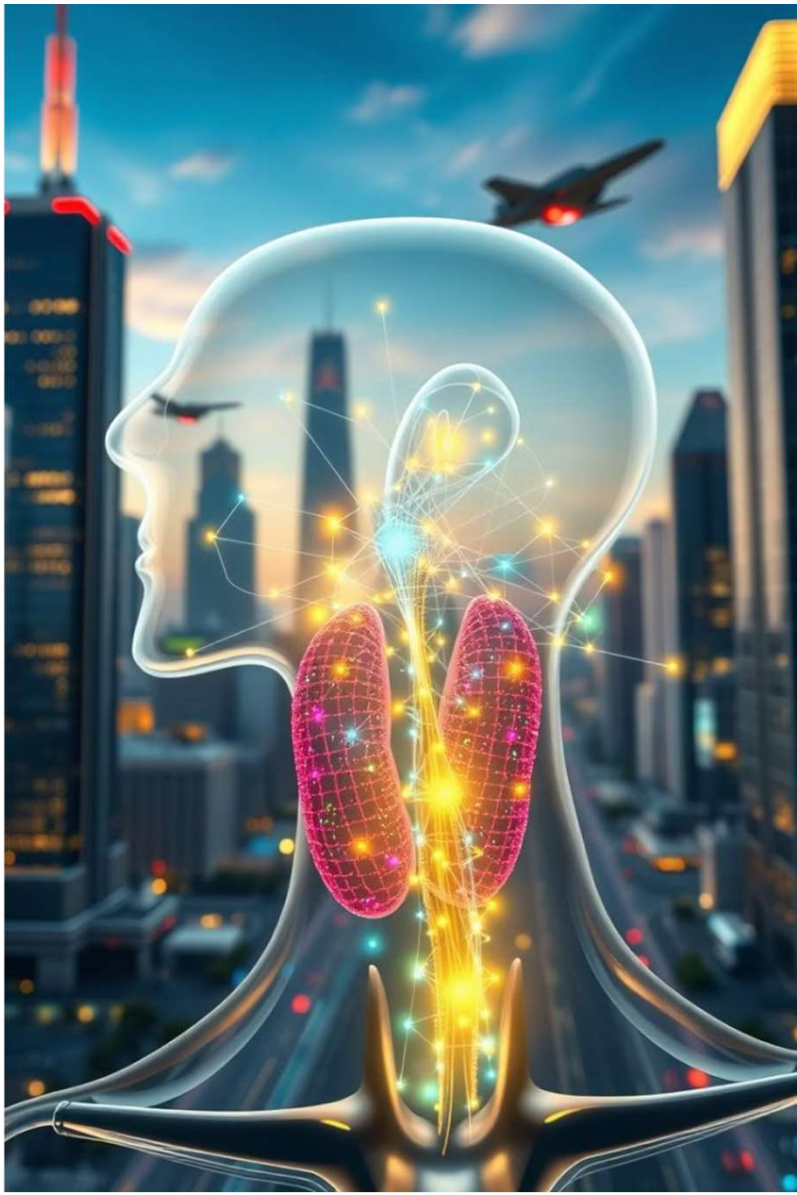
Predicting Thyroid Cancer Recurrence

Leveraging machine learning to improve patient outcomes and personalize care.

by manjiri gajmal



Why Predicting Thyroid Cancer Relapse Matters?????

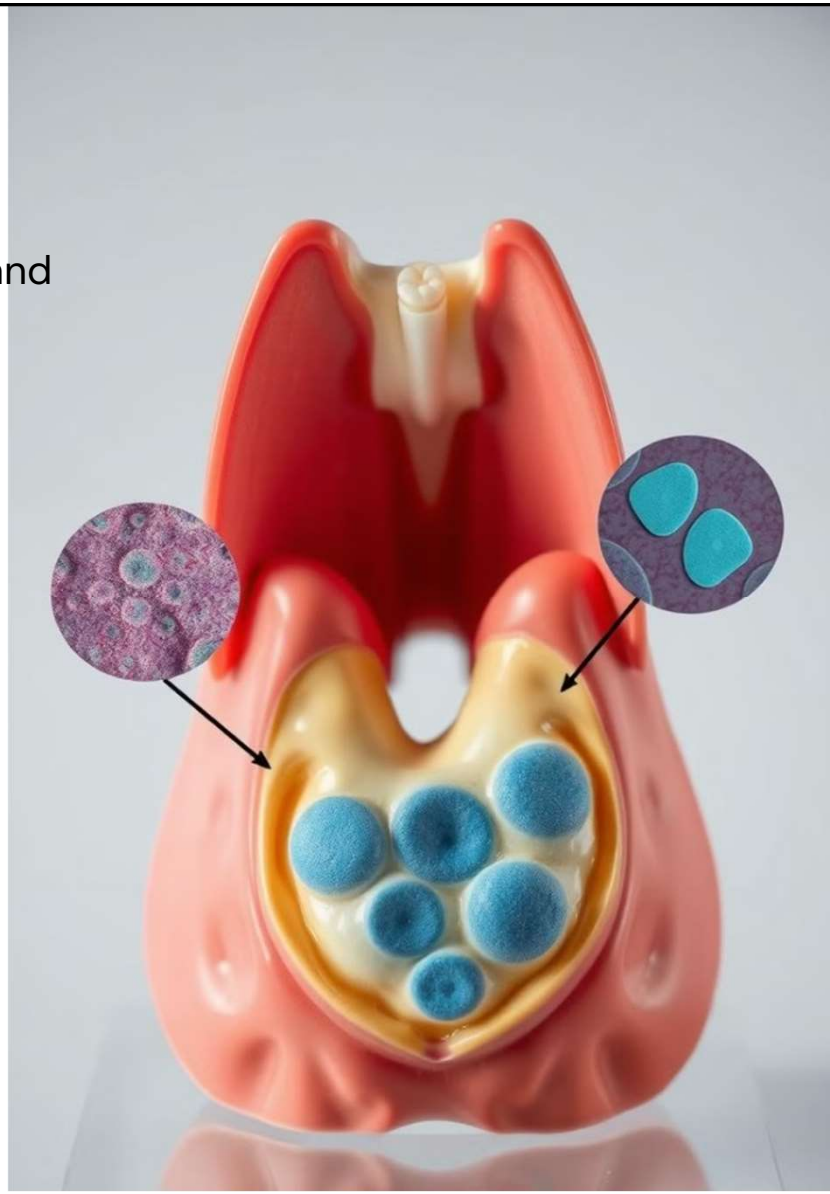
Thyroid cancer survivors are at risk of relapse, requiring timely and accurate predictions to guide interventions.

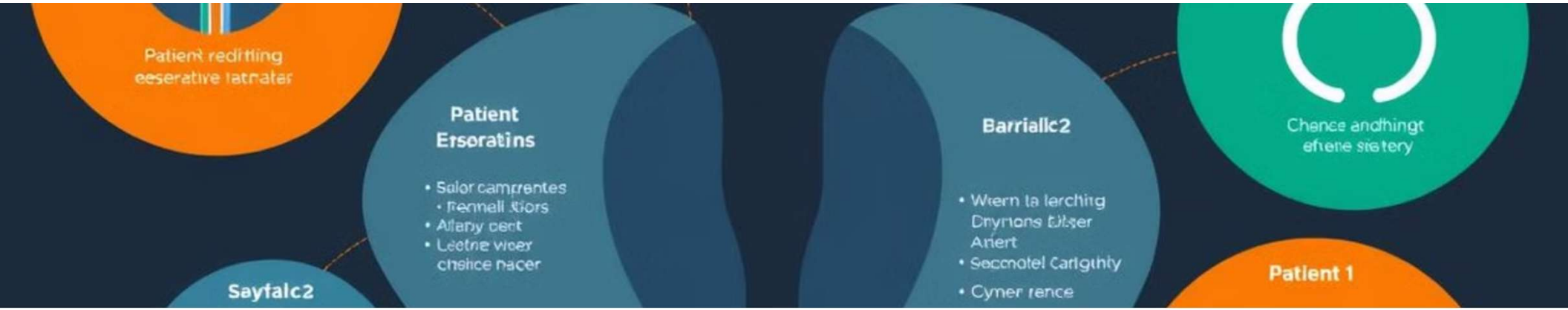
Objective:

Build a machine learning system to predict relapse using patient data.

Importance:

- Minimize false negatives to ensure no relapse is missed.
- Avoid false positives to reduce unnecessary interventions.
- Deliver interpretable and actionable insights for medical professionals.





Dataset Analysis: Understanding the Data

Dataset

Patient health and medical history data.

Key Features

- Age
- Gender
- Thyroid function tests (TSH, T3, T4 levels)
- Treatment history (e.g., radiotherapy, surgery)
- Tumor characteristics

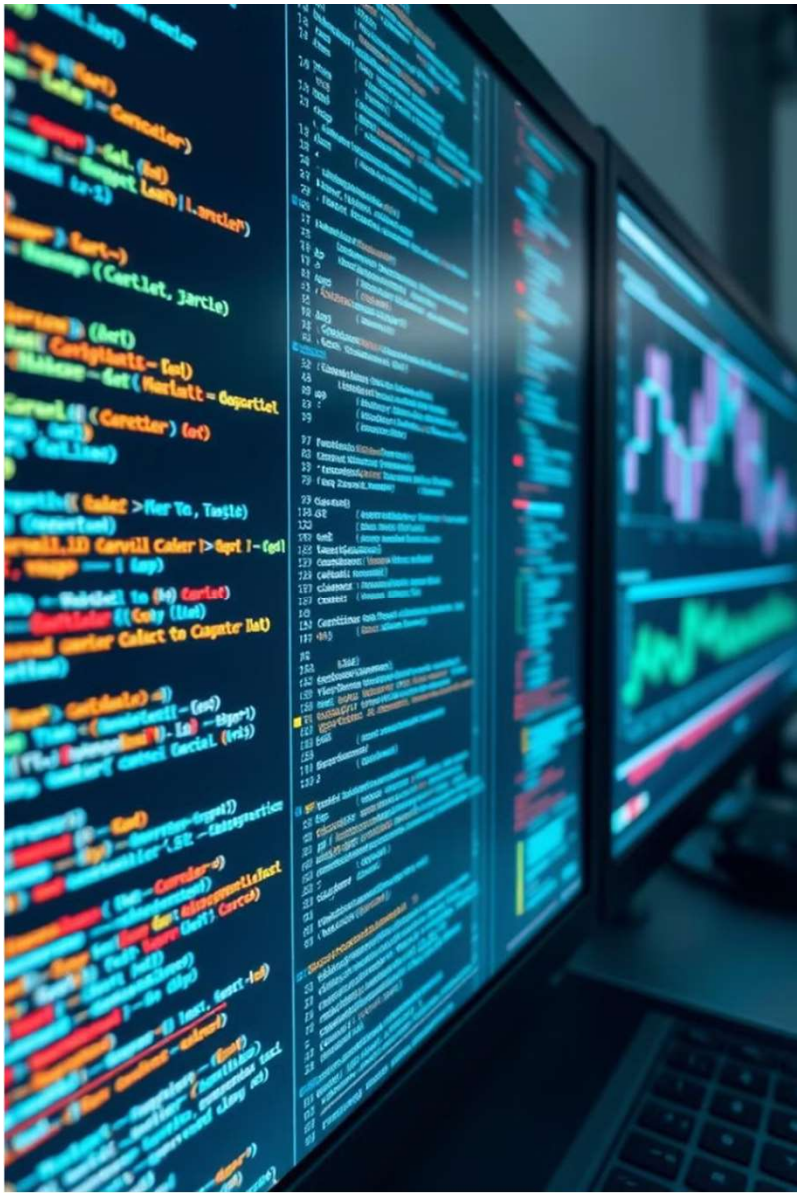
Target

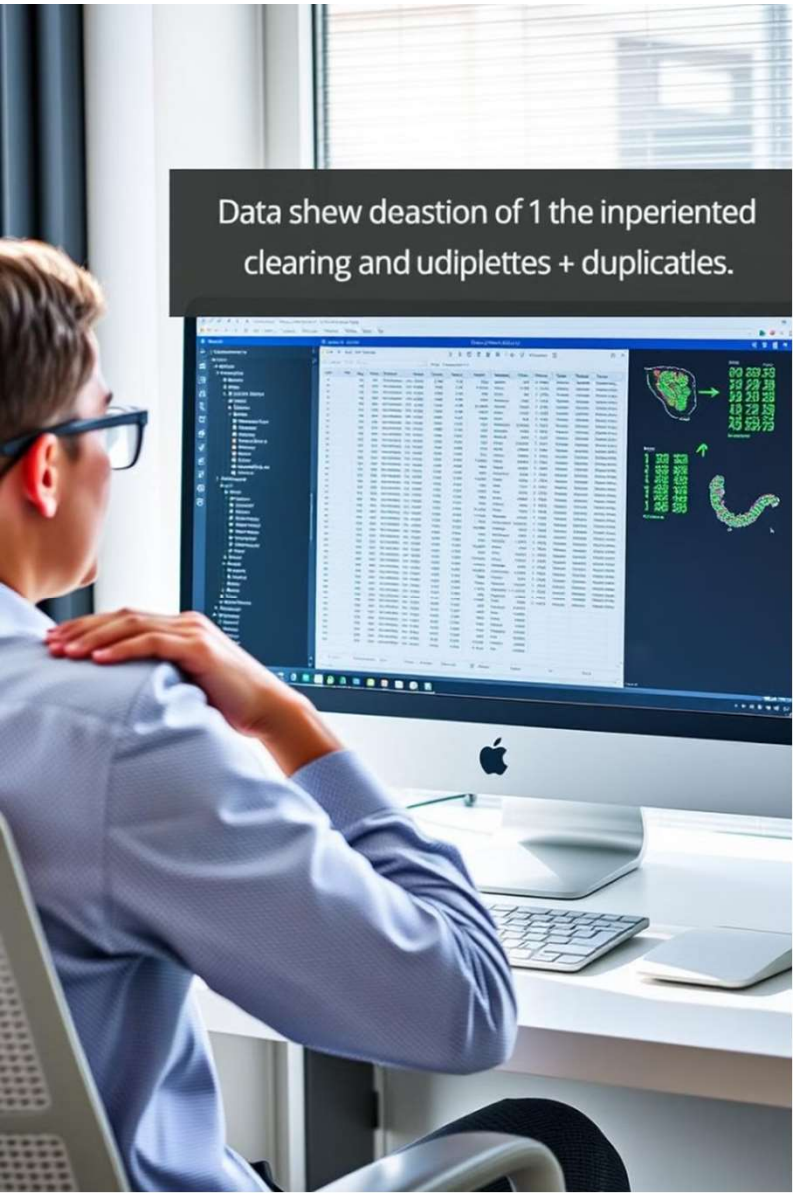
Relapse (Yes/No)

Machine Learning: A Powerful Tool for Prediction

Improve patient outcomes and streamline clinical decision-making.

Identifying subtle patterns indicating potential relapse. Reduce healthcare costs associated with late-stage recurrence treatments.





Data Cleaning: Ensuring Accuracy and Reliability

Duplicate Removal

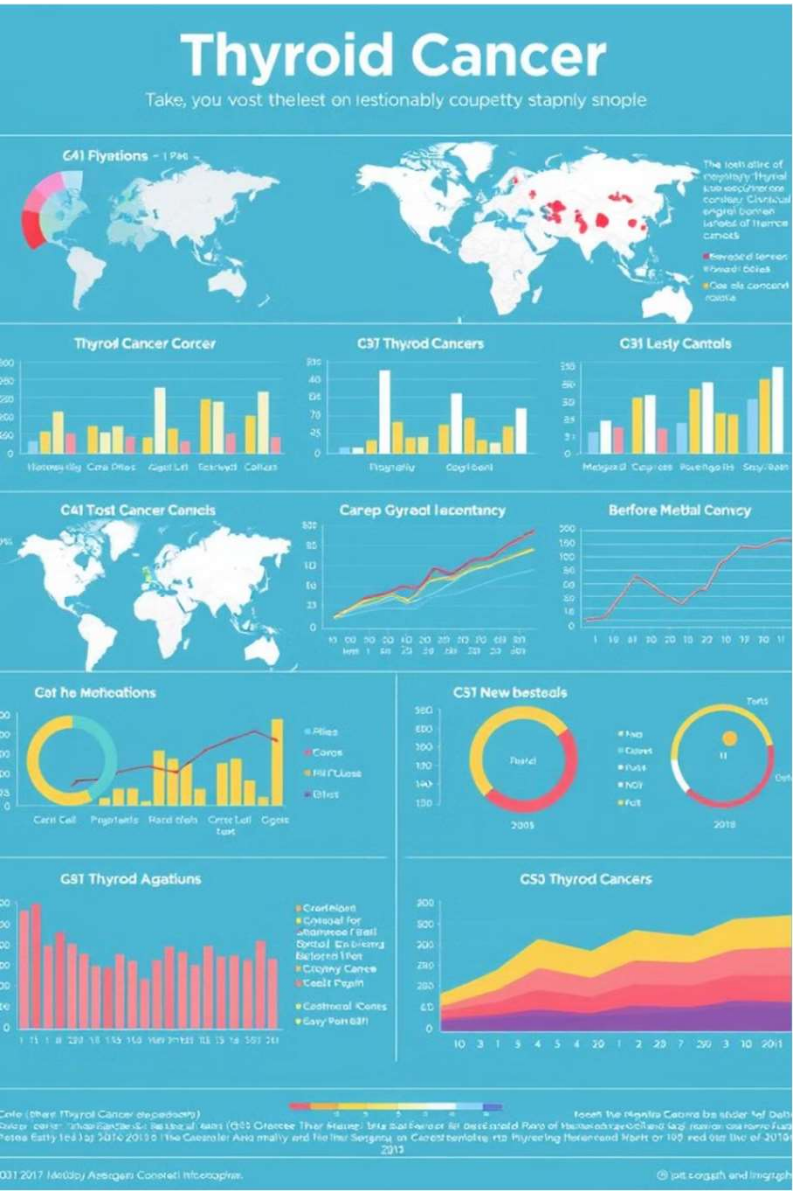
19 duplicate rows identified and removed.

Dataset Integrity

Ensuring data is ready for analysis and model building.

Tool used

Python (Pandas, NumPy, Scikit-learn).



Exploratory Data Analysis: Unveiling Insights

Age Distribution

Most individuals are aged between 30 and 50.

Smoking history

Predominantly non-smokers or no smoking history.

Cancer Staging

Most cases are classified as Stage I.

Thyroid Function

Euthyroid cases dominate, disorders are less common.

Gender Imbalance

Dataset skewed towards females.

Pathology

Papillary carcinoma is the most frequent type.

Machine Learning Models Evaluated

Algorithms Considered

- Random Forest
- Logistic Regression
- XGBoost
- Support Vector Machine (SVM)

Model Evaluation

- Accuracy
- Precision
- Recall
- ROC-AUC
- Confusion Matrix

Model Performance Summary

1

Random Forest

- Most effective model, high accuracy and precision.
- Provides feature importance for interpretability.

2

Logistic Regression

- Excellent interpretability.
- Very low false positives but slightly lower complexity handling.

3

XGBoost

- Scalable and efficient for larger datasets.

4

SVM

- Competitive performance but less interpretable than other models.

Comparing Model ROC Curves

Random Forest:

- AUC: 0.99
- Excellent performance in distinguishing relapse vs. no relapse.

Logistic Regression

- AUC: 0.98
- Slightly behind Random Forest.

XGBoost

- AUC: 0.97
- High performance for scalability.

SVM

- AUC: 0.96
- Effective but lacks transparency.

Which Model Should You Use?

Random Forest:

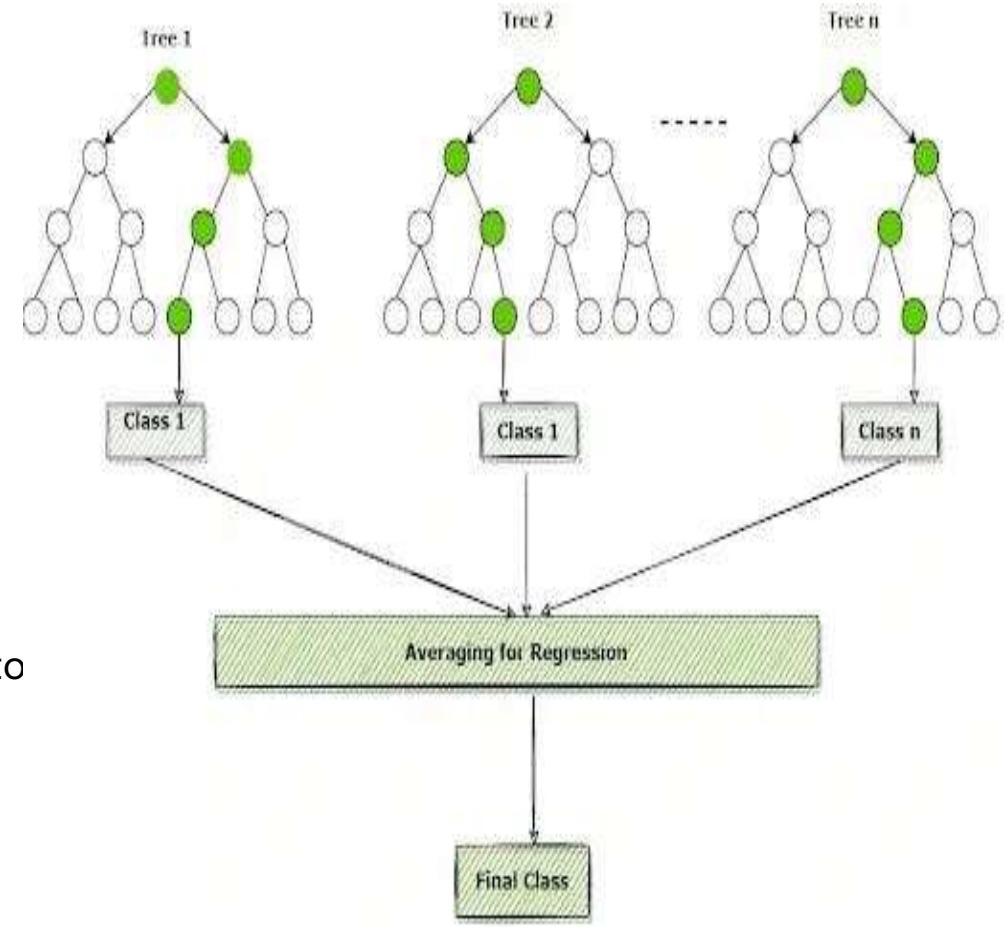
- Best balance of accuracy and interpretability.
- Handles non-linear relationships effectively.

Logistic Regression

- Ideal for high interpretability and simpler datasets.

XGBoost

- Optimal for large-scale systems requiring scalability.



Final Decision:

Random Forest for real-world implementation due to

- Low false negatives.
- Interpretable insights.

Key Advantages

Improved Patient Outcomes:

- Early detection of relapse.
- Reduced mortality rates.

Resource Optimization:

- Targeted interventions.
- Reduced unnecessary procedures.

Medical Trust:

- Transparent predictions with interpretable models.

What's next?

Challenges:

- Handling imbalanced datasets.
- Integrating the system into clinical workflows.

Future Scope:

- Use real-time patient monitoring data.
- Explore ensemble methods for further improvement.
- Ensure compliance with medical regulations and ethical standards.

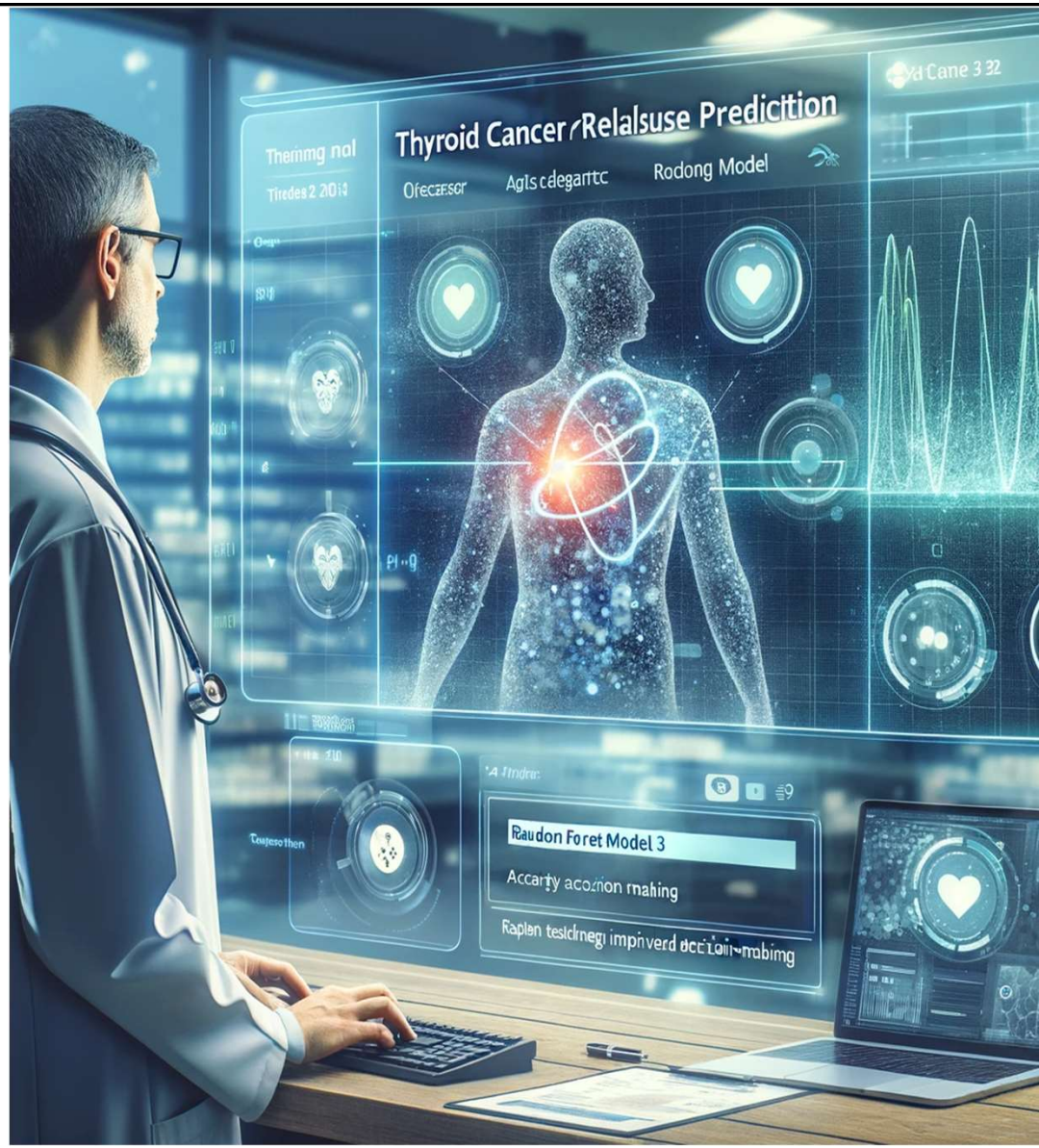
Final Thoughts!!!

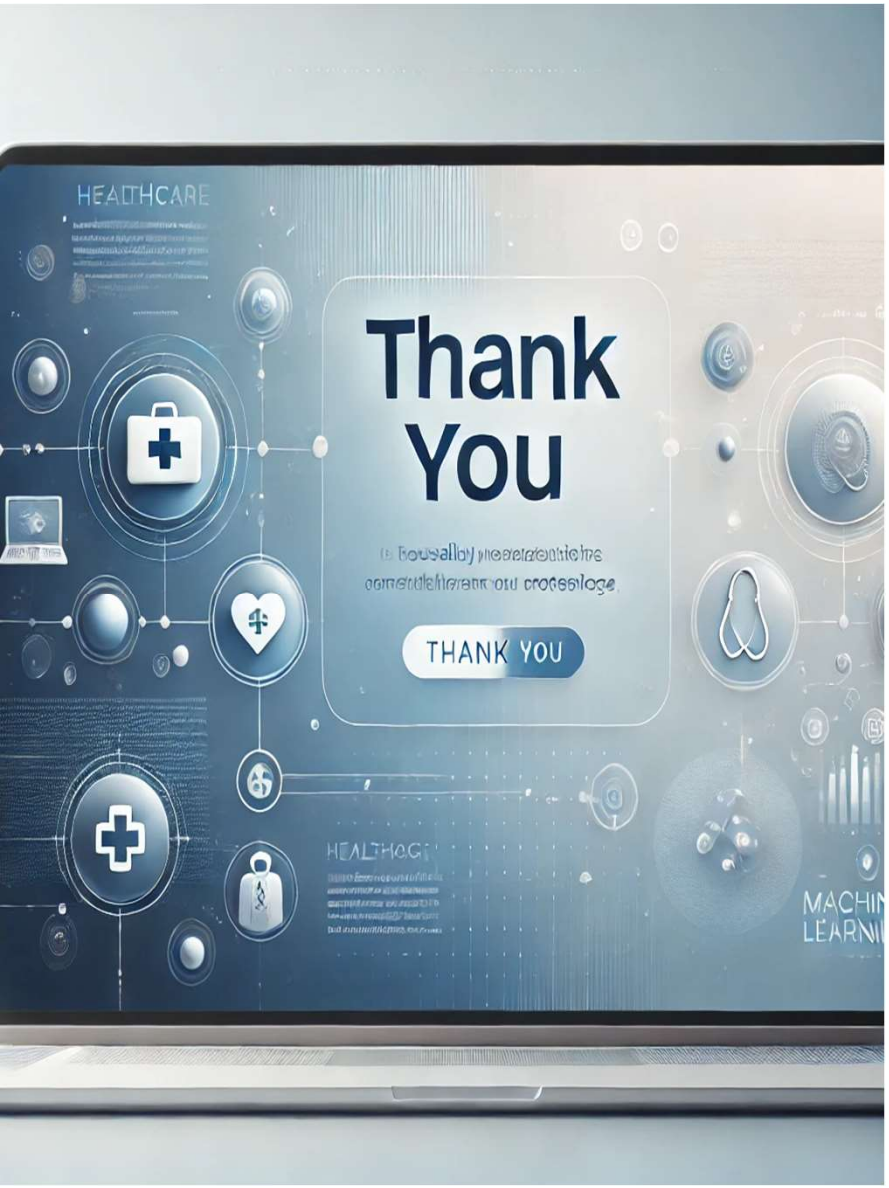
Key Takeaway::

- Random Forest provides the best balance of performance and interpretability for predicting thyroid cancer relapse.

Next Steps:

- Implementation in a clinical setting with periodic evaluation and retraining.





Any Questions?

Thank you for your attention!!!

Feel free to ask about the methodology, results, or future work.