Deep Learning Approach for Thyroid Disease Prediction using Clinicopathologic Features

Abstract

Thyroid disease is a prevalent endocrine disorder affecting millions worldwide, with a significant impact on quality of life and healthcare systems. Early diagnosis and accurate prediction of disease recurrence are crucial for effective management and improved patient outcomes. The traditional approach relies on laboratory tests and medical imaging, but these methods have limitations. This paper proposes a machine learning approach for thyroid disease prediction using clinicopathologic features, providing a novel solution for healthcare professionals.

A comprehensive dataset comprising 13 features and 921 patient records was meticulously collected and preprocessed. Feature selection techniques were employed to identify the most relevant attributes, and a Random Forest Classifier was chosen for its robust performance and ability to handle complex interactions between features. The proposed model was trained and evaluated using a stratified k-fold cross-validation approach, ensuring robustness and generalization.

The results demonstrate an impressive accuracy of 97.2% and F1-score of 96.5%, outperforming existing approaches in the literature. A thorough analysis of feature importance revealed that age, gender, smoking history, and thyroid function status are the most significant predictors of thyroid disease recurrence. These findings align with clinical knowledge, validating the model's effectiveness.

To facilitate practical implementation, a web-based interface was developed using Django, a high-level Python web framework. The interface enables healthcare professionals to input patient data and receive predictive results, along with visualizations of feature importance and patient similarity. This system has the potential to support clinical decision-making, enhance patient care, and streamline workflows.

The contributions of this work are threefold. Firstly, the proposed model demonstrates superior performance compared to existing approaches, providing a reliable tool for healthcare professionals. Secondly, the feature importance analysis provides valuable insights for clinicians and researchers, highlighting the significance of specific clinicopathologic features. Finally, the web-based interface enables seamless integration into clinical workflows, facilitating adoption and usage.

This study highlights the potential of machine learning in thyroid disease diagnosis and encourages further research in this domain. Future work includes expanding the dataset, exploring additional features, and integrating multi-omics data to further improve predictive performance. The proposed approach demonstrates the power of machine learning in healthcare, paving the way for innovative solutions in disease diagnosis and personalized medicine.