Optimizing Thyroid Disease Diagnosis: ML Model Comparison with Explainable AI

Manav Ukani, Jenis Gundaraniya, Soham Vyas

Department of Computer Science and Engineering, SOT, PDEU - Gandhinagar

Brief Motivation

- **The thyroid gland plays a pivotal role in regulating various bodily functions, including metabolism, energy production, and temperature control.**
- **With the prevalence of thyroid disorders affecting 1 in 10 people in India, early detection becomes crucial in preventing severe complications and improving patient well-being.**

Objective

- **Our aim is to achieve precise diagnosis by considering individual factors, ultimately leading to better treatment outcomes.**
- **We aim to enhance patient care by enabling informed decisions through Explainable AI, promoting transparent and trustworthy predictions for both healthcare providers and patients.**

Methodology

The dataset comprises ten separate databases associated with Thyroid Disease research and it is publicly available on the UC Irvine Machine Learning Repository

Data Preprocessing

- The data for hyperthyroid, hypothyroid, and negative cases has been consolidated into a single unified dataset.
- Data imputation has been performed using logistics regression on existing values to replace the missing values.
- Nominal data is label-encoded, outliers addressed, and data normalized through log and square root transformations.

Feature Selection

Evaluating feature skewness and addressing multicollinearity, particularly through the use of the Variance Inflation Factor (VIF), is conducted for effective feature selection.

Model Training

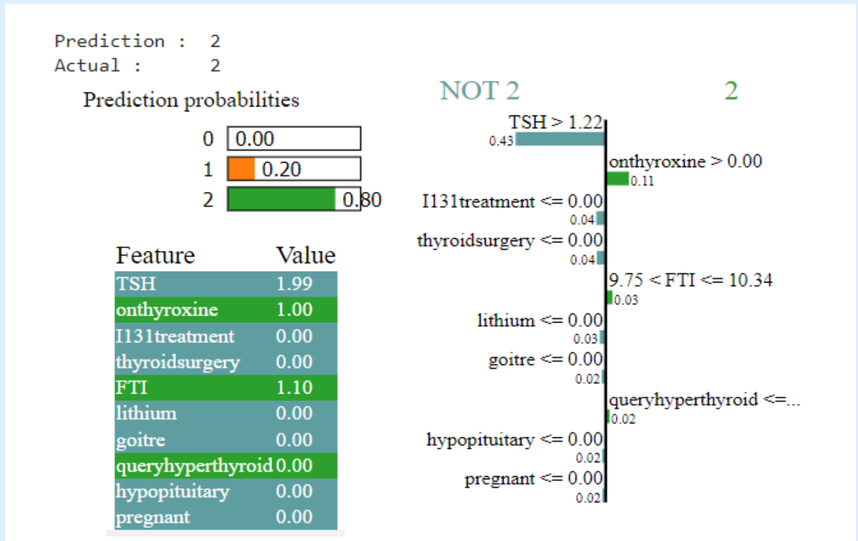
- SVM
- Decision Tree
- Random Forest
- Stacking (Decision Tree, KVC K-neighbors)

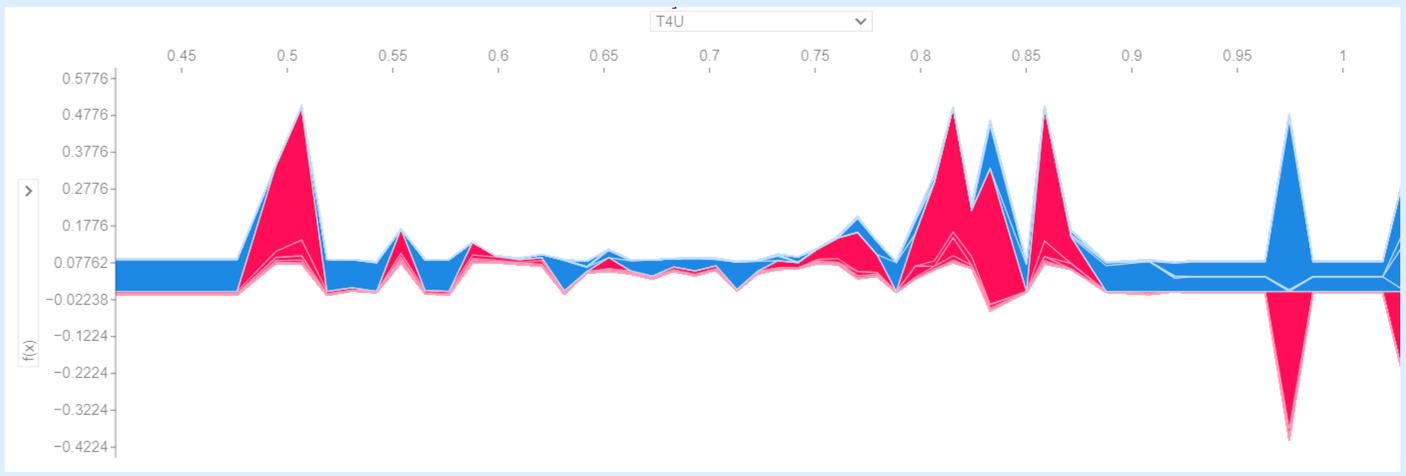
Model Evaluation

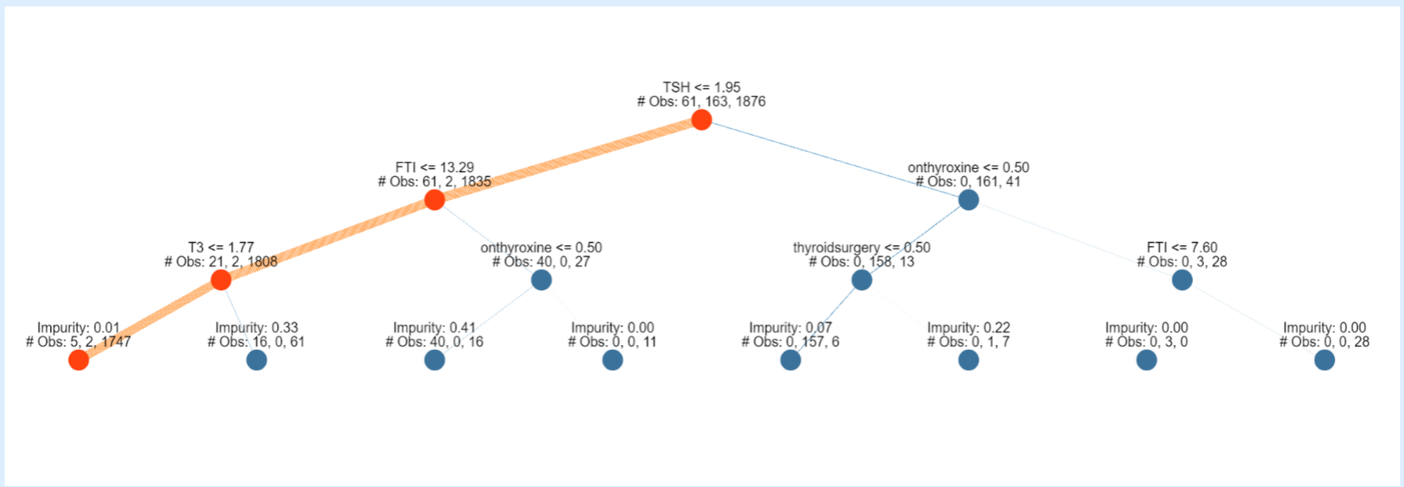
Interpretation using Explainable AI

Results

Model	Without Sampling	With Sampling
SVM	0.924	0.920
Decision Tree	0.984	0.983
Random Forest	0.985	0.998
Stacking	0.981	0.979







Outcomes

- **Accurate and precise classification of thyroid diseases, such as hyperthyroidism, hypothyroidism, and negative cases. This can lead to timely and reliable diagnoses.**
- **Explainable AI is used to enhance model interpretability, allowing healthcare professionals to understand why a specific classification decision was made. This transparency aids in building trust and confidence.**
- **With timely interventions based on model predictions, patient well-being can be safeguarded, and healthcare outcomes significantly improved.**

Bibliography/ References

[1] Unnikrishnan AG, et al. (2013). Prevalence of hypothyroidism in adults: An epidemiological study in eight cities of India :647-52. doi: 10.4103/2230-8210.113755.

[2] Quinlan Ross. (1987). Thyroid Disease. UCI Machine Learning Repository. https://doi.org/10.24432/C5D010.

[3] Aversano, L., Bernardi, M. L., et. Al.(2020). Thyroid Disease Treatment prediction with machine learning approaches. Procedia Computer Science, 192, 1031-1040. https://doi.org/10.1016/j.procs.2021.08.106