CAPSTONE PROJECT
THYROID CANCER DETECTION USING MACHINE LEARNING

Presented By:
1. Student Name – Vineetha Pavangala
College Name- D R K College of Engineering and technology affiliated J N T UH
Department- Computer Science and Engineering.

edu**net**
foundation

OUTLINE

- **Problem Statement**

- **Proposed System/Solution**

- **System Development Approach**

- **Algorithm & Deployment**

- **Result**

- **Conclusion**

- **Future Scope**

- **References**

PROBLEM STATEMENT

- Thyroid cancer is the most common cancer among women and they were facing many causes mainly pain in throat, breathing difficulties, Fatigue or weakness, Weight loss or gain and thyroid hormone level changes etc. So the early detection crucial for effective treatment.

- The treatment of this were expensive, invasive and time-consuming.

edunet
foundation

PROPOSED SOLUTION

- The proposed system aims to develop a predictive model using machine learning . This involves leveraging data analytics and machine learning techniques to predict patterns accurately. The solution will consist of the following components:

- Data Collection:

  - Collected thyroid cancer patient data from various sources(from a research studies ,medical records).

  - Gathered relevant features like demographic, clinical, imaging data. Ensured the data quality and integrity.

- Data Preprocessing:

  - Clean and preprocess the collected data to handle missing values, null values and normalize features.

  - Feature engineering to extract relevant features from the data that the data split into training and testing sets.

- Machine Learning Algorithm:

  - Trained a machine learning models that are "Random Forest,S V M, and Gradient Boosting .

  - Used above mentioned model algorithms of ML that were suitable for classification and prediction tasks.

- Deployment:

  - Developed a cost efficient thyroid cancer detection using machine learning that it was a predictive model approach.

  - Deployed the trained model in a suitable environment and integrated with existing healthcare systems or tools.

  - Ensuring the model interpretability and explain-ability.

- Evaluation:

  - Evaluated the model's performance on the testing data that metrics are 'accuracy, precision, recall, F1-score, R O C-A U C.

  - Updated the model performance and improved more.

  - Got the best accuracy and precision-recall values .

edunet
foundation

SYSTEM APPROACH

The "System Approach" section outlines the   methodology for developing and implementing thyroid cancer prediction system. We used the following process,

- Data collection and preprocessing

- Feature selection and engineering

- Model training and Evaluation

ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**

- Based on prediction model , the algorithms are Random Forest, S V M(Support Vector Machine), and Logistic Regression.

- Ensemble method for improved accuracy.

- **Data Input:**

  - The data input features includes the patient details of thyroid,stage,age,recurred and response etc. This data gives information about the people who are suffered and who are at which stage and their responses for every session of treatment.
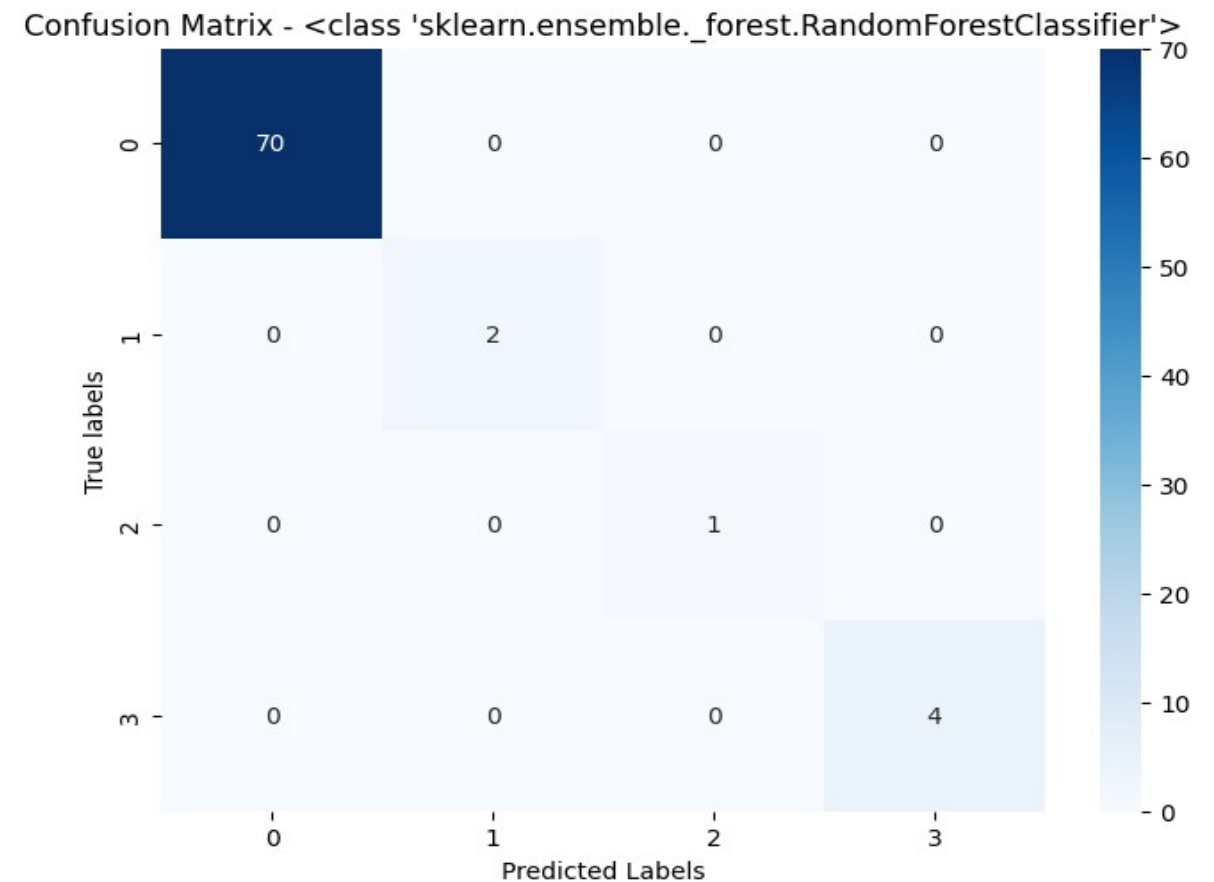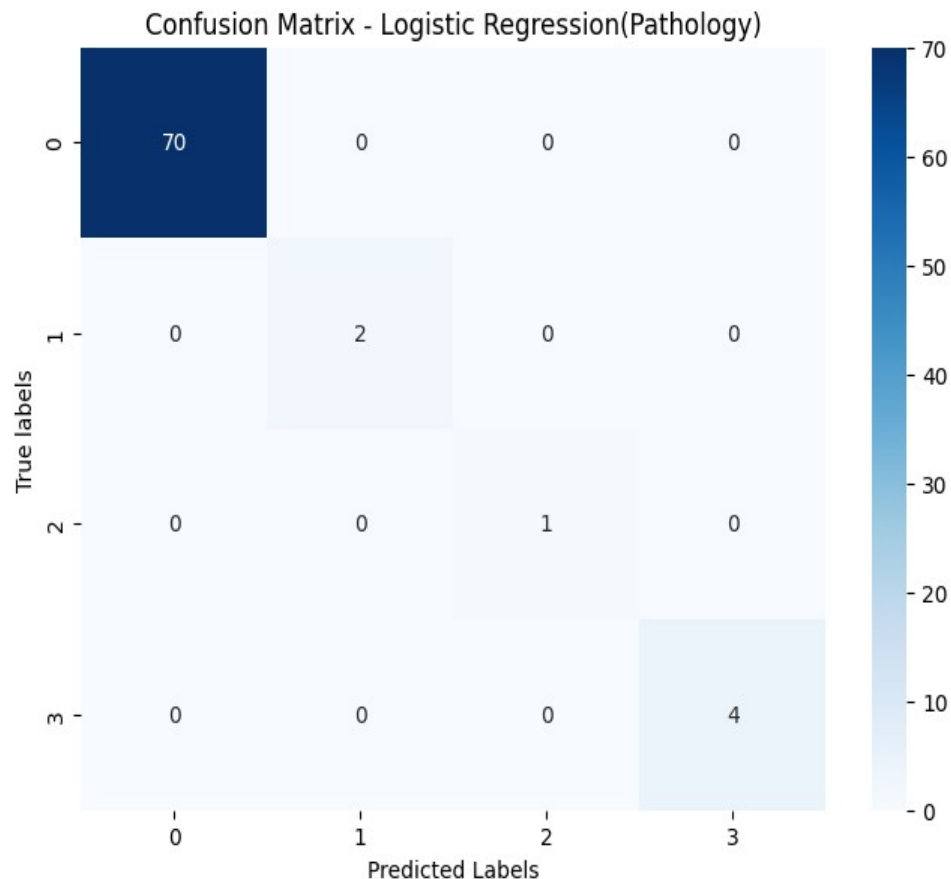
- **Training Process:**

- The training of model is done well by using the data. It was mostly depend on algorithms that it specified in multiple-classification and also trained the model in automated AI also with cross -folds and cross-validations , to check the performance of each category accuracy rate.

- **Prediction Process:**

  - By predicting the model for thyroid cancer that it can detect by using the patient details that it shows the reports of R O C -AUC, that it gives true positives and true negatives about a patient.
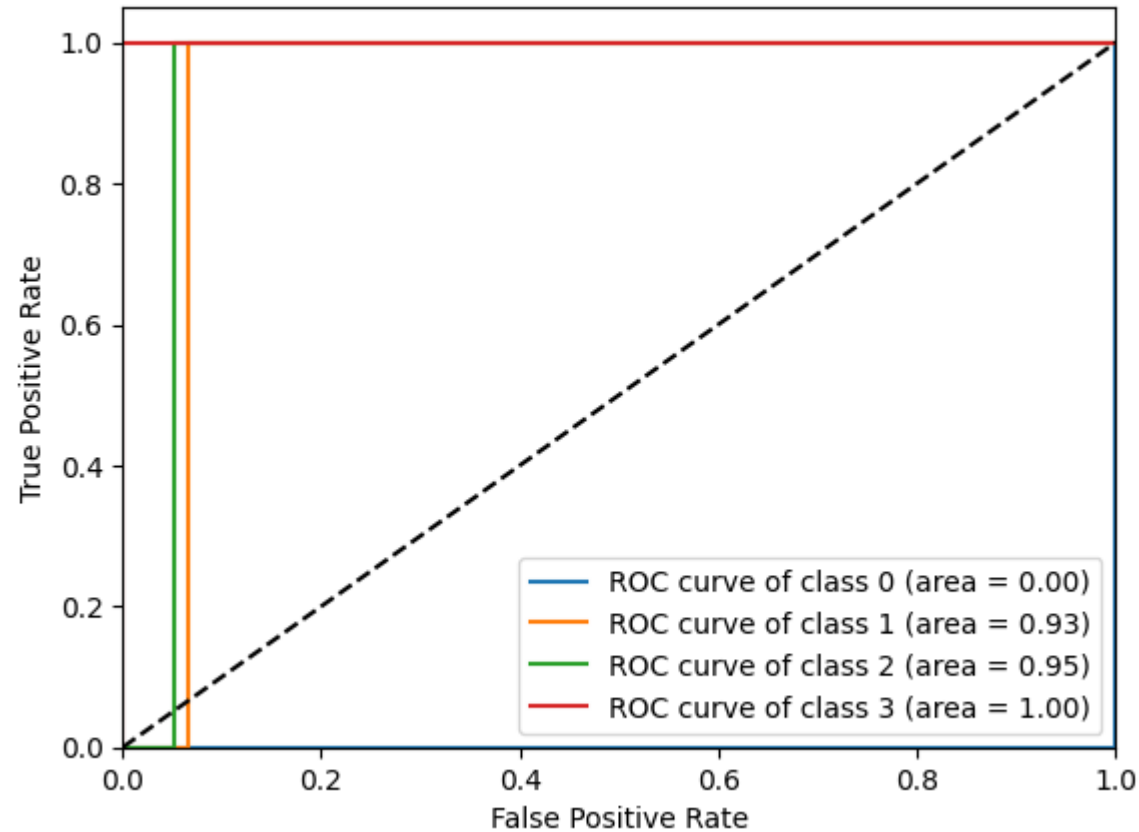
RESULT

The results of developed machine learning model of predictive model approach for detection of thyroid cancer. It includes visualizations and the values of each metric.
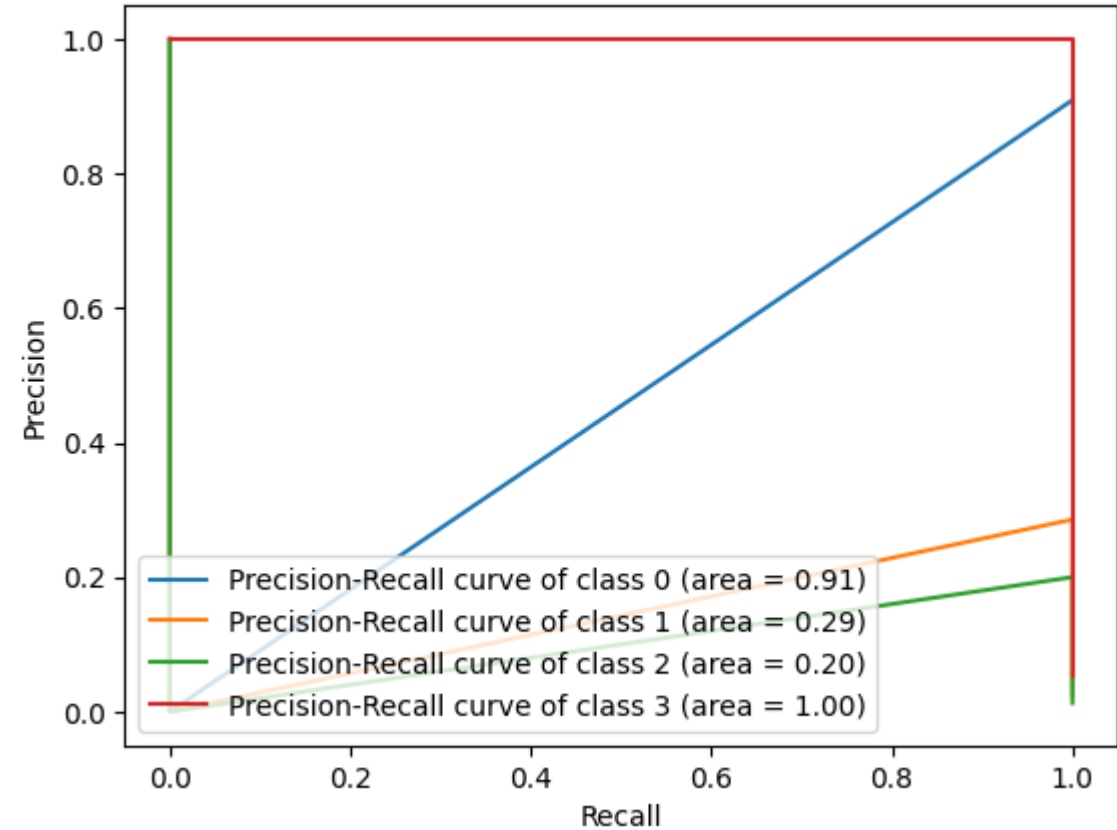

Confusion Matrix - Logistic Regression(Pathology)

Confusion Matrix - <class 'sklearn.ensemble._forest.RandomForestClassifier'>

RESULT

Receiver Operating Characteristic for Multiclass

- ROC curve of class 0 (area = 0.00)
- ROC curve of class 1 (area = 0.93)
- ROC curve of class 2 (area = 0.95)
- ROC curve of class 3 (area = 1.00)

Precision-Recall Curve for Multiclass

- Precision-Recall curve of class 0 (area = 0.91)
- Precision-Recall curve of class 1 (area = 0.29)
- Precision-Recall curve of class 2 (area = 0.20)
- Precision-Recall curve of class 3 (area = 1.00)

edunet
foundation

RESULT

Precision :: 1.0

Recall : 1.0

Classification Report:

	precision	recall	f1-score	support
0.0	1.00	1.00	1.00	70
1.0	1.00	1.00	1.00	2
2.0	1.00	1.00	1.00	1
4.0	1.00	1.00	1.00	4
accuracy			1.00	77
macro average	1.00	1.00	1.00	77
weighted average	1.00	1.00	1.00	77

Confusion Matrix:

```
[[70  0  0  0]
 [ 0  2  0  0]
 [ 0  0  1  0]
 [ 0  0  0  4]]
```

CONCLUSION

- Effective machine learning- based system for thyroid cancer detection.

- Potential to improve healthcare outcomes and save lives.

FUTURE SCOPE

- Integration with existing healthcare infrastructure.

- Expansion to other types of cancer detection.

- Continuous improvement and refinement.

REFERENCES

https://www.sciencedirect.com/science/article/pii/S1877050924011992

https://www.kaggle.com/datasets/alsaniipe/differentiated-thyroid-cancer-recurrence-dataset

edunet
foundation

COURSE CERTIFICATE 1

In recognition of the commitment to achieve professional excellence

Getting Started with
Artificial Intelligence
IBM **SkillsBuild**

IBM

Vineetha Pavangala

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: 17 JUL 2024

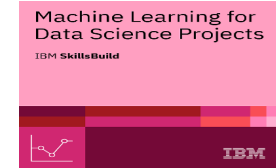Issued by IBM

Verify: https://www.credly.com/go/lpxdjBGm
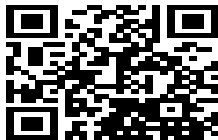
IBM.

edunet
foundation

COURSE CERTIFICATE 2

In recognition of the commitment to achieve professional excellence

Machine Learning for Data Science Projects
IBM SkillsBuild

Vineetha Pavangala

Has successfully satisfied the requirements for:

Machine Learning for Data Science Projects

Issued on: 17 JUL 2024

Issued by IBM

Verify: https://www.credly.com/go/ejJ171bg

IBM

edunet foundation

THANK YOU