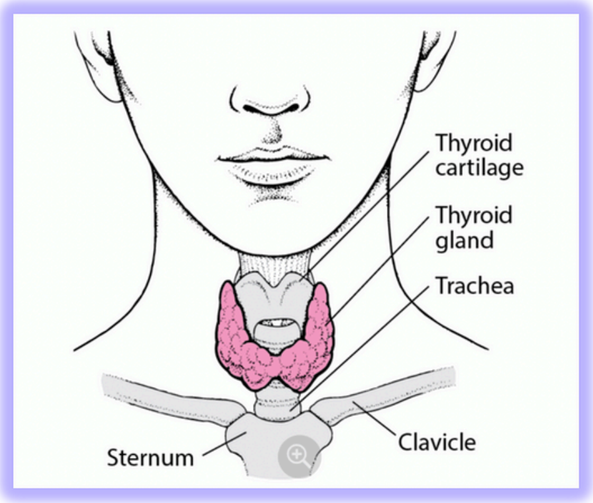
THYROID DISEASE DETECTION USING MACHINE LEARNING



A DATA-DRIVEN APPROACH

PRESENTED BY

PRATIK GOLE
202201070100

RUSHIKESH SABLE
202201070107

RAKSHITA NAIK
202201070110

JANVI PATIL
202201070113

ABSTRACT

- Thyroid gland plays a major role in maintaining the metabolism of human body. Data mining in health care industry provides a systematic use of the medical data.

- Thyroid diseases are most common today. Early changes in the thyroid gland will not affect the proper working of the gland.

- By the early identification of thyroid disorders, better treatment can be provided in the early stage thus can avoid thyroid replacement therapy and thyroid removal up to an extent.

Problem Statement:

- Thyroid disorders, including hypothyroidism and hyperthyroidism, are common and can lead to severe health complications if not diagnosed early.
- Traditional diagnostic methods are time-consuming, expensive, and often require specialized expertise.
- The growing volume of medical data calls for automated, scalable, and accurate solutions to assist in thyroid disease detection.

Objective:

- Develop a robust machine learning model to automate the classification of thyroid conditions (normal vs. hypothyroid).
- Enhance diagnostic accuracy by leveraging advanced data preprocessing and algorithm optimization.
- Minimize false positives and false negatives, ensuring reliable predictions for healthcare use.
- Provide an efficient framework for deployment in real-world healthcare applications.

PROPOSED METHOD

- In the proposed method, we are performing Data pre-processing step, in which feature engineering, feature selection, feature scaling steps are performed and then we are doing model building and performance testing.

- We are using different classification algorithms to classify the thyroid disease type.

DATASET → HYPOTHYROID.CSV

Key Features:

- Binary and categorical columns.
- Demographic details (e.g., sex, age).
- Medical indicators (e.g., TSH, T3, TT4 levels).

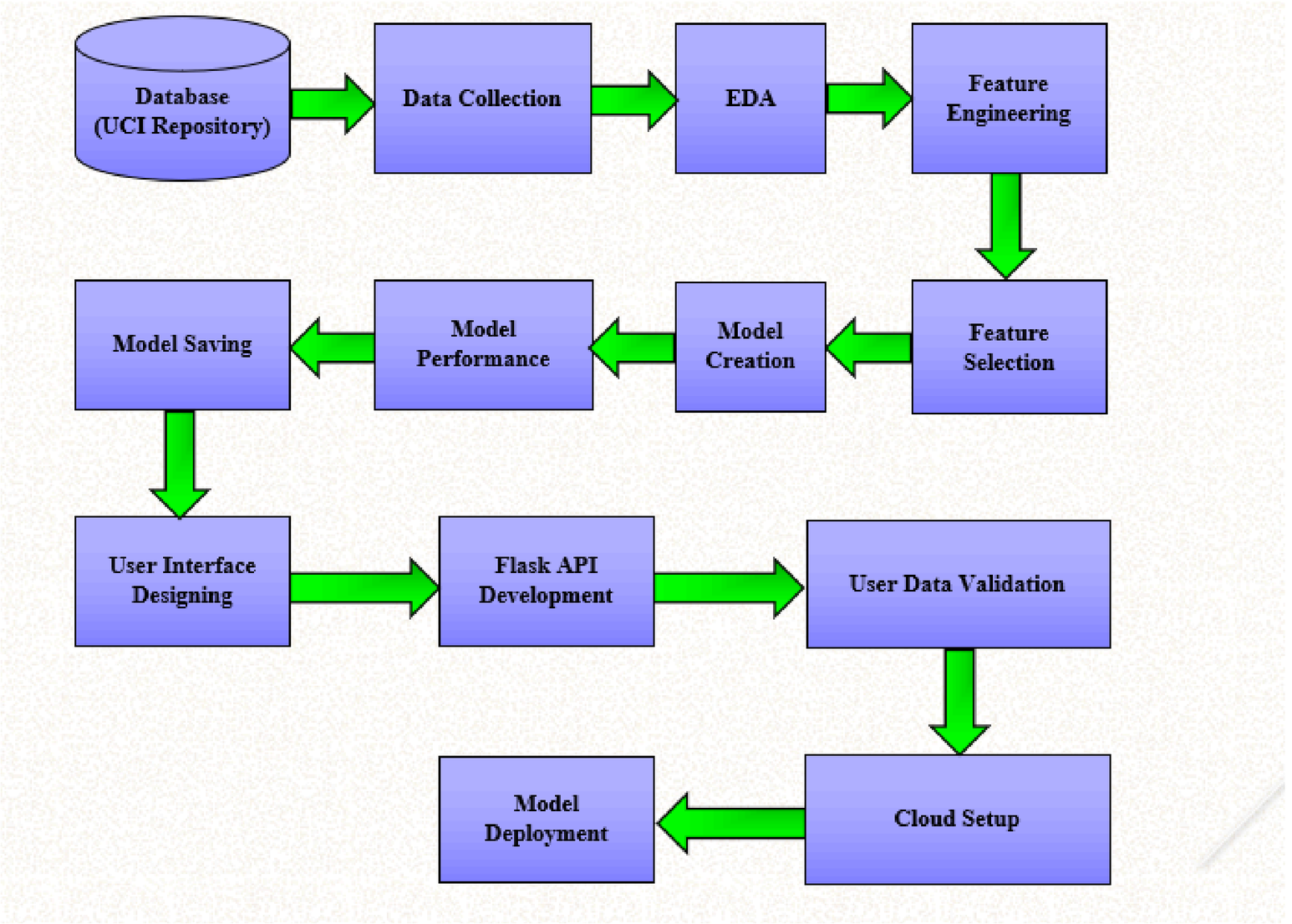
Target:

- Classification as hypothyroid or normal.

DATASET INCLUDES

- **Thyroid Disease Dataset taken from UCI Machine Learning Repository:** URL: https://archive.ics.uci.edu/ml/datasets/thyroid+disease
- **Total patients** = 2800
- **Healthy Subjects** = 2503
- **Abnormal Subjects** = 297
- **Numerical Features:** age, TSH, T3, T4U, FTI
- **Categorical Features:** sex, on_thyroxine, query_on_thyroxine, on_anti_thyroid_medication, sick, pregnant, thyroid_surgery, I131_treatment, query_hypothyroid, query_hyperthyroid, lithium, goitre, tumor, hypopituitary, psych, labels

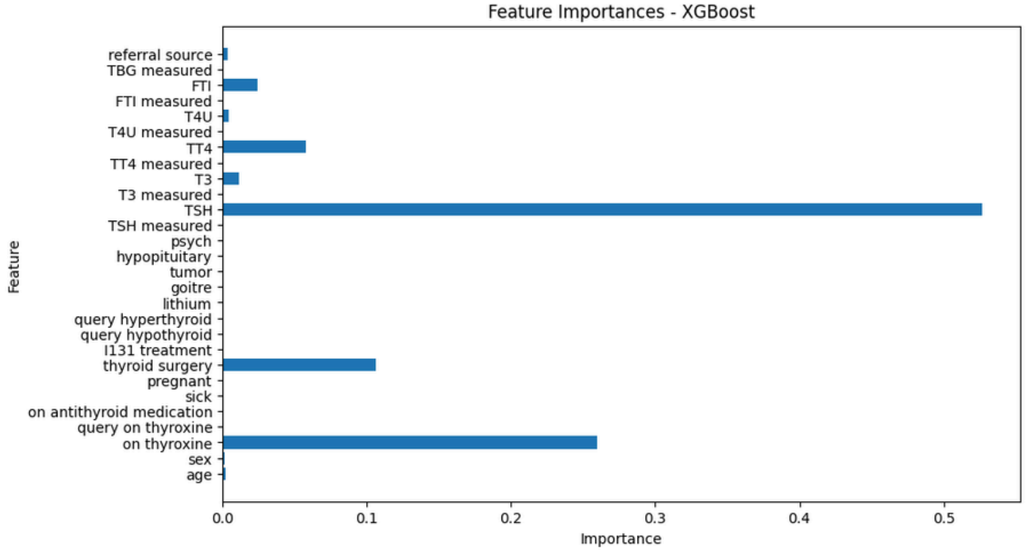
ARCHITECTURE

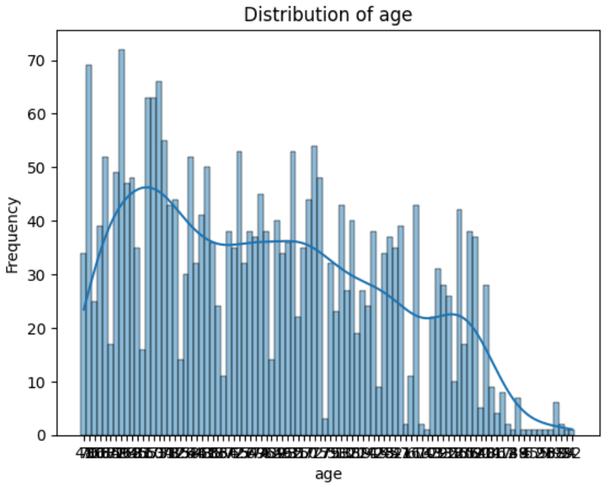


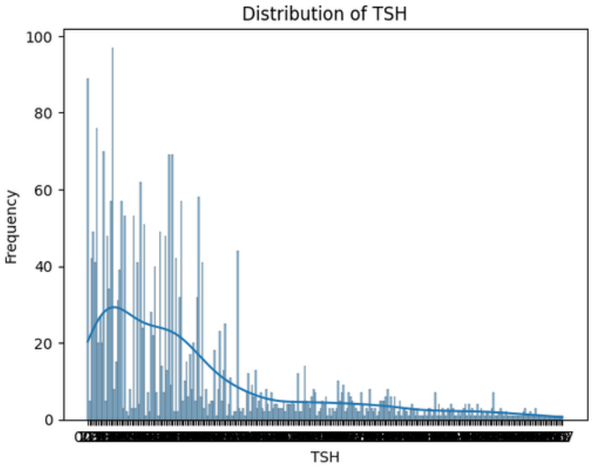
DATA PREPROCESSING

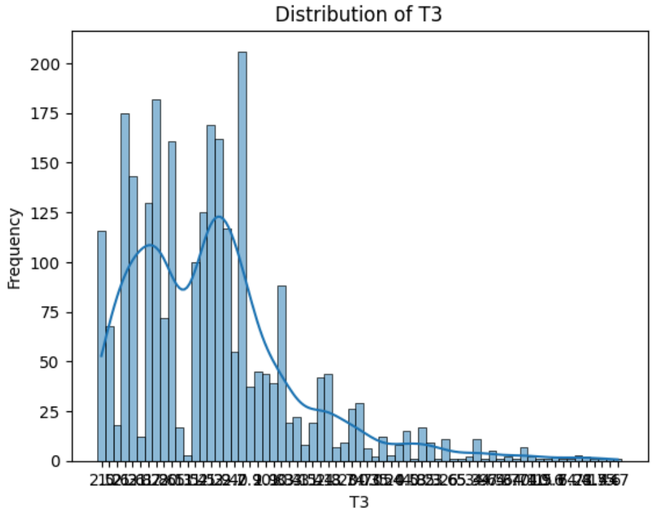
Steps Taken:

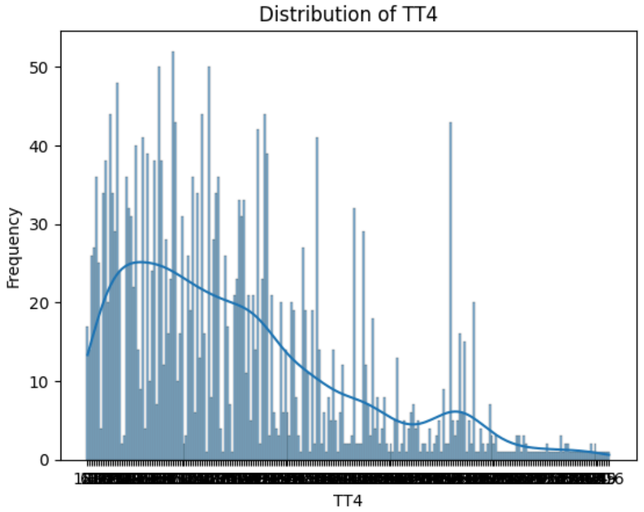
- **Removed missing values ("?").**
- **Encoded categorical variables using label encoding.**
- **Standardized numerical data.**
- **Handled imbalances with appropriate splitting.**

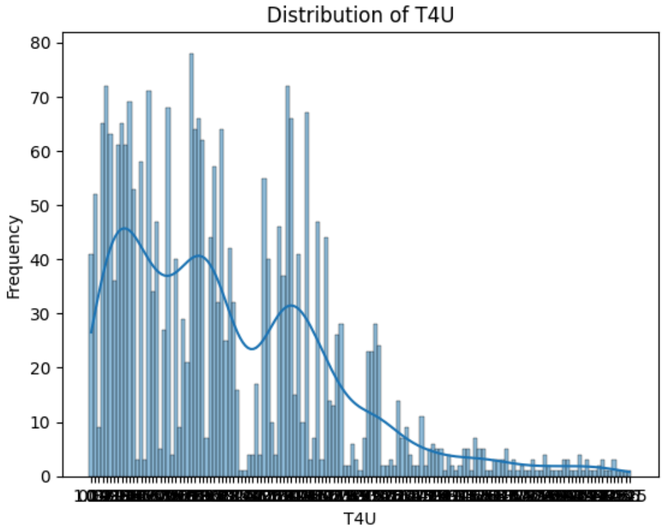

Feature Importances - XGBoost
referral source
TBG measured
FTI
FTI measured
T4U
T4U measured
TT4
TT4 measured
T3
T3 measured
TSH
TSH measured
psych
hypopituitary
tumor
goitre
lithium
query hyperthyroid
query hypothyroid
I131 treatment
thyroid surgery
pregnant
sick
on antithyroid medication
query on thyroxine
on thyroxine
sex
age
Feature
0.0
0.1
0.2
0.3
0.4
0.5
Importance

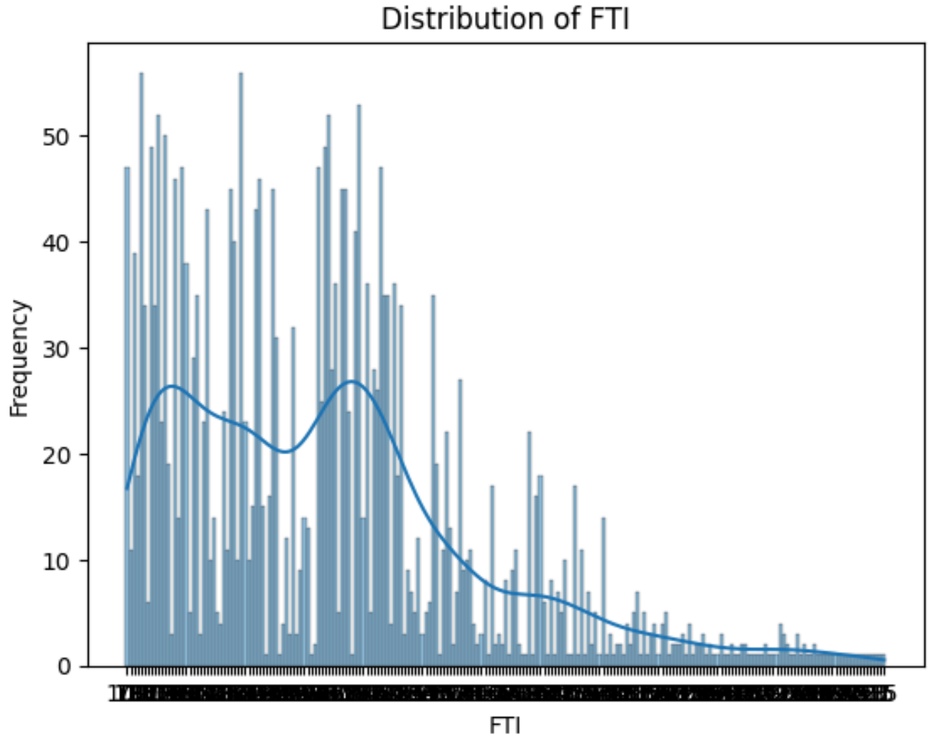

Distribution of age
Frequency
0
10
20
30
40
50
60
70
age

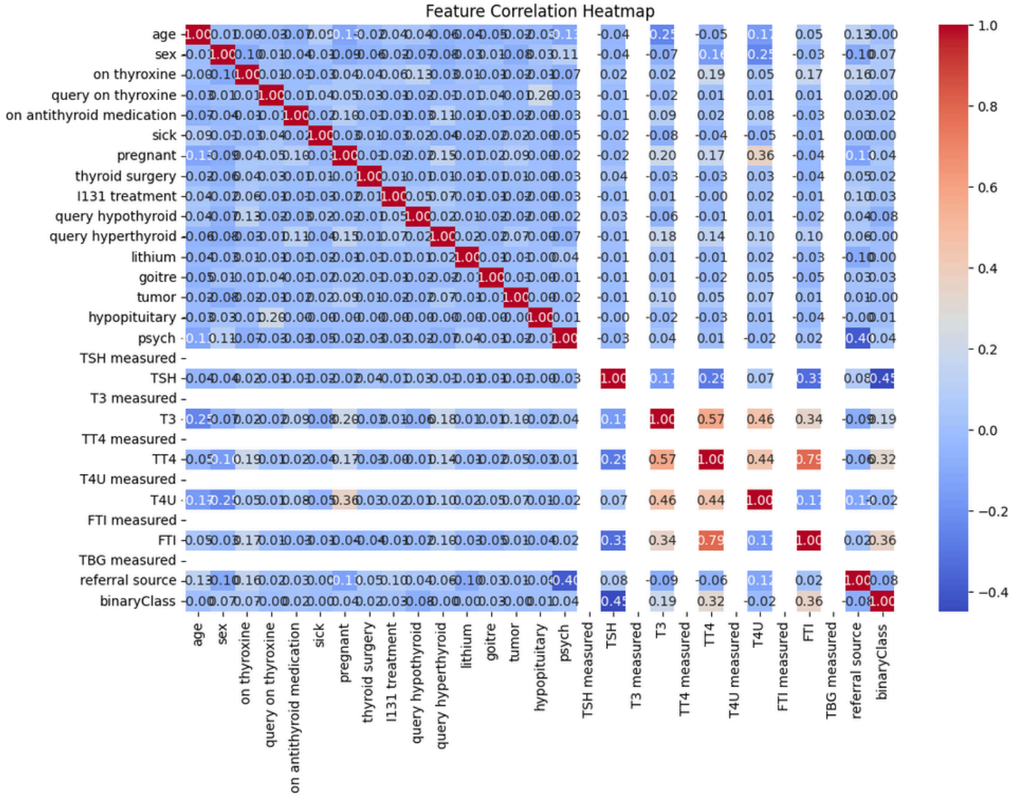

Distribution of TSH
Frequency
0
20
40
60
80
100
TSH


Distribution of T3
Frequency
0
25
50
75
100
125
150
175
200
T3


Distribution of TT4
Frequency
0
10
20
30
40
50
TT4


Distribution of T4U
Frequency
0
10
20
30
40
50
60
70
80
T4U


Distribution of FTI
Frequency
FTI
0
10
20
30
40
50


Feature Correlation Heatmap
age
sex
on thyroxine
query on thyroxine
on antithyroid medication
sick
pregnant
thyroid surgery
I131 treatment
query hypothyroid
query hyperthyroid
lithium
goitre
tumor
hypopituitary
psych
TSH measured
TSH
T3 measured
T3
TT4 measured
TT4
T4U measured
T4U
FTI measured
FTI
TBG measured
referral source
binaryClass
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

MACHINE LEARNING MODEL

MODEL USED: XGBOOST CLASSIFIER

- Random Forest, XGBoost, Decision Tree were produced good result
- XGBoost as the best model chosen for training and testing purpose
- Model performance evaluated based on Accuracy, Confusion Matrix, Recall

MODEL TRAINING

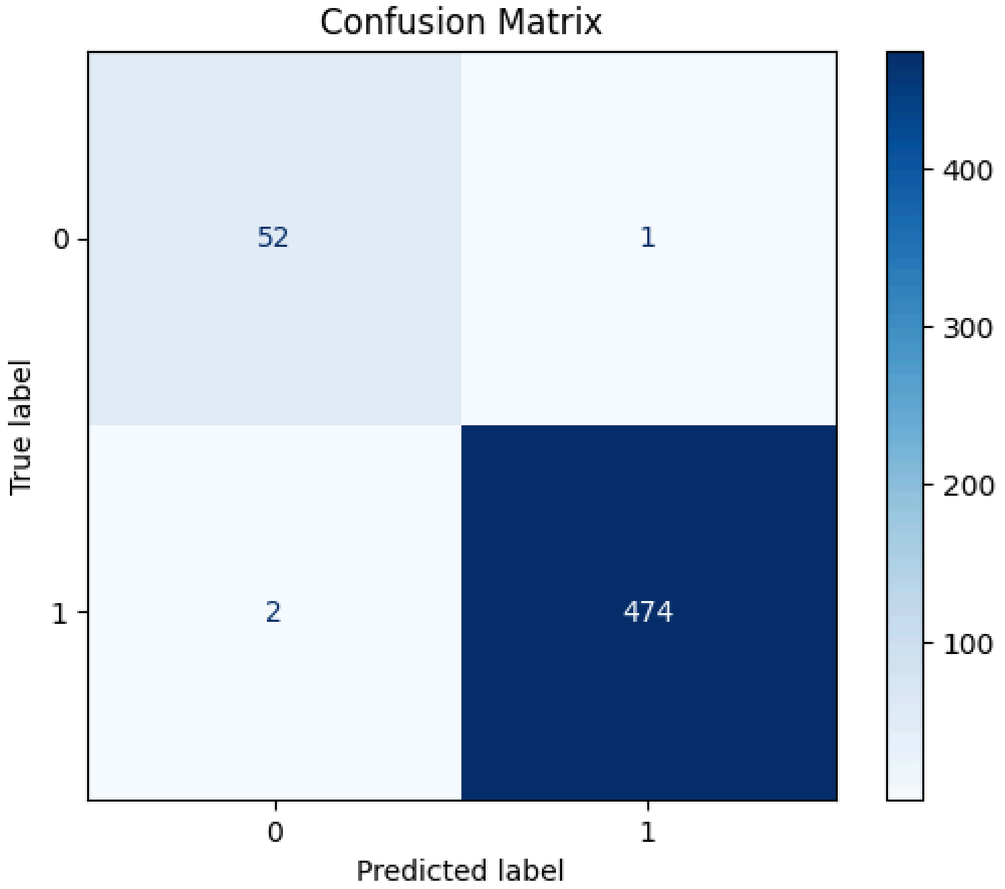
Training Metrics:

- **Accuracy: 99.43%**
- **Precision: 99.79%**
- **Recall: 99.58%**
- **F1 Score: 99.68%**

MODEL EVALUATION

Confusion Matrix

True Positives, True Negatives, False Positives, False Negatives.



REFERENCES

[1] K. Chandel, V. Kunwar, S. Sabitha, T. Choudhury, S. Mukherjee, “A comparative study on thyroid disease detection using K-nearest neighbor and Naive Bayes classification techniques”, pp. 1-7, 2017.

[2] H. Kumar H S, “A Novel Approach of SVM based Classification on Thyroid Disease Stage Detection”, pp. 836-841, 2020.

[3] S. Razia, M. R. N. Rao, “Machine Learning Techniques for Thyroid Disease Diagnosis - A Review”, vol. 9(28), pp. 1-9, 2016.

[4] Z. Parry and R. Macnab, “Thyroid disease and thyroid surgery,” Anaesthesia & Intensive Care Medicine, vol. 18, no. 10, pp. 488–495, 2017.View at: Publisher Site | Google Scholar

[5] G. Serpen, H. Jiang, and L. Allred, “Performance analysis of probabilistic potential function neural network classifier,” in Proceedings of Artificial Neural Networks in Engineering Conference, vol. 7, pp. 471–476, St. Louis, MO, USA, 1997.View at: Google Scholar

[6] Ozyilmaz and T. Yildirim, “Diagnosis of thyroid disease using artificial neural network methods,” in Proceedings of International Conference on Neural Information Processing, pp. 2033–2036, Singapore, 2002.View at: Google Scholar

[7] E. Dogantekin, A. Dogantekin, and D. Avci, “An expert system based on generalized discriminant analysis and wavelet support vector machine for diagnosis of thyroid diseases,” Expert Systems with Applications, vol. 38, no. 1, pp. 146–150, 2011.View at: Publisher Site | Google Scholar

[8] H.-L. Chen, B. Yang, G. Wang, J. Liu, Y.-D. Chen, and D.-Y. Liu, “A three-stage expert system based on support vector machines for thyroid disease diagnosis,” Journal of Medical Systems, vol. 36, no. 3, pp. 1953–1963, 2011.View at: Publisher Site | Google Scholar