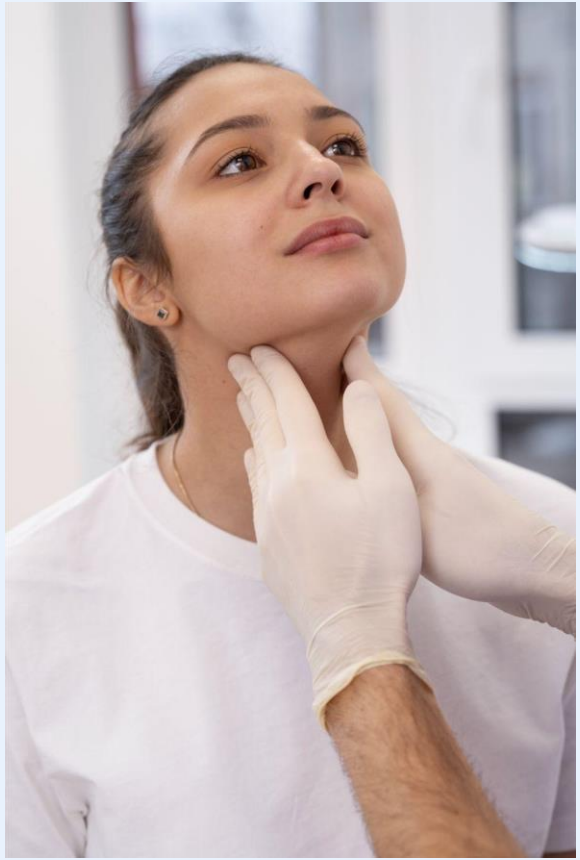
Thyroid Detection Using Machine Learning

Nikhila Baby
S3 MCA
MAC23MCA-2044

Contents

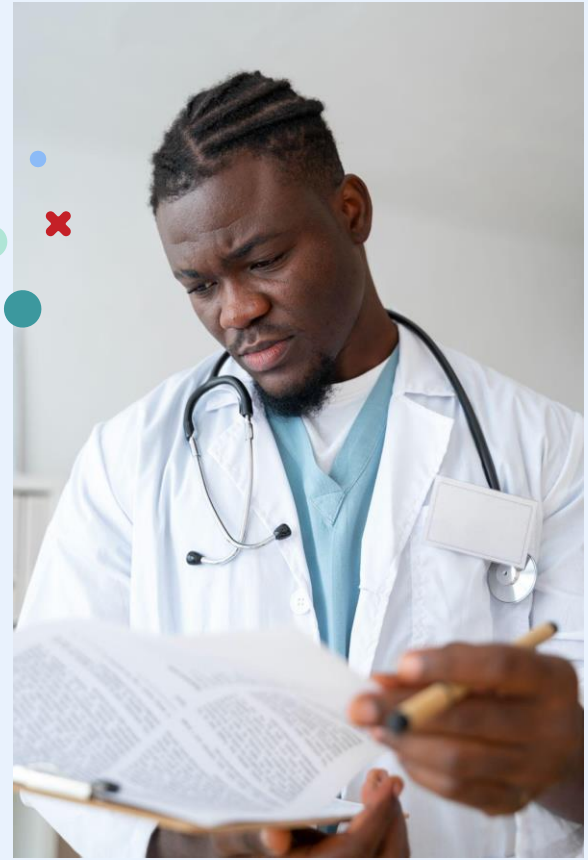
1. INTRODUCTION
2. LITERATURE REVIEW 1
3. LITERATURE REVIEW 2
4. LITERATURE REVIEW 3
5. SUMMARY
6. PROJECT PROPOSAL
7. CONCLUSION



Introduction

- The thyroid gland has one of the most important functions in regulating metabolism.
- Detection and accurate diagnosis of hypothyroidism and hyperthyroidism.
- Aims to utilize machine learning techniques to improve the detection and diagnosis of thyroid diseases.
- Develop and train machine learning models using relevant medical data.

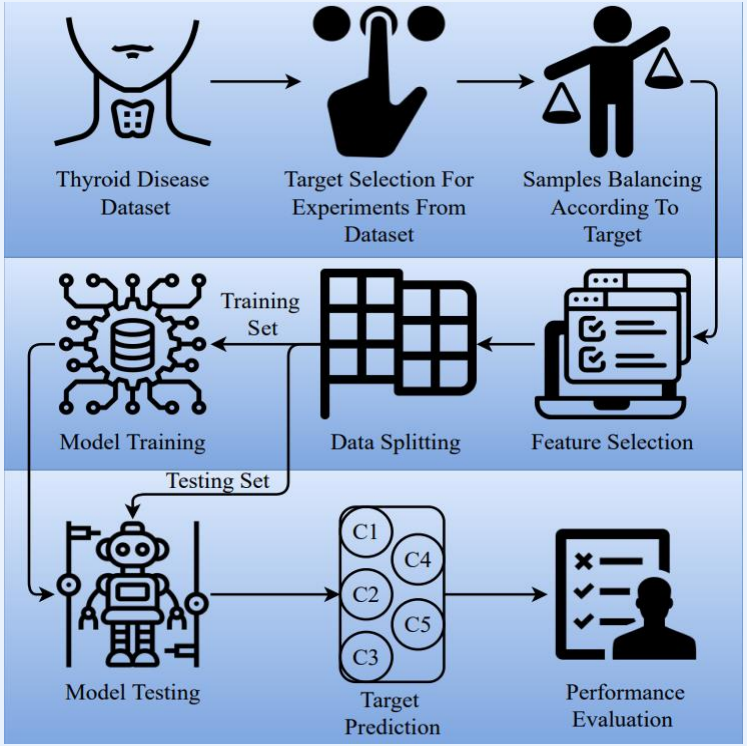
Literature Review



1 Thyroid disease prediction using selective features and machine learning techniques

- Focuses on enhancing feature engineering techniques for machine learning.
- Machine learning models with enhanced feature engineering are more accurate and computationally efficient for detecting thyroid diseases.
- Accuracy of all proposed algorithm is arrived to show the best model.

1 Work Flow


Thyroid Disease Dataset
Target Selection For Experiments From Dataset
Samples Balancing According To Target
Training Set
Model Training
Data Splitting
Feature Selection
Testing Set
Model Testing
C1
C2
C3
C4
C5
Target Prediction
Performance Evaluation

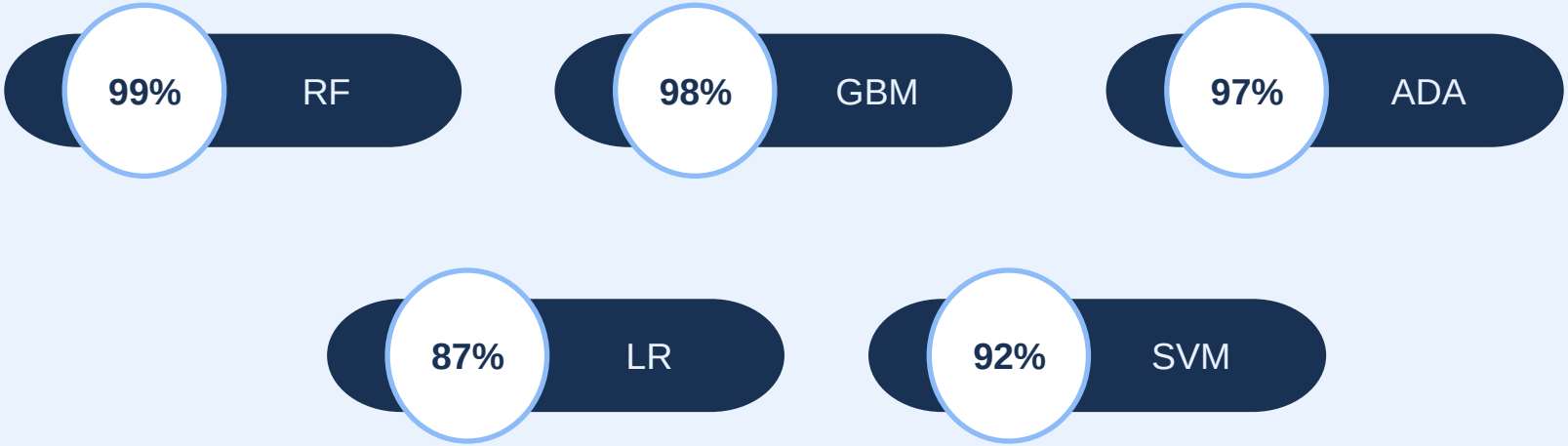
1 Dataset

- Dataset is collected from UCI repository. The dataset contains 9172 sample observations and each sample is represented by 31 features.

Data sample attribute Types

age	pregnant	goiter	T3_measured	FTI_measured	patient_id
sex	thyroid_surgery	tumor	T3	FTI	
on_thyroxine	I131_treatment	hypopituitary	TT4-measured	TBG_measured	
query_on_thyroxine	query_hypothyroid	psych	TT4	TBG	
on_antithyroidmeds	query_hyperthyroid	TSH_measured	T4U_measured	referral_source	
sick	lithium	TSH	T4U	Target	

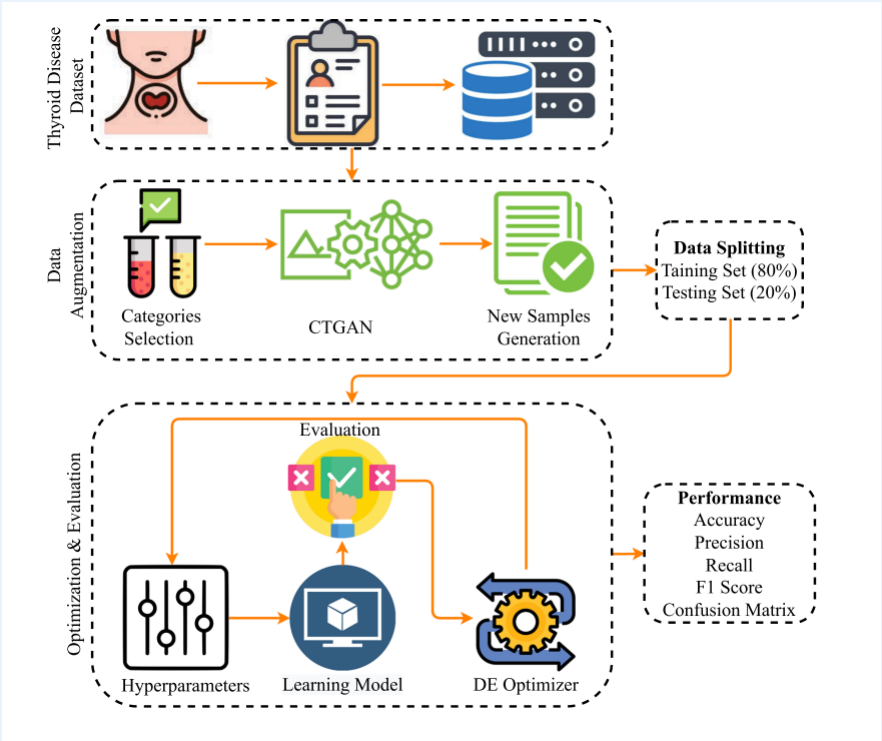
1 Accuracy



2 Detecting thyroid disease using optimized machine learning model based on differential evolution

- Conditional Generative Adversarial Networks (CGANs) are employed for data augmentation to handle class imbalance effectively.
- Hyperparameter optimization is performed using Differential Evolution (DE) algorithm. This algorithm helps to find the best hyperparameter settings for the machine learning models.
- Machine learning models are trained on the augmented data.

2 Work Flow


Thyroid Disease Dataset
Data Augmentation
Categories Selection
CTGAN
New Samples Generation
Data Splitting
Taining Set (80%)
Testing Set (20%)
Optimization & Evaluation
Evaluation
Hyperparameters
Learning Model
DE Optimizer
Performance
Accuracy
Precision
Recall
F1 Score
Confusion Matrix

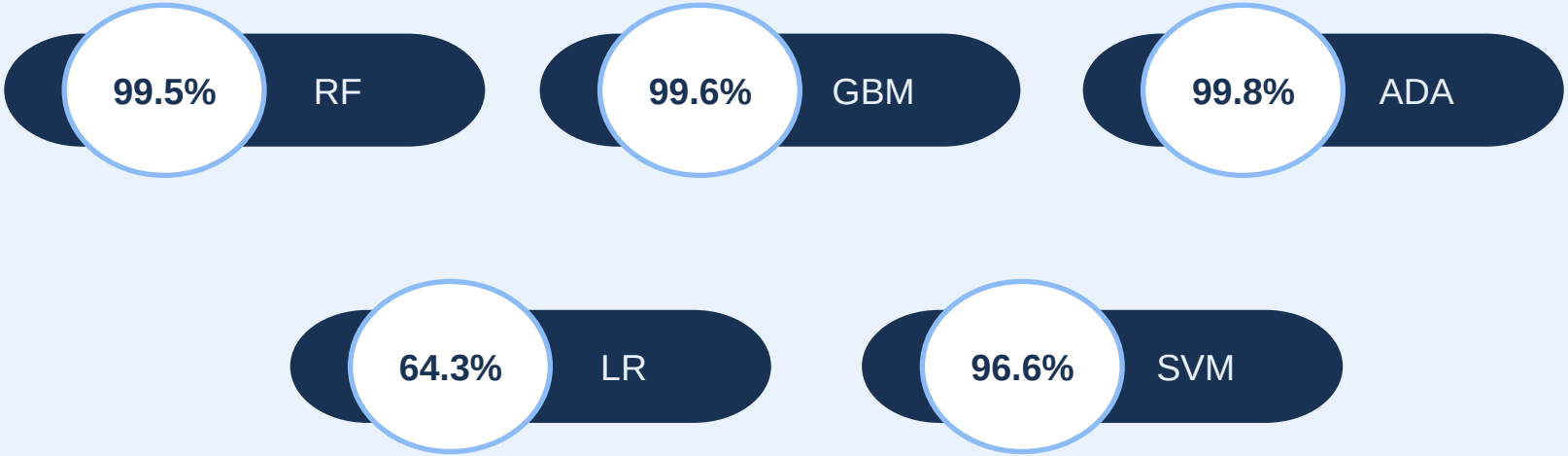
2 Dataset

- Dataset is collected from Kaggle. The dataset contains 9172 sample observations and each sample is represented by 31 features.

Data sample attribute Types

age	pregnant	goiter	T3_measured	FTI_measured	patient_id
sex	thyroid_surgery	tumor	T3	FTI	
on_thyroxine	I131_treatment	hypopituitary	TT4-measured	TBG_measured	
query_on_thyroxine	query_hypothyroid	psych	TT4	TBG	
on_antithyroidmeds	query_hyperthyroid	TSH_measured	T4U_measured	referral_source	
sick	lithium	TSH	T4U	Target	

2 Accuracy

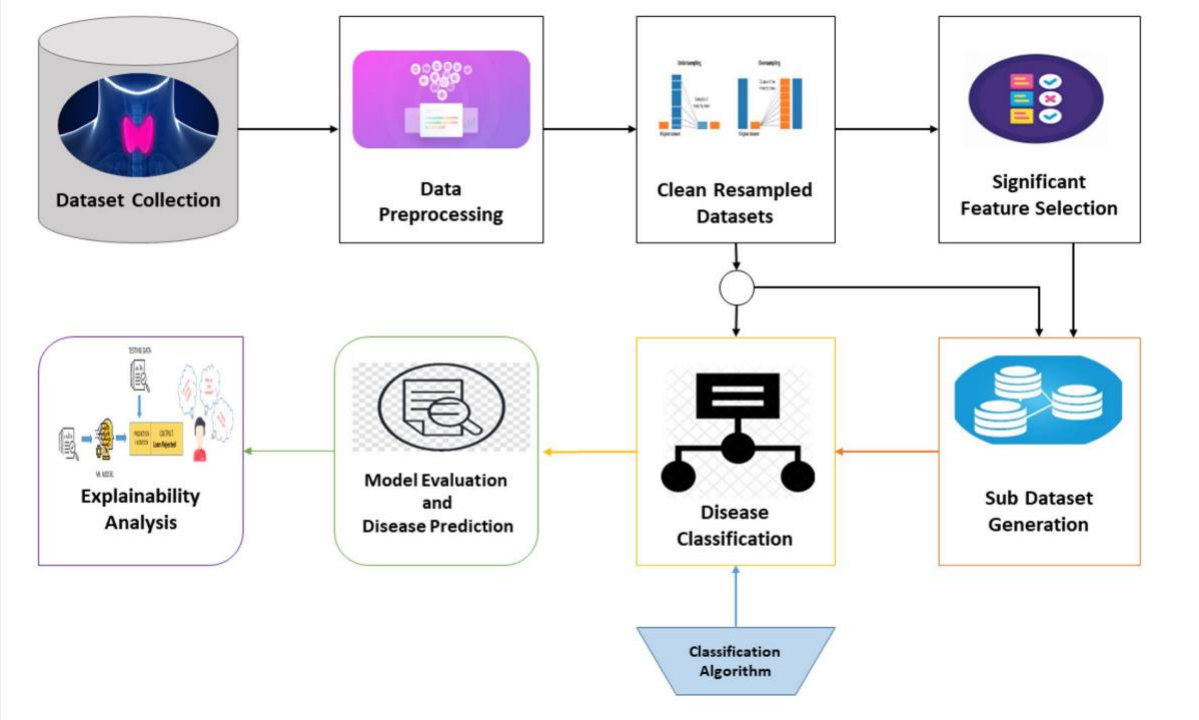


3

An explainable artificial intelligence framework for the predictive analysis of hypo and hyper thyroidism using machine learning algorithms

- Use machine learning algorithms to predict hypothyroidism and hyperthyroidism based on medical data.
- Identifying significant features that enhance the accuracy of disease detection.
- Various classification models are tested to classify thyroid diseases.

3 Work Flow


Dataset Collection
Data Preprocessing
Clean Resampled Datasets
Significant Feature Selection
Explainability Analysis
Model Evaluation and Disease Prediction
Disease Classification
Sub Dataset Generation
Classification Algorithm

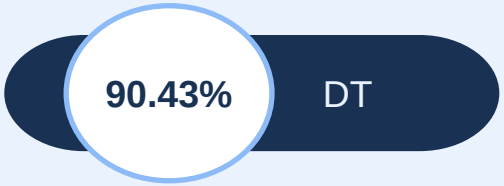
3 Dataset

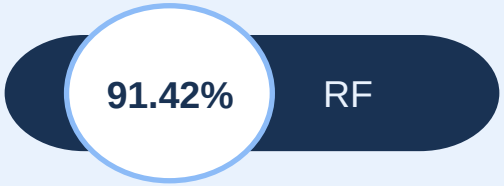
- Dataset is collected from UCI repository. The dataset contains 3221 instances with a total of 30 features.

Data sample attribute Types

age	pregnant	goiter	T3_measured	FTI_measured
sex	thyroid_surgery	tumor	T3	FTI
on_thyroxine	I131_treatment	hypopituitary	TT4-measured	TBG_measured
query_on_thyroxine	query_hypothyroid	psych	TT4	TBG
on_antithyroidmeds	query_hyperthyroid	TSH_measured	T4U_measured	referral_source
sick	lithium	TSH	T4U	category

3 Accuracy





Summary

	Title	Year	Journal Name	Summary
Paper 1	Thyroid disease prediction using selective features and machine learning techniques	2022	Cancers(MDPI)	RF – 0.99 GBM – 0.98 ADA – 0.97 LR – 0.87 SVM – 0.92
Paper 2	Detecting thyroid disease using optimized machine learning model based on differential evolution	2024	International Journal of Computational Intelligence Systems(Springer)	RF – 0.995 GBM – 0.996 ADA – 0.998 LR – 0.643 SVM – 0.966
Paper 3	An explainable artificial intelligence framework for the predictive analysis of hypo and hyper thyroidism using machine learning algorithms	2023	Human-Centric Intelligent Systems(Springer)	DT – 90.43 RF – 91.42 GBM – 90.5 NB – 67.86 KNN – 86.22 LR – 73.15 SVM – 73.7

Project Proposal

- **Comparative study of three algorithms.** RF LR SVM

Why Random Forest ?

- Accuracy
- Figure out which features are most important.

Why Logistic Regression ?

- Gives probabilities for each class, helping to understand how confident the model is in its predictions

Why Support vector machine?

- Consistently delivers good results especially with high-dimensional data
- avoiding overfitting

Project Proposal

- Classification of thyroid disease under three classes.

Hypothyroid

Hyperthyroid

No Thyroid

- Helpful to medical experts to make disease predictions without any human mistakes.
- Patients can diagnose their condition without the assistance of a medical expert.

Dataset

- Dataset is taken from Kaggle Repository.
- 9172 samples and every sample has 31 features.
- Contains numeric values and Boolean values.
- Class labels include letters from A to T which indicates different thyroid conditions.
- Dataset has missing values.

https://www.kaggle.com/datasets/emmanuelfwerr/thyroid-disease-data

Dataset Attributes

age	age of the patient	**I131_treatment**	whether patient is undergoing I131 treatment
sex	gender of the patient	**query_hypothyroid**	whether the patient believes they have hypothyroid
on_thyroxine	whether patient is on thyroxine medication	**query_hyperthyroid**	whether the patient believes they have hyperthyroid
query_on_thyroxine	Whether the patient is queried for thyroxine medication	**lithium**	whether patient is on lithium
on_antithyroidmeds	whether the patient is on antithyroid medication	**goiter**	whether patient has goitre
sick	whether patient is sick	**tumor**	whether patient has tumor
pregnant	whether patient is pregnant	**hypopituitary**	whether patient has hyperpituitary gland
thyroid_surgery	whether patient has undergone thyroid surgery	**psych**	whether patient has psychological conditions

TSH_measured	whether TSH was measured in the blood	**FTI_measured**	whether FTI was measured in the blood
TSH	TSH level in blood from lab work	**FTI**	FTI level in blood from lab work
T3_measured	whether T3 was measured in the blood	**TBG_measured**	whether TBG was measured in the blood
T3	T3 level in blood from lab work	**TBG**	TBG level in blood from lab work
TT4-measured	whether TT4 was measured in the blood	**referral_source**	The source of referral
TT4	TT4 level in blood from lab work	**Target**	target classification for the patient's thyroid condition.
T4U_measured	whether T4U was measured in the blood	**patient_id**	unique id of the patient
T4U	T4U level in blood from lab work		

Class Labels

Class labels include letters from A to T which indicates different thyroid conditions

Hyperthyroid

- A Hyperthyroid
- B T3 toxic
- C Toxic goitre
- D Secondary toxic

Hypothyroid

- E Hypothyroid
- F Primary hypothyroid
- G Compensated hypothyroid
- H Secondary hypothyroid

No Thyroid

- I Increased binding protein
- J Decreased binding protein
- K Concurrent non-thyroidal illness
- L Consistent with replacement therapy
- M Underreplaced
- N Overreplaced
- O Antithyroid drugs
- P I131 treatment
- Q Surgery
- R Discordant array results
- S Elevated TBG
- T Elevated thyroid hormones

THANK YOU