iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** VII **Month of publication:** July 2024

DOI: https://doi.org/10.22214/ijraset.2024.63582

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue VII July 2024- Available at www.ijraset.com

Thyroid Disease Classification Using Machine Learning Algorithms

Mrs.S.Krishnaveni[1], E Poojitha[2], M Anthara Reddy[3], G Harshith[4]
[1]*Associate Professor,* [2, 3, 4]*UG Student, Dept. of ECE, CMR College of Engineering & Technology, Telangana*

Abstract: ***The thyroid is very important because it helps our body function properly. It does this by secreting special hormones into our bloodstream that control how our body uses energy. Sometimes there may be problems with the thyroid gland secreting too much or too little hormone. This is because not having enough iodine, swelling of the tumor, or the body's immune system attacking the tumor. Doctors can diagnose these problems with blood tests, but sometimes the results may be confusing. Using special technology, doctors can make information easier to understand and detect how a person may have thyroid problems. This analysis includes analysis of the distribution pattern of thyroid disease and data based on the UCI Machine Learning Repository. Detection of thyroid diseases involves machine learning, which plays an vital role in detecting thyroid disease. Using this technique,doctors can better understand and treat thyroid problems.***

Index Terms: Artificial intelligence; AI; machine-based learning; identification; categorization; image processing; computer vision; pattern recognition

I. INTRODUCTION

Endocrinology incorporates thyroid malady is one of the foremost misjudged and undiscovered maladies. Thyroid organ maladies are the moment most common endocrine disarranges after diabetes concurring to WHO. It influences almost 1% and 2% of people with hyper work hyperthyroidism and hypothyroidism separately. The proportion for men is approximately one per ten ladies. Too the thyroid organ is inclined to different sorts of tumors, and it can turn into an risky area where immune system cells happen (autoantibodies). Physicians say that early malady discovery, conclusion and administration are imperative for halting its movement coming about in passing indeed. Early distinguishing proof and differential conclusion improve chances of favorable treatment in numerous distinctive shapes of anomalies.Even despite of the truth that a few cases have done to discover this, clinical determination is continuously assumed to be difficult task.A butterfly-shaped organ found at the base of our throat; this can be how the Thyroid organ shows up like.Thyroid clutter is classified into two sorts: hypothyroidism and hyperthyroidism. Information mining could be a semi-automated strategy of seeking out for relationships in enormous datasets. Machine learning calculations are is the leading arrangements to numerous issues that are troublesome to unravel Classification could be a information extraction procedure(machine learning) utilized to foresee and recognize numerous illnesses, such as thyroid disease, which we investigated and classified here since machine learning calculations play a critical part in classifying thyroid malady andsince these calculations are tall performing and effective and help in classification . Despite the case that computer learning application and manufactured insights in pharmaceutical can be followed back to the time when it begun, there has been a modern development towards considering machine-learning driven healthcare needs. As a result, investigators anticipate that machine learning will gotten to be common place in healthcare within the near future. Hyperthyroidism could be a clutter in which the thyroid organ discharges so numerous thyroid hormones.

Hyperthyroidism is caused by an increment in thyroid hormone levels. Dry skin, lifted temperature affectability, hair diminishing, weight misfortune, expanded heart rate, tall blood weight, heavy sweating, neck extension, anxiety, menstrual cycles shortening unpredictable stomach movements. Hypothyroidism could be a condition in kind of the thyroid organ is underactive Hypothyroidism is caused by a decrease in thyroid hormone generation. Hypo implies insufficient or less in restorative terms. Aggravation and thyroid organ harm are the two essential causes of hypothyroidism. Corpulence, moo heart rate, expanded temperature affectability, neck swelling, dry skin,hand deadness, hair issues, overwhelming menstrual cycles, and intestinal issues are a few of the side effects. On the off chance that not treated, these side effects can raise over time

II. LITERATURE SURVEY

So in order develop a new project or to develop a new product onemust do some surveys which are available so that the upcoming project will be not having any negative feedback.

©IJRASET: All Rights are Reserved

The survey is so important so that one can know the advantages as well as disadvantages of the existing methods. Surveys helps one to get to know the tools which are required with the aim to implement the method successfully. Surveys also helps the proper understanding of the existing methods and any single can also learn something different from the existing methods. Following are the obtained information from the various papers as portion of the literature review.

Paper [1], In this paper ,The Thyroid Dataset is analyzed using SVM machine learning technique with logistics regression. Based on ROC, RMS error, F measure, Precision, Recall, and F measure,these two methods were compared. The best classifier, it turned out, was logistic regression.Using a radiomics-based approach yields classification accuracy, sensitivity, and specificity of 66.81%, 51.19%, and 75.77%, respectively.

Paper [2], In this paper proposed algorithm is applied to the images in the dataset and required region of thyroid gland is extracted .A total of 45 training patterns, including 20 thyroid tissues and 25 non-thyroid tissues were extracted by an experienced physician, which are used in training the feed forward neuralnetwork.

Accuracy(96.51%),Sensitivity(89.06%),Specificity(98.90%),False negative rate (10.93%), False positive rate(1.90%) respectively

Paper [3], Using datasets from the UCI library, they trained multiclass support vector machine classifiers to build a model and categorize thyroid disorders into four groups: hypothyroid, hyperthyroid, and euthyroid condition. Based on the results of the proposed model, SVM is recognized as a well-known technique in the machine learning community for binary classification problems.Excellent accuracy in high dimensional space is provided by SVM.

Paper [4], In this prospective thyroid disease surveillance pattern to policy makers for the ensuing ten years (2013–2022) by providing data on the prevalence of thyroid illness throughout the previous ten years (2002–2012). The system's methodology consists of three primary processes. The first is the application ofdata preparation techniques. The decision model is then built in the second stage using Time Series Regression (TSR) in R software, and the outcomes are ultimately shown on a geographic map drawn in QGIS.Based on the approach's results, they draw the conclusion that, among women, thyroid disease is mostly similar to impact them and that, over the next ten years, its prevalence may rise by excess 15% in the age group of 21 to 30.

Paper [5], In this the focus dimension deduction techniques are tried .And reduced dimension data input to classifiers.Data augmentation is applied so as to able to generate sufficient data to deep neural network model.The technique for dimension reductionand data augmentation are used and given as input bold indicates the best result obtained to two classifiers. The techniques of feature reduction shows an accuracy of 98.7% while the second with data augmentation technique gives an accuracy of 99.95%.

.

III. EXISTING SYSTEM

1) Hypothyroidism Prediction And Detection Using MachineLearning

It is an analysis and classification model that considers all the variables that are involved in hypothyroidism prediction. Toobtain the best forecast, they used a variety in machine learning methods, including ensemble, Nave Bayes, decision trees, and support vector machines (SVM). At 90 percent and 976 percent, respectively, the decision tree algorithm had the best cost and accuracy. Three classes—compensated hypothyroidism, primaryhypothyroidism, and negative—were formed by the results.

- Disadvantages

It can only classifier the disease established on hypothyroidism but don't detect hyperthyroidism

2) An Interactive Machine Learning-Based Thyroid IllnessPrediction System

Several classification algorithms, including the k-Nearer-Neighbor algorithm, Artificial Neural Network, Decision Tree, and Support Vector Machine, are used in this work. Classification and prediction were carried out using the data set acquired from the UCI Repository, and accuracy was determined by looking at the results generated. They have examined the accuracy of the employed algorithms and comparedthem to determine which method is the most accurate.

Here the output depends on various parameter more vast datais required in dataset.

- Disadvantages:

-Low accuracy

-High power computing.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue VII July 2024- Available at www.ijraset.com

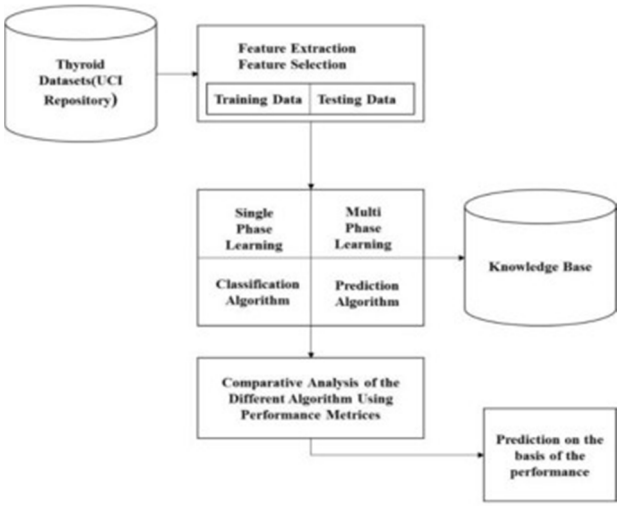


Fig1. Thyroid Prediction System

IV. PROPOSED SYSTEM

The data set is collected from Internet. For balancing the dataset ,we employ resampling technique.We spilled the dataset into training and testing data set . 80 out of 100 data wasfor training ,and 20 out of 100 is used for testing . The trainedata set was used to fit the model and train the algorithm .

Then the data is preprocess is to enhance the standard of the dataset obtained for further analysis. The train data set must be trained using several machine learning methods, such as SVM,Decision Tree, and Random Forest, with the intention to predict the future., and MLP to predict thyroid diseases. Each algorithm istrained on a thyroid dataset containing 3 different categories such as normal, hyperthyroid and hypothyroid. The performance of all algorithms is analyzed depending on accuracy, precision, recall, FSCORE and confusion matrix, and in all algorithms Random Forest and Decision Tree give the,precision ,recall, FSCORE and confusion matrix and compare them to get the highest accuracy algorithm. The most accurate machine learning system is then fed the test data in order to forecast the thyroid disease. Then the output is labeled .

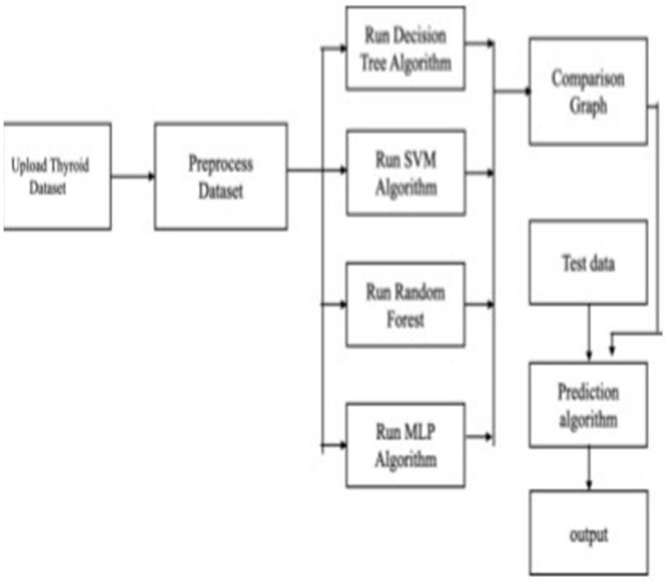


Fig. 2.Block diagram for classifying thyroid diseases using amachine learning method

A. Advantage

1) access to a significant amount of data
2) risk factors influence disease progression
3) more accurate

©IJRASET: All Rights are Reserved

B. Modules

1) Upload Dataset
2) Preprocess Dataset
3) Run Decision Tree
4) Run SVM Algorithm
5) Run Random Forest
6) Run MLP Algorithm
7) Comparison Graphs
8) Predict Thyroid from Test Data

C. Functional Requirements

The system should handle the following:
File management (filename)
Deep learning accuracy tracking (deep_learning_acc)Classification (classifier)
Coordinate tracking (X,Y)

D. Non-Functional Requirements

Usability: The system should be user-friendly, ensuring easeof operation.
Security: Implementation should adhere to security protocols,ensuring data integrity and user privacy.
Readability: The system's interface and outputs should beeasily understandable.
Performance: The system need to exhibit efficient executionand response times.
Availability: The system need to be accessible and operationalwhenever required.
Scalability: The system should be scalable to accommodatepotential future expansions or modifications.

E. Process Model

We adopt the SDLC umbrella model, encompassing stages such as Requirement Gathering, Analysis, Designing, Coding, Testing, and Maintenance. Each stage contributes refinement of the system, ensuring its robustness and effectiveness.

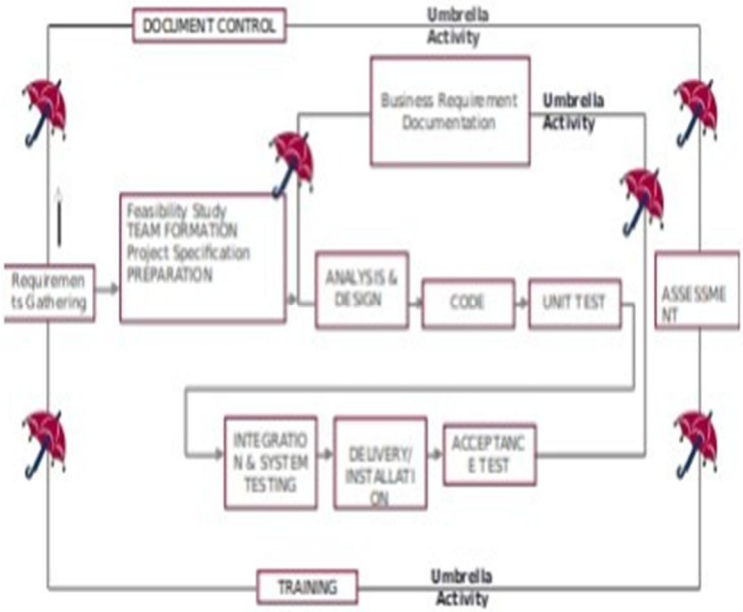


Fig. 3. SDLC Umbrella Model and Architecture .

F. Software Requirements Specification (SRS)

Requirement study: Conducted to identify and address projectneeds and challenges.

Feasibility analysis: Evaluates technical, operational, andeconomic feasibility of the project.
Operational feasibility: Ensures the system aligns withoperational requirements and user needs.
Technical feasibility: Validates the system's technical viabilityand compatibility with existing infrastructure.

G. External Interface Requirements

User Interface: Python Graphical User Interface for userinteraction.
Hardware Interfaces: Interaction facilitated through Pythoncapabilities.
Software Interfaces: Python environment for system operation.Operating Environment: Windows 10.

H. Hardware Requirements

Processor: Intel i3 or i5 (1.1 GHz minimum)RAM: 256 MB (minimum)
Hard Disk: 20 GB
Input Devices: Standard Windows Keyboard, Two or ThreeButton Mouse
Monitor: SVGA

I. Software Requirements

Operating System: Windows 10 Programming Language: Python 3.7.0 or later
Dependencies: OpenCV, Keras, TensorFlow, Protobuf, H5py,Scikit-learn, Numpy, Pandas
Our system's overall goal is to offer a complete solution for classifying satellite images that can accommodate different user requirements and operational conditions.

J. Implementation and Testing

The implementation stage of every project is critical because it transforms the idea into a functional system. During implementation, significant care is required to ensure that all the components function as a whole. In order to confirm the programs integrate as specified in the program specification, each application is tested independently using sample data during development. To satisfy user needs, the complete computer system and its surroundings are tested.



Fig. 4. Data Flow Diagram (DFD) , An illustration thatexplicates the passage of information in a process.

1) Implementation

The implementation phase primarily focuses on user training and file conversion. User training may be extensive, and initial system parameters might require adjustments based on programming outcomes. Simple operating procedures are given to facilitate user understanding of system functions. Overall, the suggested system is straightforward to implement, involving the conversion of a newor revised system design into an operational one.

2) Testing

Testing involves preparing test data and using it to test individual modules and validate fields. System testingconforms that every part works as a whole. Test data should cover various conditions to ensure thorough testing. The objective of testing is to verify that the apparatus functionsaccurately and efficiently before actual operation commences.

3) System Testing

Testing is crucial to cope up the information technology to ensure the reliability and readiness of a system before deployment. Various testing types are employed to guarantee software reliability. Logical and pattern testing are conducted to estimate the program's execution and outcomes for different data sets.

4) Module Testing

Each module undergoes individual testing to detect and fix mistakes without having an impact on other modules. The system's modules are being tested sequentially, starting from the smallest and lowest-level modules and progressing upwards. For instance, modules such as job classification and resource allocation are tested individually to ensure efficient system performance.

5) Integration Testing

After module testing, integration testing is performed toidentify and rectify mistakes that could happen while linking modules. All modules are interconnected and tested to ensure correct functioning of the entire system. Integration testing confirms the accurate mapping of jobs with resources.

6) Acceptance Testing

Once users confirm the system's accuracy and functionality, itundergoes final acceptance testing. This test ensures that the system converge with the original goals, objectives, and requirements established during analysis. Acceptance testing validates the system's readiness for operation without executing actual tasks, thereby avoiding time and cost wastage. These test cases ensure comprehensive testing of the system's functionalities, permitting for the identification and resolution of any issues before deployment.

V. RESULTS

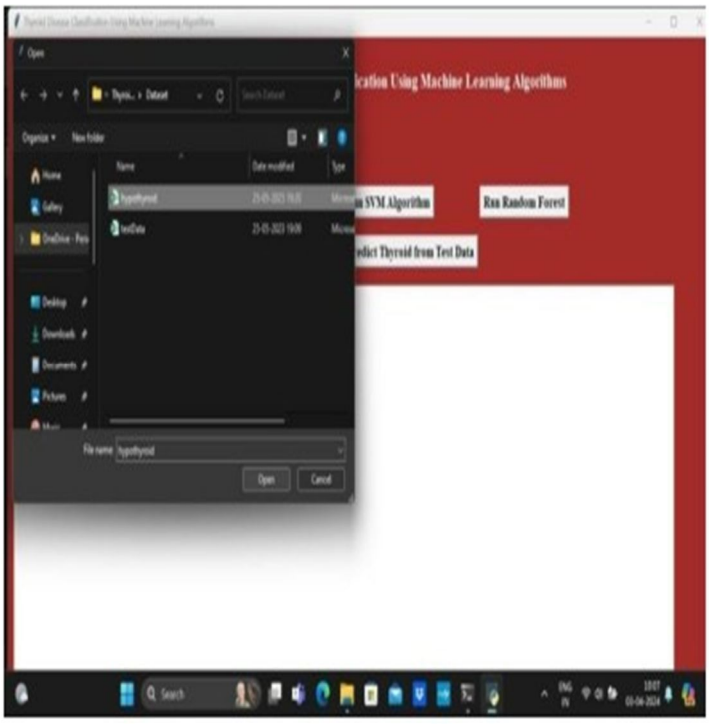


Fig. 5. Entity diagram for Uploading the data set is shown below

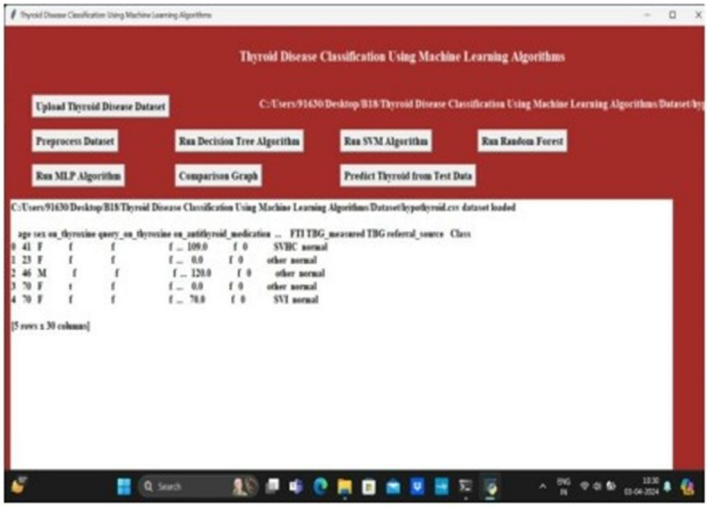


Fig. 6 .Entity diagram displaying that the data is beingpreprocessed

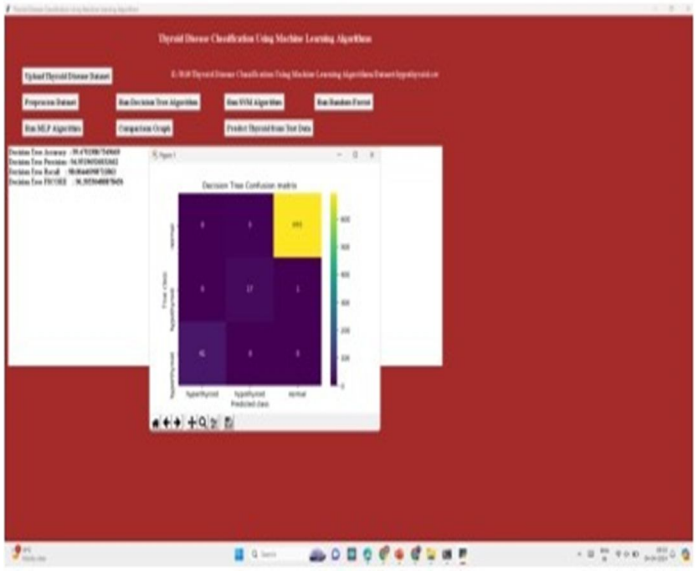


Fig. 7. Entity diagram showing decision Tree has runnedsuccessfully with accuracy of 99%.

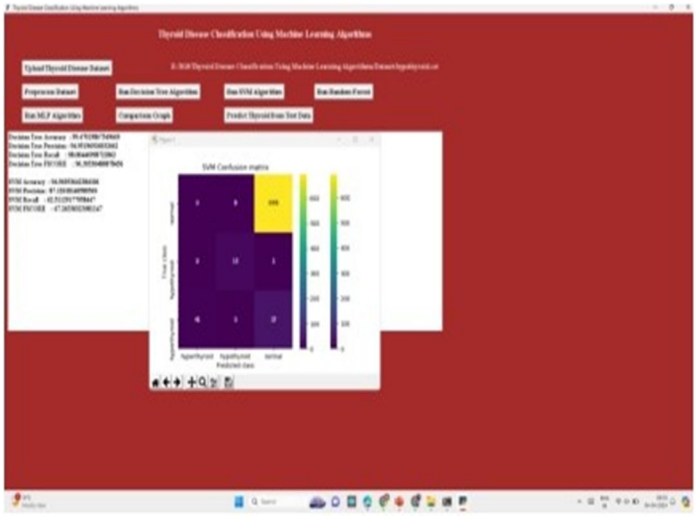


Fig. 8. Entity diagram showing SVM has runned successfullywith accuracy of 94%.

©IJRASET: All Rights are Reserved

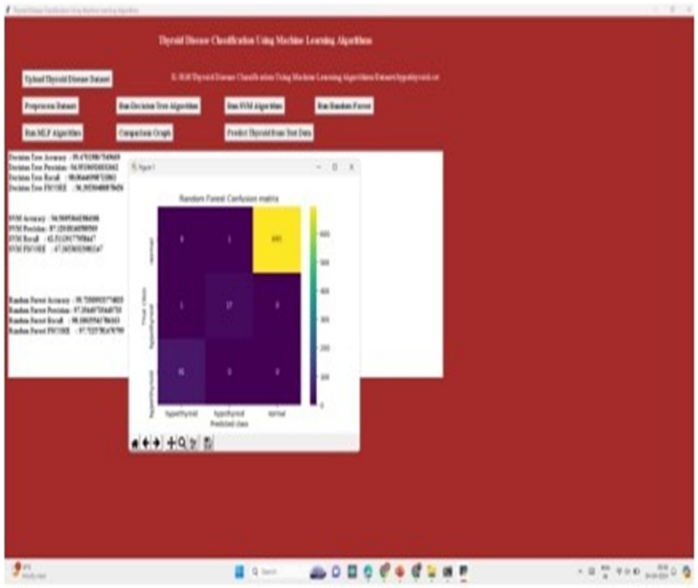


Fig. 9. Entity diagram showing Random Forest has runnedsuccessfully with accuracy of 97%.

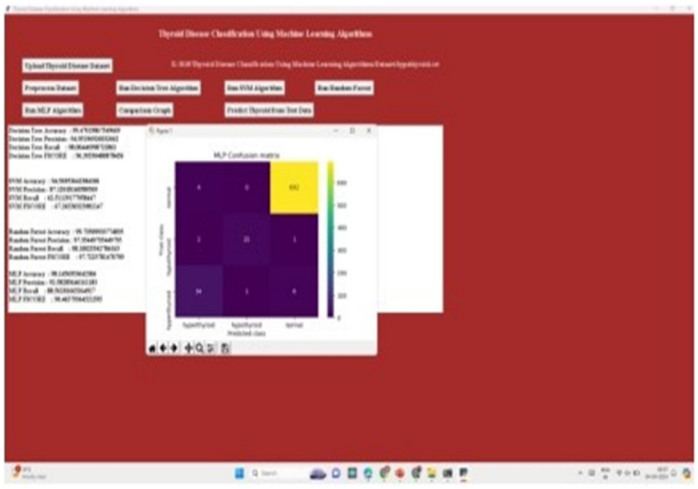


Fig. 10. Entity diagram showing MLP has runnedsuccessfully with accuracy of 98%.

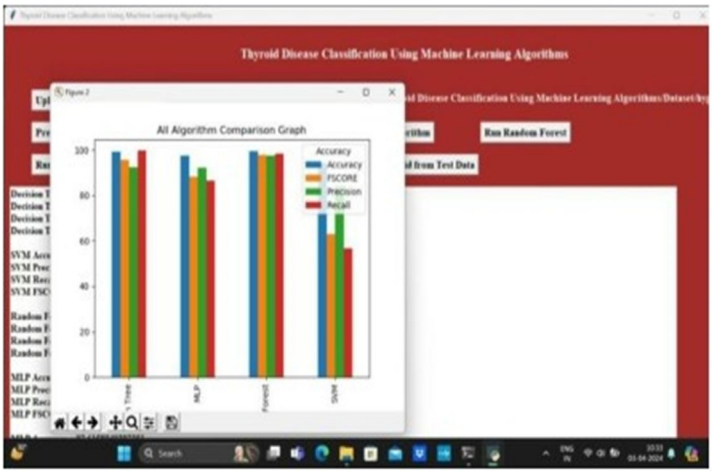


Fig. 11. Entity diagram showing Random forest has highestaccuracy.

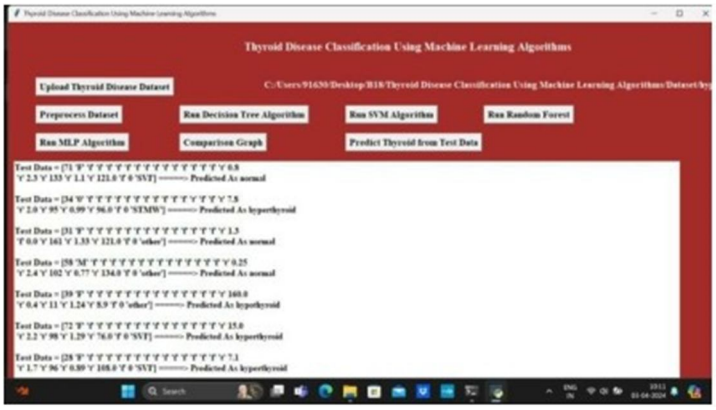


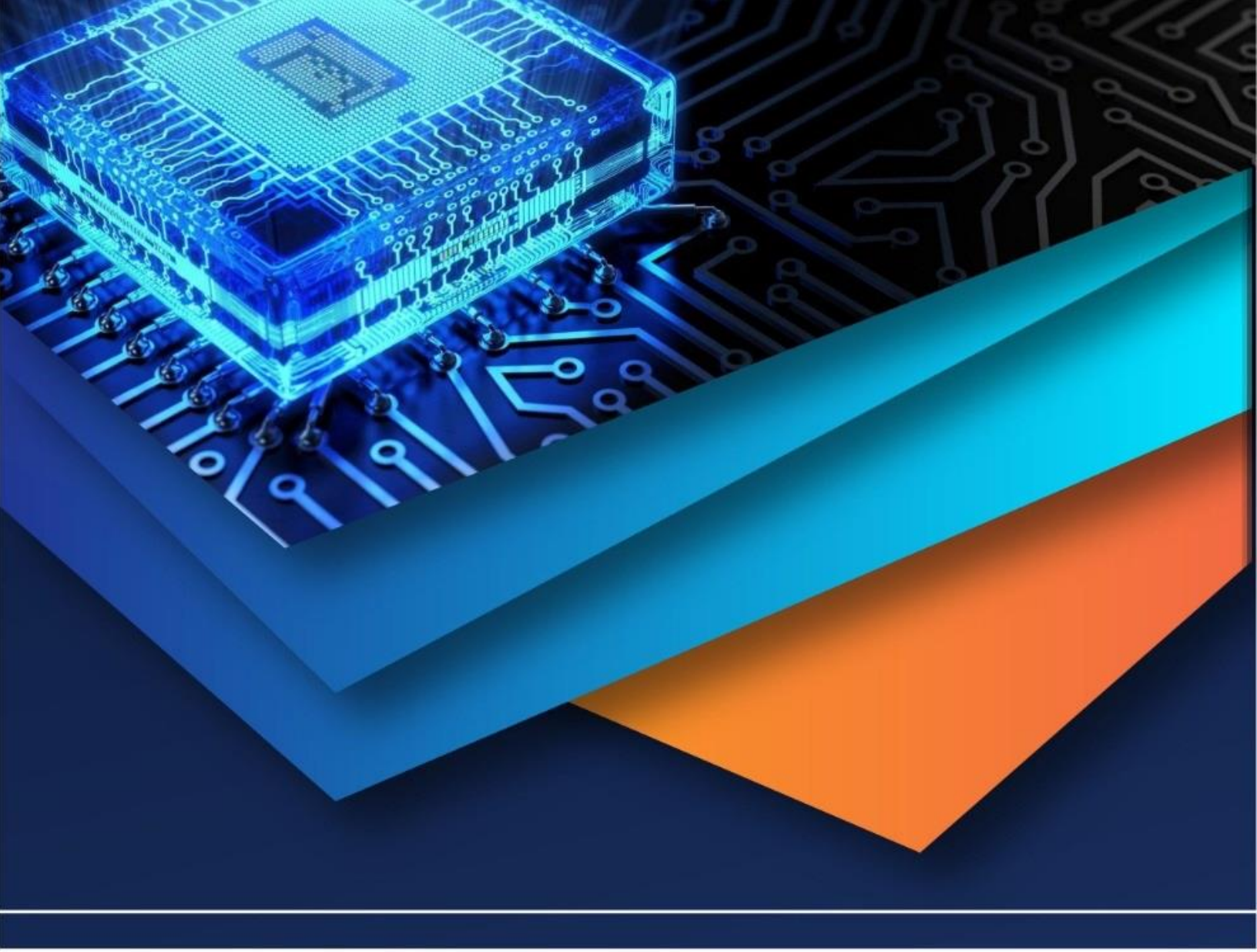
Fig 12. The algorithm that has the highest accuracy is used inpredicting the thyroid from the test data successfully

VI. CONCLUSION

One of the illnesses affecting people worldwide is thyroiddisease, and the number of cases of this condition is rising. Ourstudy focuses on the distinction between hyperthyroidism and hypothyroidism in thyroid disease due to medical reports demonstrating severe imbalances in thyroid disorders.Algorithms were used to classify this illness. With the help of multiple algorithms, machine learning produced impressive results for us and was implemented as two models. With 16 inputs and 1 output, all the characteristics were taken intoaccount in the first model. The random forest method produced the best accuracy of all the algorithms, 98.93. The furtherfeatures of the second embodiment were omitted based on aprevious study. The removed attributes were 1- query thyroxine 2- query_hypothyorid 3-query_hyperthyroid. Here we haveincluded the increased accuracy of some algorithms, in additionthe retention of the accuracy of others. The highest precision ofthe MLP algorithm was 96.4 accuracy.

REFERENCES

[1] Ritika Mehra and Ankita Tyagi (2018).The paper "InteractiveThyroid Disease Prediction System using Machine Learning Techniques" been released on the IEEE website.
[2] Hiteh Garg (2013). "Thyroid Gland Segmentation in Ultrasound Image" using Neural Network"published on Institute of Electrical And Electronics Engineers(IEEE)
[3] Hemanth H S Kumar (2020)"A Novel Method of SVM-based Classification for Determining the Stage of Thyroid Disease" published on Institute of Electrical And Electronics Engineers(IEEE)
[4] Sunila Godara, 2018]. The Institute of Electrical and Electronics Engineers (IEEE) published a paper titled"Prediction of Thyroid Disease Using Machine Learning Techniques."
[5] The article "Computer Aided Diagnosis of Thyroid Disease Using Machine Learning Algorithms" was published on the Institute of Electrical and Electronics Engineers (IEEE) website in 2021 by Md. Asfi-Ar Raihan Asif and Mirza Muntasir Nishat.
[6] A Machine Learning Approach to Predict Thyroid Disease at Early Stages of Diagnosis, Amulya R. Rao and B.S. Renuka (2020) published on the Institute of Electrical and Electronics Engineers (IEEE).
[7] Awasthi, A. K., Antony, A. (2018). An sophisticated approachfor diagnosing and classifying thyroid diseases. 2018 saw theSecond International Conference on Computational Technologiesand Ingenious Communication (ICICCT). IEEE,pp 1261–1264.
[8] Wang, YongFeng (2020). The article "Comparison Study of Radiomics and Deep-Learning Based Methods for Classifying Thyroid Nodules Using Ultrasound Images" was posted on the IEEE Access website.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)