomas based solely on cellular characteristics, posing a considerable diagnostic dilemma.

The field of thyroid cancer detection is being revolutionized by the adoption of Artificial Intelligence (AI) and Machine Learning (ML). These cutting-edge technologies offer an unparalleled opportunity to overcome the long-standing limitations of traditional diagnostic approaches and techniques.

Human assessments of thyroid nodule malignancy are prone to errors and may not consistently provide an accurate preoperative diagnosis. This highlights a crucial need for more objective and reliable diagnostic tools. ML algorithms are particularly useful in identifying complex patterns and subtle relationships within vast multidimensional datasets that often escape human observation. This makes them well-suited for analyzing the diverse clinical, imaging, and molecular data associated with thyroid cancer. In [4], a clinical decision support model was acquired using the Bagged Classification and Regression Trees (Bagged CART) model to predict thyroid cancer. This study utilized a dataset of 724 patients from Shengjing Hospital, China, including data on nodular malignancies, demographics, ultrasound characteristics, and blood test results. The Bagged CART model was gauged using metrics such as accuracy (99.1%), balanced accuracy (98.7%), sensitivity (99.7%), specificity (97.7%), positive predictive value (99.1%), negative predictive value (99.2%), and F1-score (99.4%). The most important variables influencing the predictions were identified as nodule size, TSH levels, enriched blood flow, multilateral presence, FT4 levels, nodule site (isthmus), and patient age. The study concluded that this model effectively detected thyroid cancer, aiding in faster decision-making and enhancing preventive medicine practices. In [5], an ML system/framework was developed to detect thyroid nodule malignancy using a clinical dataset of 724 patients and 1232 nodules. Six distinct ML approaches were implemented: Gradient Boosting Machine (GBM), Logistic Regression (LR), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) with both radial and linear kernels, and Random Forest (RF). The models were evaluated using ten-fold cross-validation, bootstrap analysis, and permutation predictor importance. RF achieved the highest prediction accuracy of 79.31% and an AUC/ROC of 0.8541, demonstrating superior performance over expert assessments. GBM showed the highest sensitivity at 87.50%, highlighting the model's ability to outperform human judgment in predicting nodule malignancy and emphasizing the significance of variables such as calcification, laterality, blood flow, and location.

In [6], ML methods were also investigated for thyroid cancer prediction, utilizing established classification algorithms to determine cancer stages and probabilities. This study achieved exceptional classification and clustering accuracy (98.72%) while requiring fewer input variables than comparable existing systems. The study included a comprehensive analysis and comparison of various feature selection methods. In [7], a thyroid cancer classification method used a Deep Learning (DL) framework on ultrasound images from the Digital Database of Thyroid Ultrasound Images (DDTI). This study utilized ultrasound images with 61 benign and 289 malignant nodules based on TI-RADS classifications. Multiple DL models were fine-tuned to improve diagnostic accuracy. VGG16, MobileNet, and Inception ResNet V2 achieved the highest precision of 99.9%, showing their potential as reliable diagnostic tools for thyroid cancer. In [8], a new approach was proposed for classifying thyroid textures in ultrasound images using autoregressive features with three different ML classifiers: Artificial Neural Network (ANN), Random Forest Classifier (RFC), and Support Vector Machine (SVM). This study utilized two datasets consisting of 2D thyroid ultrasound images. Autoregressive modeling computed 30 spectral energy-based features to distinguish between thyroid and non-thyroid textures. Training the classifiers using these features resulted in classification accuracies of around 90%. ANN functioned slightly better than other classifiers with a Dice Coefficient (DC) of 0.930 on one dataset. In [9], an AI-based method was proposed for thyroid nodule classification using both spatial and frequency domain information. This study utilized the Thyroid Digital Image Database (TDID), which includes ultrasound images of 298 patients. Features were retrieved from the image using a combination of DL with Fast Fourier Transform (FFT) techniques. A cascade classifier scheme was employed, where the FFT-based method first classified the images into benign, benign-malignant, or malignant categories, followed by a CNN-based method for further classification of benign-malignant cases. Pre-trained ResNet models (ResNet18, ResNet34, and ResNet50) were fine-tuned to enhance performance. The ResNet50-based model achieved the highest accuracy of 90.883%, with sensitivity and specificity of 94.933% and 63.741%, respectively.

In [10], a Computer-Aided Diagnosis (CAD) system was developed using Multiple-Instance Learning (MIL) to classify thyroid nodules as benign or malignant based on ultrasound images. This study utilized data from 99 cases, including 33 benign and 66 malignant nodules. Image preprocessing involved median filtering and binarization, followed by segmentation using the Grey-Level Co-occurrence Matrix (GLCM) to extract features from ultrasound images. The dataset was divided into 87% training and 13% validation sets. The SVM algorithm attained an accuracy of 96%, sensitivity of 78.66%, and specificity of 100%, outperforming ANN, which had an accuracy of 75%. In [11], an ML-based approach was developed to predict malignant and metastatic thyroid cancer using pre-operative demographic and laboratory data from 1735 patients [11]. ML models included LR, Ridge Regression (RR), and Extreme Gradient Boosting (XGBoost). The XGBoost model displayed superior performance with an AUC of 0.84 for malignancy and 0.72-0.77 for metastasis. Significant risk factors involved age, obesity, prothrombin time, fibrinogen, and hepatitis B virus 'e' antibody (HBeAb), while protective factors included monocyte count, D-dimer, T3, FT3, and albumin. Tumor size was highlighted as a critical indicator of metastasis. The XGBoost model outperformed predictions based solely on TI-RADS. In addition, an online tool was developed for clinicians to predict the risk of malignancy and metastasis using these models. In [12], a comprehensive review of AI techniques for thyroid cancer diagnosis was provided, covering supervised, unsupervised, and ensemble learning methods. This review included DL models, namely, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), traditional ML classifiers such as SVM and Decision Trees (DT), and Probabilistic Models (PMs). The study highlighted the importance of high-quality datasets and feature selection methods to improve diagnostic accuracy. In [13], a retrospective cross-sectional study was conducted to increase the detection of Papillary Thyroid Cancer (PTC) using computerized cytological features of FNA Cytology (FNAC) samples. The study included 240 cases, with 110 histologically confirmed PTCs and 130 benign cases. Eight significant cytological features were quantified, including mean nuclear size, nuclear elongation, nuclear-cytoplasmic saturation ratio, and inclusion index. Stepwise LR identified six significant features for predicting malignancy. The LR model achieved an AUC/ROC of 87.7% in differentiating PTC from benign nodules, with 84.5% sensitivity and 81.0% specificity. The model also showed AUC/ROCs of 81.3% for atypia and 78.7% for suspicious malignancy categories in indeterminate FNAC results. In [14], ML models were used in data from 2,444 patients with PTC to predict structural relapse in PTC. The analysis covered 29 perioperative variables in four dimensions: demographic characteristics, tumor-related variables, Lymph Node (LN)-related variables, and metabolic and inflammatory markers. Five ML algorithms were employed, namely LR, SVM, XGBoost, RF, and a Neural Network (NN). The RF model achieved the best results with an AUC of 0.766, a sensitivity of 0.676, a specificity of 0.784, and an accuracy of 0.775, exceeding the ATA risk stratification system (AUC = 0.620). The key predictors included thyroglobulin (Tg), LN Ratio (LNR), and stage N. In [15], Deep CNN (DCNN) models were developed to detect thyroid cancer from ultrasound images, using ultrasound images of 42,952 patients, including 17,627 with thyroid cancer and 25,325 controls, obtained from three hospitals in China. DCNN models based on the ResNet-50 and Darknet-19 architectures were evaluated, achieving AUC values of 0.947 for the internal validation set, 0.912 for the Jilin external validation set, and 0.908 for the Weihai external validation set. These models outperformed experienced radiologists in specificity and showed comparable sensitivity. The DCNN model achieved a sensitivity of 93.4% and specificity of 86.1% in the Tianjin internal validation set.

In [16], feature selection and ML algorithms were applied to the SEER database to detect thyroid cancer, analyzing 34 clinical variables to build classifiers that distinguished between patients with more than ten years of survival and those who did not survive at least five years. Multilayer Perceptrons (MLPs) were used using different numbers of independent variables. MLP-1, which utilized seven variables, achieved the highest accuracy of 94.5%. Feature selection algorithms identified age, primary disease extent, and location of nodal disease as the most predictive variables, leading to MLP-2, which used three variables and achieved 91.1% accuracy. A third model based on the TNM staging system achieved 80.87% accuracy.

In [17], an ML-assisted diagnostic system was developed to improve the accuracy of differentiating malignant from benign thyroid nodules using Ultrasound and Real-Time Elastography (RTE) data. This study analyzed 2064 thyroid nodules from 2032 patients. Nine different ML algorithms were assessed, including L2-LR, LDA, RF, SVM, AdaBoost, k-nearest neighbours, NN, Naive Bayes, and CNN. RF achieved the highest performance, with 0.938 AUC, 89.1% sensitivity, 85.3% specificity, and 85.7% accuracy when combining ultrasound and RTE features. In [18], the progress in DL methods for the diagnosis of thyroid conditions was evaluated, emphasizing various models that included CNNs, Generative Adversarial Networks (GANs), autoencoders, Long Short-Term Memory (LSTM), Deep Belief Networks (DBNs), and RNNs. CNNs demonstrated better performance in classifying thyroid nodules from ultrasound images, with models such as ResNet50 and VGG16 achieving high accuracy and sensitivity. GANs were recognized for generating high-quality synthetic images, which aids in data augmentation. The difficulties in incorporating DL approaches into medical settings were discussed, such as requirements for extensive labelled data, privacy issues regarding patient information, and the absence of uniform assessment criteria. In [19], an ML-based approach was developed to detect thyroid diseases using selective features from a dataset consisting of 9172 samples and 31 features [20]. Four feature selection techniques were investigated: Forward Feature Selection (FFS), Backward Feature Elimination (BFE), Bidirectional Feature Elimination (BiDFE), and ML-based Feature Selection (MLFS) using an extra tree classifier. The selected features were used to train ML models, including RF, LR, SVM, AdaBoost, and GBM, as well as DL models such as CNN, LSTM, and CNN-LSTM. The RF classifier with MLFS achieved the highest accuracy of 99%. In [21], a literature review was performed on six online databases, adhering to PRISMA-ScR guidelines. This review focused on English-language publications from 2014 to 2024 that investigated ML applications in thyroid cancer, and 21 of the most pertinent full-text articles were selected. The authors concluded that ML techniques demonstrated significant potential in thyroid cancer research, providing innovative solutions for diagnosis, predicting metastasis, forecasting prognosis, and personalizing treatment. The authors were also of the view that ML also held promise for drug discovery, tailoring treatment strategies, and long-term patient monitoring, areas that could profoundly reshape thyroid cancer care.

These works used datasets sourced from different sources and different ML models. In addition, results were evaluated using different metrics, such as F1-score, Recall, AUC, etc. When looking at the existing literature on ML techniques for thyroid cancer detection/prediction, almost all existing works have used homogeneous datasets having the same structure. The novelty of this study lies in the fact that it uses heterogeneous datasets in CSV and XML formats. This study extracts the required features from diverse datasets and combines them into a homogeneous one as a preprocessing step. Then, this dataset is used to train and test ML models.

II. METHODOLOGY

It is a well-known fact that AI algorithms struggle with accuracy due to the lack of enough labeled clinical outcome data. Experts agree that neural networks need large amounts of data to give reliable results. This shortage of comprehensive and well-annotated datasets on how cancer occurs and spreads slows the development and training of AI models. The primary challenge in thyroid cancer diagnosis is the integration and analysis of diverse data sources, including clinical, demographic, and imaging data. Existing ML techniques use limited datasets, which can lead to suboptimal accuracy. The need for a robust, data-driven diagnostic tool using datasets of different formats is evident, particularly to enhance early and timely detection and to develop better patient management strategies.

This study used diverse datasets in different formats from Zenodo [22], Kaggle [23], and the UCI Machine Learning Repository [20]. The Zenodo dataset is in CSV format and containes the following features: id, age, FT3, FT4, TSH, TPO, TGAb, site (the nodule location), echo_pattern (thyroid echogenicity), multifocality (if multiple nodules exist in one location), size (the nodule size in cm), shape (the nodule shape), margin (the clarity of the nodule margin), calcification, echo_strength (the nodule echogenicity), blood_flow (the nodule blood flow), composition (nodule composition), multilateral (if nodules occur in more than one location), and mal (the nodule malignancy).). The Kaggle dataset is in XML format, containing the features: case_number, age, sex, composition, echogenicity, margins, calcifications, tirads, and polygons, along with SVG image coordinates. The thyroid cancer dataset in [20] is in CSV format and contains a class column and 5 unnamed features, which were labeled after carrying out extensive EDA on the dataset. Using a diverse dataset that includes multiple CSV and XML files containing relevant clinical features, this study aimed to generate a comprehensive diagnostic tool capable of analyzing various aspects of patient health to ensure robust and reliable diagnosis, capable of generalizing across different patients. This initiative seeks to enhance the precision and speed of thyroid cancer diagnosis to facilitate prompt and accurate therapeutic measures and improve patient outcomes. The key objective of this study was to create an ML system capable of predicting thyroid cancer, focusing on:

- Data Integration: The dataset used comprises multiple CSV files and an XML file, which include clinical features and patient demographics. The data were preprocessed, including handling missing values using RF imputation, converting categorical variables into numerical formats, and normalization.
- EDA, Preprocessing and Feature Selection: Implementing rigorous preprocessing steps, including missing data handling, normalization, and selecting relevant features to ensure high-quality input for model training. EDA was performed to investigate the connections between the target variable and different features. This step involved generating statistical summaries and visualizing data distributions. Feature selection was performed using *p*-value filters to ensure that the most important features were included for model training. A correlation matrix was used to identify significant relationships, and p-value analysis was employed to find the features that were most crucial to the model. This process ensured the use of only the most impactful predictors, enhancing the accuracy and efficiency of the model.
- Model Development: Using an RF algorithm to generate a predictive model for the accurate classification of thyroid cancer cases.
- Model Evaluation: Performance metrics such as accuracy, precision, recall, and F1-score were used to evaluate the model's performance. Cross-validation was applied to ensure robustness and generalizability.

Figure 1 shows a heat map that illustrates the relationships among features, with color intensity indicating the strength of the correlation. This visual representation helps in identifying which features are highly correlated and potentially redundant.

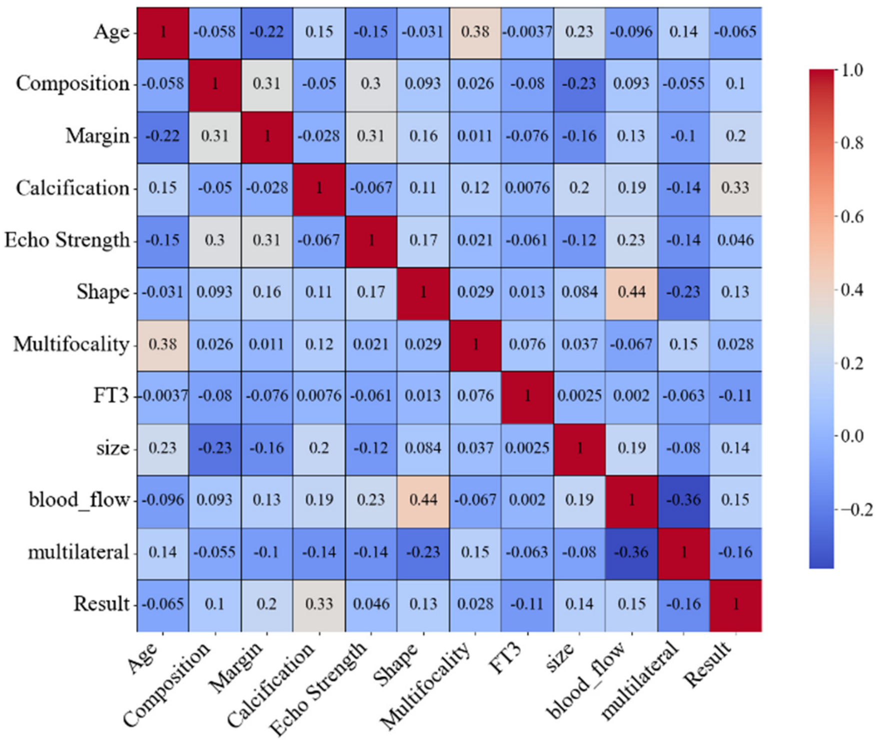

Fig. 1. Correlation matrix.

Following the correlation analysis, the distribution of each feature was explored regarding the target variable, "Result". As seen in Figure 2, features with higher importance scores are more influential in predicting the target variable. Feature selection was performed using *p*-value filters to ensure that the most important features were included for model training.

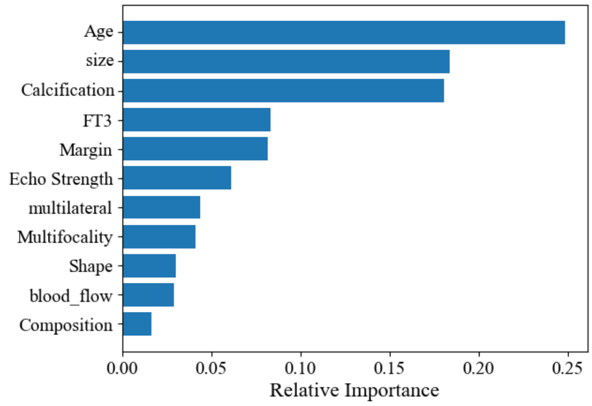

Fig. 2. Feature importance graph.

A. *Model Development and Evaluation*

Various ML techniques were explored, namely, LR, SVM with linear kernel, Gradient Boosting (GBM), and RF. The models were compared using Leave-One-Out Cross-Validation (LOOCV) and stratified k-fold cross-validation strategies to select the best model suitable for accurate classification of patient data as benign or malignant. The models were compared using the Root Mean Square Error (RMSE), finding that LR had 0.14, SVM with linear kernel had 0.16, GBM had 0.17, and RF had 0.12. Thus, the RF model was selected due to its robustness, high performance, and lower RMSE. The model was trained using cross-validation techniques to ensure its generalization capabilities. Performance measures such as accuracy, sensitivity, and specificity were employed to evaluate the model against independent test datasets. This rigorous evaluation process ensured the model's reliability and effectiveness in detecting thyroid cancer.

Figure 3 shows detailed statistics for each class. The model using LOOCV achieved an accuracy of 89.29%, while using stratified k-fold cross-validation achieved 80.65%. The results show a precision of 0.60 and a recall of 0.54 for Class 0, and a precision of 0.86 and a recall of 0.89 for Class 1.

```
Leave-One-Out Cross Validation Accuracy: 0.8929
Stratified K-Fold Cross Validation Accuracy: 0.8065
Confusion Matrix:
[[ 55  47]
 [ 36 291]]
Classification Report:
              precision    recall  f1-score   support

           0       0.60      0.54      0.57       102
           1       0.86      0.89      0.88       327

    accuracy                           0.81       429
   macro avg       0.73      0.71      0.72       429
weighted avg       0.80      0.81      0.80       429
```

Fig. 3. Classification report.

In Figure 4, it can be observed that the model has strong performance, as shown by its substantial quantity of correct positive and negative predictions. However, the model produces certain incorrect positive and negative classifications that require examination to enhance its precision and minimize errors.

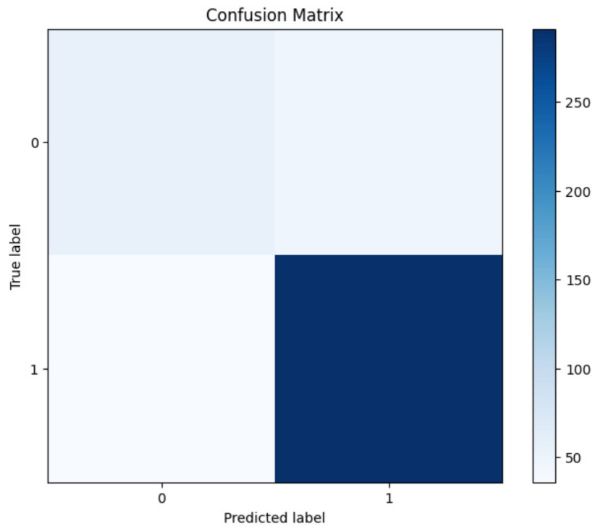

Fig. 4. Confusion matrix.

Figure 5 shows the ROC curve and the AUC/ROC score of 0.81, indicating the model's ability to effectively differentiate between positive and negative classes.

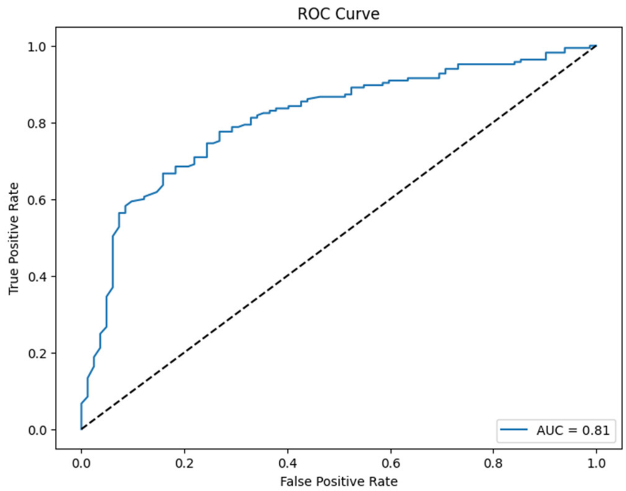

Fig. 5. ROC curve and AUC/ROC score.

B. *Front-end Implementation*

To improve early thyroid cancer identification and treatment planning for better patient care outcomes, a responsive Flask-based web application was developed, providing physicians with immediate diagnostic predictions through an intuitive and user-friendly interface that does not require technical expertise.

III. RESULTS AND DISCUSSIONS

The proposed comprehensive approach involved data preprocessing, feature selection, model training, and deployment to generate a robust and reliable thyroid cancer detection system. This method aimed to set an example for future research and development in medical diagnostics, with possible applications to other disease areas. The combined dataset, along with the performance evaluation of the ML models, yielded significant insights. The results highlighted the predictive capabilities of the RF model and the relevance of various attributes in the dataset. The model's accuracy was evaluated using stratified k-fold cross-validation and LOOCV, achieving 0.8065 and 0.8929, respectively, reflecting a more thorough assessment that is particularly useful for smaller datasets. A correlation matrix was employed to examine the relationships among the features, helping to identify highly correlated and potentially redundant variables. The feature distribution analysis, represented by stacked bar charts, revealed notable patterns. For example, the age distribution showed peaks for benign and malignant cases around ages 35-50, with a higher frequency of malignant cases in most age groups. Correlation values closer to 1 were predominantly associated with malignant cases, indicating a strong correlation with malignancy. Lower margin values were more frequently associated with benign cases.

A comparison of ML models was based on initial accuracy findings from previous studies. RF showed a significant enhancement in performance, reaching an accuracy of 0.8929 with LOOCV, compared to 0.7931 in a previous study. This improvement underscores the effectiveness of the feature selection mechanism and dataset collection processes. The usage of visualizations, such as heatmaps and distribution plots, facilitated a deeper perception of the data complexity and the model's effectiveness. These visual tools were crucial in interpreting how different features influence the prediction outcomes, boosting the interpretability of the model's results.

Table I shows a comparison of the results of this study with previous ones. The accuracy of 89.29% in this study is quite comparable with previous works and even better than many of them. It should be noted that the Bagged CART model [4] achieved an accuracy of 99.1% which is very high. However, all other works used homogeneous data formats (CSV, XML, or images), whereas this one used heterogeneous formats (combining CSV and XML) with good accuracy.

TABLE I. COMPARISON WITH PREVIOUS STUDIES

Ref.	Dataset - Format	Model/Method	Accuracy (%)
[4]	Zenodo - CSV	Bagged CART	99.1
[5]	Zenodo - CSV	RF	79.3
[6]	UCI Repository -CSV	Bi-directional RNN	98.72
[7]	DDTI - XML	EfficientNetB0-B6	82.3
[8]	Open-CAS - CSV	SVM, ANN	88.7, 89.4
[10]	Thyroid Images from Universidad Nacional de Colombia	ANN, SVM	75.0, 96.0
[15]	Thyroid Images from Tianjin Cancer Hospital	DCNN	Specificity: 86.9
[19]	UCI Repository - CSV	CNN, MLFS	89.0, 99.0
This study	Zenodo, Kaggle, UCI - CSV and XML	RF	89.29

IV. CONCLUSION

This study developed a predictive model for thyroid cancer using ML techniques, specifically an RF algorithm, after a comparative analysis with LR, SVM with linear kernel, and GBM based on RMSE, using thyroid cancer datasets from Zenodo [22], Kaggle [23], and the UCI Machine Learning Repository [20]. The proposed model demonstrated strong performance metrics, including a peak accuracy of 89.29% with LOOCV. This high level of accuracy underscores the potential of ML in improving thyroid cancer diagnosis. Overall, the accuracy of the proposed model is in the order of existing works, making it a suitable candidate for thyroid cancer prediction.

The key contributions of this study include the integration of multiple data sources in CSV and XML formats. Another important contribution is the rigorous preprocessing steps and the feature selection technique, which takes into account the heterogeneous nature of the input data. The feature selection method combines the data into a single uniform dataset, which is used to train and test the ML models. Another contribution is the development of a user-friendly Flask-based frontend for real-time predictions. Together, these components improve the diagnostic process, making it more accurate and accessible to healthcare professionals.

Future efforts will be directed toward further optimizing the model, exploring additional ML techniques, and expanding the dataset to incorporate more diverse patient populations. Ultimately, by continuously innovating and improving these methods, the objective is to enhance the detection and treatment of thyroid cancer, resulting in better patient outcomes.

REFERENCES

[1] Z. Lyu, Y. Zhang, C. Sheng, Y. Huang, Q. Zhang, and K. Chen, "Global burden of thyroid cancer in 2022: Incidence and mortality estimates from GLOBOCAN," *Chinese Medical Journal*, vol. 137, no. 21, pp. 2567–2576, Nov. 2024, https://doi.org/10.1097/CM9.0000000000003284.

[2] S. Hu, X. Wu, and H. Jiang, "Trends and projections of the global burden of thyroid cancer from 1990 to 2030," *Journal of Global Health*, vol. 14, 2024, Art. no. 04084, https://doi.org/10.7189/jogh.14.04084.

[3] D. F. Sigmon and S. Fatima, "Fine Needle Aspiration," in *StatPearls*, Treasure Island, FL, USA: StatPearls Publishing, 2025.

[4] İ. B. Çiçek and Z. Küçükakçalı, "Machine Learning Approach for Thyroid Cancer Diagnosis Using Clinical Data," *Middle Black Sea Journal of Health Science*, vol. 9, no. 3, pp. 440–452, Aug. 2023, https://doi.org/10.19127/mbsjohs.1282265.

[5] N. M. Xi, L. Wang, and C. Yang, "Improving the diagnosis of thyroid cancer by machine learning and clinical data," *Scientific Reports*, vol. 12, no. 1, Jul. 2022, Art. no. 11143, https://doi.org/10.1038/s41598-022-15342-z.

[6] M. A. Begum, I. M. Tresa, S. Sandhya, S. Vidhya, and G. Vinodhini, "Machine learning based dysfunction thyroid cancer detection with optimal analysis," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 7, pp. 818–823, 2021.

[7] S. A. Nasr, H. M. Abdel-Fattah, M. M. Abdelsalam, and H. El-Din Moustafa, "An Accurate Deep Learning Based Framework for Detection of Thyroid Cancer Using Ultrasound Images," *International Journal of Chemical and Biochemical Sciences*, vol. 24, no. 12, pp. 455–468, 2023.

[8] P. Poudel, A. Illanes, E. J. G. Ataide, N. Esmaeili, S. Balakrishnan, and M. Friebe, "Thyroid Ultrasound Texture Classification Using Autoregressive Features in Conjunction With Machine Learning

Approaches," *IEEE Access*, vol. 7, pp. 79354–79365, 2019, https://doi.org/10.1109/ACCESS.2019.2923547.

[9] K. E. Setiawan, "Predicting recurrence in differentiated thyroid cancer: a comparative analysis of various machine learning models including ensemble methods with chi-squared feature selection," *Communications in Mathematical Biology and Neuroscience*, vol. 2024, Apr. 2024, Art. no. 55, https://doi.org/10.28919/cmbn/8506.

[10] V. V. Vadhiraj, A. Simpkin, J. O'Connell, N. Singh Ospina, S. Maraka, and D. T. O'Keeffe, "Ultrasound Image Classification of Thyroid Nodules Using Machine Learning Techniques," *Medicina*, vol. 57, no. 6, Jun. 2021, Art. no. 527, https://doi.org/10.3390/medicina57060527.

[11] J. Gu *et al.*, "A machine learning-based approach to predicting the malignant and metastasis of thyroid cancer," *Frontiers in Oncology*, vol. 12, Dec. 2022, https://doi.org/10.3389/fonc.2022.938292.

[12] Y. Habchi *et al.*, "AI in Thyroid Cancer Diagnosis: Techniques, Trends, and Future Directions," *Systems*, vol. 11, no. 10, Oct. 2023, Art. no. 519, https://doi.org/10.3390/systems11100519.

[13] S. R. Shih *et al.*, "Computerized Cytological Features for Papillary Thyroid Cancer Diagnosis—Preliminary Report," *Cancers*, vol. 11, no. 11, Nov. 2019, Art. no. 1645, https://doi.org/10.3390/cancers11111645.

[14] H. Wang *et al.*, "Development and validation of prediction models for papillary thyroid cancer structural recurrence using machine learning approaches," *BMC Cancer*, vol. 24, no. 1, Apr. 2024, Art. no. 427, https://doi.org/10.1186/s12885-024-12146-4.

[15] X. Li *et al.*, "Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study," *The Lancet Oncology*, vol. 20, no. 2, pp. 193–201, Feb. 2019, https://doi.org/10.1016/S1470-2045(18)30762-9.

[16] M. Mourad *et al.*, "Machine Learning and Feature Selection Applied to SEER Data to Reliably Assess Thyroid Cancer Prognosis," *Scientific Reports*, vol. 10, no. 1, Mar. 2020, Art. no. 5176, https://doi.org/10.1038/s41598-020-62023-w.

[17] B. Zhang *et al.*, "Machine Learning–Assisted System for Thyroid Nodule Diagnosis," *Thyroid®*, vol. 29, no. 6, pp. 858–867, Jun. 2019, https://doi.org/10.1089/thy.2018.0380.

[18] S. Anari, N. Tataei Sarshar, N. Mahjoori, S. Dorosti, and A. Rezaie, "Review of Deep Learning Approaches for Thyroid Cancer Diagnosis," *Mathematical Problems in Engineering*, vol. 2022, no. 1, 2022, Art. no. 5052435, https://doi.org/10.1155/2022/5052435.

[19] R. Chaganti, F. Rustam, I. De La Torre Díez, J. L. V. Mazón, C. L. Rodríguez, and I. Ashraf, "Thyroid Disease Prediction Using Selective Features and Machine Learning Techniques," *Cancers*, vol. 14, no. 16, Jan. 2022, Art. no. 3914, https://doi.org/10.3390/cancers14163914.

[20] R. Quinlan, "Thyroid Disease." UCI Machine Learning Repository, 1986, https://doi.org/10.24432/C5D010.

[21] I. O. Lixandru-Petre *et al.*, "Machine Learning for Thyroid Cancer Detection, Presence of Metastasis, and Recurrence Predictions—A Scoping Review," *Cancers*, vol. 17, no. 8, Jan. 2025, Art. no. 1308, https://doi.org/10.3390/cancers17081308.

[22] N. M. Xi, L. Wang, and Chuanjia Yang, "Improving The Diagnosis of Thyroid Cancer by Machine Learning and Clinical Data." Zenodo, Apr. 16, 2022, https://doi.org/10.5281/ZENODO.6465436.

[23] "DDTI: Thyroid Ultrasound Images." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/dasmehdixtr/ddti-thyroid-ultrasound-images.