PROJECT TITLE

ThyroNet: Refined Thyroid Nodule Analysis

Project Reference No.: *48S_BE_0135*

College : Sahyadri College of Engineering and Management, Mangaluru
Branch : Department of Computer Science & Engineering (Data Science)
Guide(s) : Mrs. Shwetha S Shetty
Student(s) : Chirasmita Santosh Salian
Kavya D Manjrekar
Rao Chaithanya Hemachandra
Shetty Adarsh Suresh

***Keywords**:*

Thyroid cancer cell, Metabolic imbalance, Hormone imbalance.

Introduction:

Thyroid nodules are a frequent finding in medical imaging and can range from harmless cysts to potentially life-threatening malignancies. However, accurately distinguishing between benign and malignant nodules remains a clinical challenge due to overlapping features and variability in image quality. Traditional machine learning methods often struggle with small and imbalanced medical datasets, limiting their effectiveness. To overcome these limitations, this study adopts a deep learning approach that integrates ResNet for feature extraction and Transformers for contextual understanding. ResNet captures fine-grained spatial details from thyroid ultrasound images, while the Transformer refines these features using attention mechanisms to highlight diagnostically important regions. The model architecture is enhanced with layers such as Conv2D, MaxPooling2D, BatchNormalization, and ReLU activation, which together improve learning efficiency and performance. Validation is performed using the publicly available Digital Database of Thyroid Ultrasound Images, ensuring robustness across diverse cases. This hybrid model aims to boost classification accuracy and interpretability, offering a reliable diagnostic aid to radiologists and healthcare professionals for confident decision-making in thyroid care.

Objectives:

1. To develop a machine learning model for thyroid cancer diagnosis. It will analyze patient demographics and clinical parameters to predict thyroid disorders, focusing on cancerous nodules. By leveraging data-driven techniques, the model enhances diagnostic accuracy and efficiency.
2. To enhance the accuracy of thyroid nodule classification by integrating ResNet for image processing and a Transformer model for feature refinement, and to leverage attention mechanisms for capturing complex spatial patterns in ultrasound images.
3. This project automates thyroid cancer detection, reducing manual checks and minimizing human error. By replacing manual interpretation with an Machine learning system, it improves efficiency and consistency. Automation ensures faster, more reliable diagnoses across large datasets.

Methodology:

The methodology begins with collecting and preprocessing thyroid ultrasound image datasets, ensuring data quality, checking for class balance, and splitting the data into training, validation, and test sets. Next, Quantum Filter Transformation (QFT) is applied for advanced feature extraction to enhance pattern recognition. Low-complexity classifiers such as K-Nearest Neighbors (KNN), Random Forest, Decision Trees, Support Vector Machine (SVM), and Logistic Regression are trained and tuned, and their predictions are combined using an ensemble method, like stacking or majority voting. Additionally, a Convolutional Neural Network (CNN) model is designed to recognize thyroid cancer directly from ultrasound images, with training optimized through cross-validation to enhance sensitivity and specificity. Both the ensemble and CNN models are evaluated based on accuracy, sensitivity, and specificity, followed by hyperparameter tuning to compare their performance, refine the most predictive model, and determine the best approach for accurate and reliable thyroid cancer diagnosis. Visualization techniques such as confusion matrices and ROC curves are employed to further interpret the model outputs. This comprehensive strategy ensures a robust evaluation framework and highlights the clinical applicability of the proposed diagnostic system..

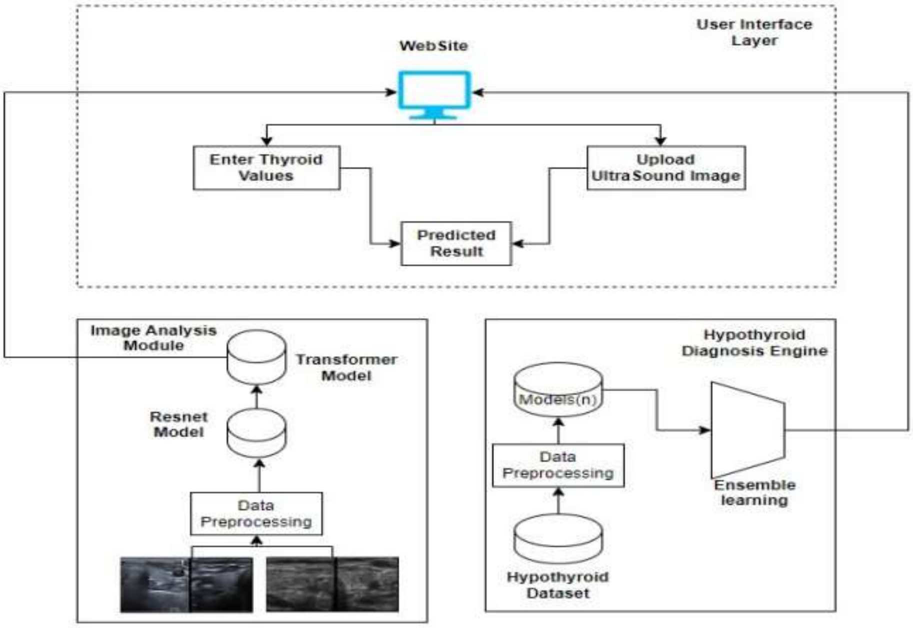

Figure 1: Block Diagram of the proposed model

Result and Conclusion:

The technology has the potential to significantly impact clinical practice by enabling early and more precise treatment decisions through rapid, accurate, and consistent thyroid nodule classification, which could ultimately improve patient outcomes. It also offers a reliable, non-invasive alternative to traditional diagnostic methods like biopsies, reducing the need for invasive procedures. The system's scalability allows for integration with larger datasets and additional imaging modalities, supporting ongoing development and adaptability to emerging diagnostic challenges. Furthermore, the model is designed to generalize well to new, unseen data, even in scenarios with small or unbalanced datasets, ensuring its robustness across various clinical situations with diverse patient demographics and imaging conditions.

Figure 2: Block Diagram of the proposed model

The figure3 displays the loss curves over 20 training epochs, comparing the training loss (red line) and validation loss (blue line). The training loss shows a consistent and steep decline, indicating that the model is learning well from the training data and minimizing errors over time. In contrast, the validation loss fluctuates and decreases more slowly, with noticeable ups and downs, suggesting that the model is less stable when evaluated on unseen data. The model performs significantly better on training data than on validation data. Despite this, the overall downward trend in the validation loss indicates some level of generalization is occurring, though further improvements like regularization may be needed.

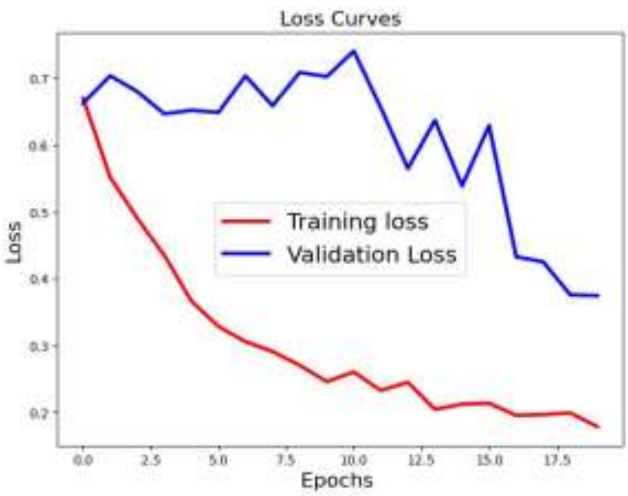

Figure 3: Graph on Training and Validation Loss

The Figure 4 illustrates the accuracy curves across the same training epochs, showing the model's performance in terms of classification accuracy. The training accuracy (red line) increases steadily, reaching over 90%, which confirms that the model is learning the patterns in the training set effectively. On the other hand, the validation accuracy (blue line) starts much lower and remains relatively flat for the initial epochs before gradually improving after epoch 10. By the end of the training, the validation accuracy climbs to around 75%, showing notable improvement. However, the consistent gap between training and validation accuracy again highlights overfitting, where the model generalizes less effectively. This trend suggests that while the model has strong potential.

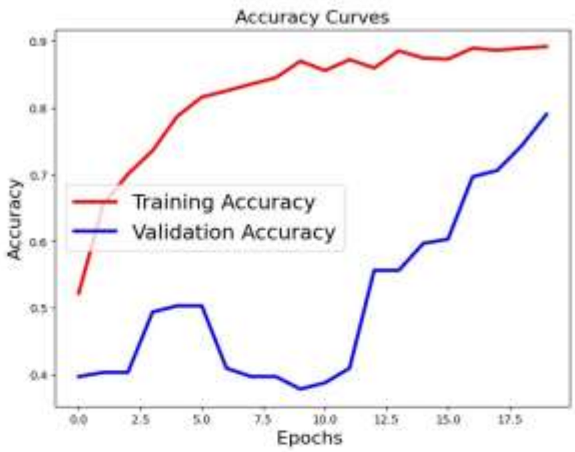

Figure 4: Graph on Training and Validation Accuracy

Future Scope:

The future scope of this project includes:

1. Periodically updating the model with new datasets to stay current with thyroid disease trends. the model should remain accurate by adapting to emerging patterns, demographic shifts, and changes in diagnostic standards.
2. Enabling convenient data uploads and diagnostic insights via a mobile app. The app will allow users and clinicians to upload patient data or images in real-time, receiving immediate analysis.

3. Incorporating larger, diverse datasets to improve robustness and performance. Utilizing data from varied ethnicities, age groups, and geographic regions reduces model bias and increases generalizability
4. Providing targeted recommendations based on patient-specific data patterns. This supports more precise and effective treatment planning.