**INTRODUCTION:**

According to recent statistics, approximately 1 in 10 adults in India are suffering with thyroid disease. Small butterfly shaped Thyroid gland controls the metabolism of our body which is present at the front part of our neck under our skin. The thyroid gland mainly produces two hormones they are-T3 (**Triiodothyronine),** T4 (**Thyroxine)**. The pituitary gland in the brain will help the body to keep a check on thyroid hormones. If thyroid hormones are less it will secrete a hormone called TSH (**Thyroid Stimulating Hormone),** which will stimulate the thyroid gland in producing thyroid hormones. There are mainly two types of thyroid disorders, they are-hypothyroidism(underactive), hyperthyroidism(overactive). Thyroid is most common disease among adults, however it shows comparatively higher impact on women. Even though thyroid is one of the most common diseases it can cause heavy damage for the metabolism of the body if not detected and diagnosed properly. It may cause infertility, mood disorders like depression, menstrual problems and also may increase the risk of heart attacks. So it is necessary to detect thyroid in order to avoid major damages. The diagnosis of thyroid disease can be done using tradition clinical tests or using advanced machine learning models. There are certain disadvantages in clinical diagnosis like Diagnosis can vary between physicians due to differences in experience and judgement, The process of clinical testing is time consuming. To avoid these kind of issues we can use machine learning models in place of traditional clinical diagnosis. Machine learning models provide high accuracy and produce consistent and quick results by reducing variability between different physicians. In our machine learning model we will be using various classification algorithms like KNN(K-nearest neighbours) classifiers, random forest classifier, Naive bayes classifier, random forest classifier, SVM(support vector machines) and decision trees. Among all the above classification algorithms random forest have produced more accurate results. We will be taking the data of different patients from the laboratories and convert into .csv file. We will train the model using the above algorithms and the performance of the model is calculated on the basis of Accuracy, Precision, Recall and F1-Score. In our paper we will be using explainable AI (XAI) to explain the results and output produced by our model. There are several methods and processes in the field of XAI, here we are using an open source python library called SHAP(shapely additive explanations). The main advantage of SHAP is that it will help the user to understand which feature is most influential in the model.

**Related work:**

**Thyroid Cancer:**

the collected medical data is storing in a large content from various cities labouratory

T Medical data is developing rapidly, and it originates from hundreds of data sources. The amount of data is huge, the speed of next generation is quick and easy and the format is unique. Because case data supplies a wealth of medical knowledge with regard to illnesses, medications, treatments etc., it is highly useful. For the advancement of smart medicine, it can offer crucial support. A knowledge graph is a type of graph-based data structure that forms a semantic network by accurately representing the relationships between these medical facts in real life. In order to help with disease diagnosis, this research connects disparate and insignificant knowledge in different medical information systems using knowledge graph technology. An semantic network that depicts the concepts, entities, and connections between them in the real world is called a knowledge graph.

It makes full use of visualization technology and has the ability of analyzing and describing knowledge as well as the relationships between knowledge in addition to describing knowledge resources and carriers. The majority of medical knowledge graph studies in the past were literature-based, with the majority of the knowledge derived from publicly available medical literature and seldom use of data from electronic health records. On the other hand, the electronic medical record data contains a wealth of medical knowledge and documents the entire process of diagnosing and treating patients in different hospital departments.

Compared to the information in the subject area, the electronic medical record contains more disorganized data, with the majority of the data being unstructured text. Analyze the data in the electronic medical record using semantic analysis, then take out the knowledge unit that was utilized to create the knowledge graph. Utilize the knowledge graph technology to connect the dispersed and fragmented knowledge in the electronic medical records of medical resources by recognizing the connections between the knowledge units.

This can offer complete services to both physicians and patients. Medical knowledge graphs are also commonly used in medical decision support systems, medical question and answer systems, and medical information search engines.

**Explainable AI:**

Artificial Intelligence (AI) is one of the emerging technologies. In recent years, artificial intelligence (AI) has gained global acceptance in a variety of fields, inclusive virtual support, healthcare, and security. Human-Computer Interaction (HCI) is a area that has been combining AI and human-computer engagement over many years in order to create an interactive intelligent system for user. Artificial Intelligence is the intelligence took over by the machines under which they perform various function. With the help of A.I machines will be able to learn, solve problems, plan things, think, etc. Artificial Intelligence for example, is the simulation of human intelligence by machines. In the field of technology Artificial Intelligence is evolving rapidly day by day and it is believed that in the near artificial intelligence is going to change human life very drastically and will most probably end all the crises of the world by sorting out the major problems. Artificial Intelligence may seem to be a new technology but if we do a bit of research, we will find that it has roots deep in the past. In Greek Mythology, it is said that the concepts of AI were used.

2. MODEL AND METHODOLOGY DATA PREPROCESSING:

Data preprocessing is making the raw data into something useful that can be used by the machine learning model. The raw data contains missing values, and unstable formatted data which cannot be useful for the analysis. So Data preprocessing is required to convert this type of unstable and raw data into useful data. It is the crucial step for creating a machine learning model.

1. Missing Value Filtration: The Data collected from different labs through the cloud is inconsistent and unstable. For the Null Value filtration dropna() function, removes the rows with the null data. For the missing values, the mean of the column is calculated and replaced with the missing values. Calculating the mean by

Mean=ΣX/n

ΣX is the sum of the parameter of the column data and is the number of the data rows for which we are calculating the mean.

1. Correlation Matrix: The fig2. shows the correlation of the data which is predicted to the data which is true, the factors of the data are age, sex, TSH, T3, T4, on\_thyroxine, query\_on\_thyroxine, on\_antithyroid\_medication, sick, pregnant, thyroid\_surgery, I131\_treatment, query\_hypothyroid, query\_hyperthyroid, tumor, psych, Class. The correlation matrix shows how two values are correlated to each other. If the correlation value is high then values are highly correlated and tend to each other. If the correlation value is less then the values don’t tend to each other much. Fig1. is the formula for the correlation coefficient.

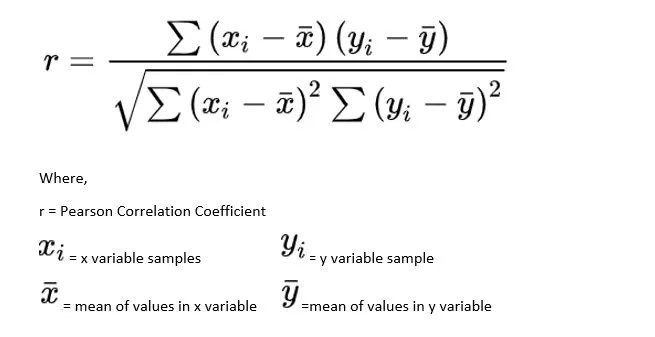


Fig 1. Correlation coefficient formula

Fig 2. Correlation Matrix of XRF model for true and predicted values.

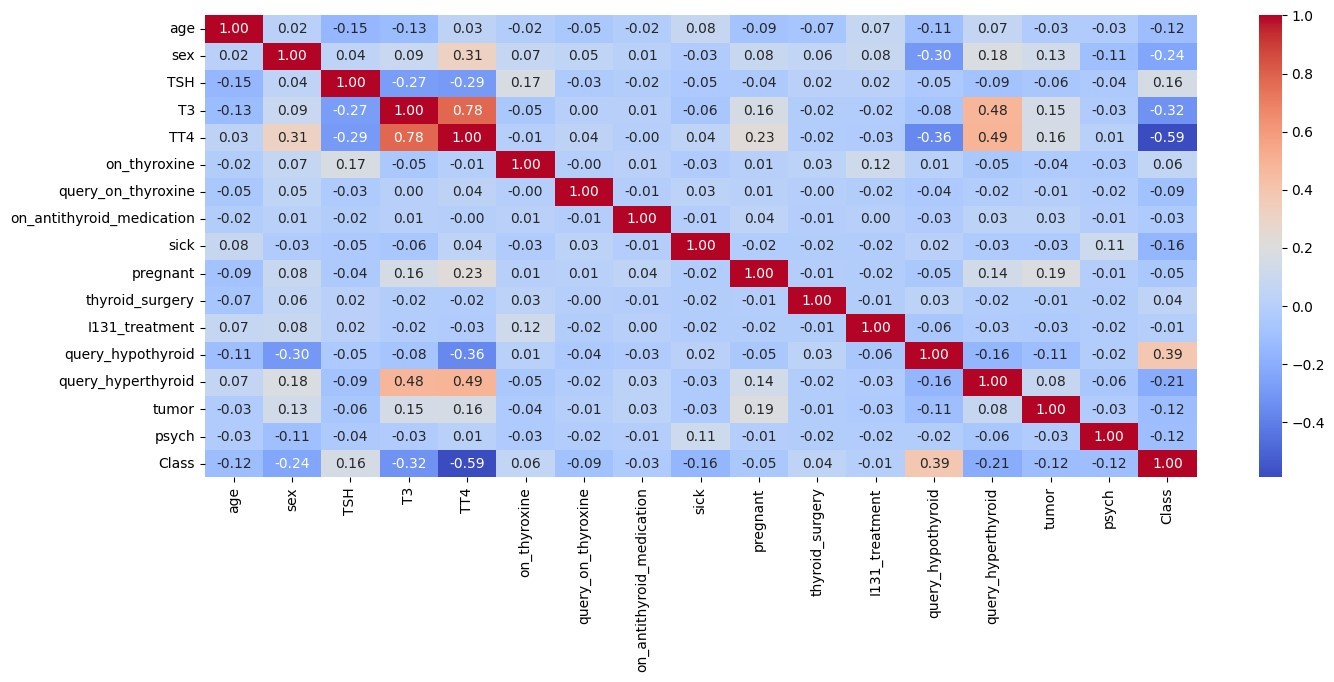
# EXPERIMENTAL EVALUATION

The Experimental evaluation explains the experimental setup, dataset, performance analysis and comparative analysis of the XRF Model.

1.Experimental Setup: The XRF model comprises a software system and cloud.

The prototype of the XRF model is built using the MacBook Air of Apple M2 chip of 64-bit, 8-core CPU with macOS 14.5 version. The programming language used for the models is Python 3.11.4 and Scikit-learn 1.3.0. For the machine learning model to validate the dataset, different machine learning algorithms are used which are Logistic Regression[LR], Gaussian Naive Bayes[GNB], Support Vector

Machine[SVM], K-Nearest Neighbour[KNN], Decision Tree Classifier[DT] and

Random Forest Classifier[RF]. Among the six algorithms, the Random Forest Classifier[RF] has shown the most accuracy. We have used Explainable AI[XAI] techniques called SHAP to interpret the results. For training and testing the models, the data is split in the ratio of 80:20.

2.Dataset and Evaluation Metric: The dataset used consists of rows 26495. The data were collected from different test laboratories, and the test reports were converted. CSV files, and uploaded to the cloud. The CSV file was then taken from the cloud and used for analysis. The data uploaded into the cloud is the patient data that was taken at the time of the taking samples and which is needed by the laboratory; in this process, different features of data from different laboratories are combined. In this process, a large amount of data was missing, null values were created, and the data became unstable. The dataset was processed using preprocessing techniques, and the quality of the data was evaluated with the help

of a correlation matrix. The performance of the XRF model is calculated using the measures of Accuracy, Precision, Recall, and F1-Score.

The Accuracy is defined as how often the machine learning model correctly predicts

the output, it is the ratio of the sum of the True Positive[TP] and the True Negative[TN] to the sum of True Positive[TP], True Negative[TN], False Positive[FP] and False Negative[FN].

Accuracy = TP+TN/TP+TN+FP+FN.

Precision is defined as the ratio of True Positive[TP] to the sum of True Positive [TP] and the False Positive[FP].

Precision = TP/TP+FP

Recall is defined as whether the model can find all the objects of the target class, it can also be defined as the ratio of True Positive[TP] to the sum of True Positive[TP] and False Negative[FN].

Recall = TP/TP+FN

F1-Score is the harmonic mean of precision and recall.

F1 = 2\*TP/2\*TP+(FP+FN)

1. Performance Analysis: The results of the XRF model which is whether the model is predicting or not cannot be understood by everyone, so the results need to be interpreted using Explainable AI(XAI). The Python library that is used in the XRF model is SHAP. By using the SHAP library, the results can be interpreted easily. The factors affecting thyroid disease are mainly TSH, T3, T4, thyroxine levels, tumor presence, and pregnancy presence. For normal range conditions, the TSH value is in the range of 0.7-6.4 µIU/ml, the T4 value is in the range of 5.5-11.1 µg/dL, and T3 value is in the range of 0.86-1.92 ng/ ml. For thyroid, the value will be changed based on the type of thyroid.

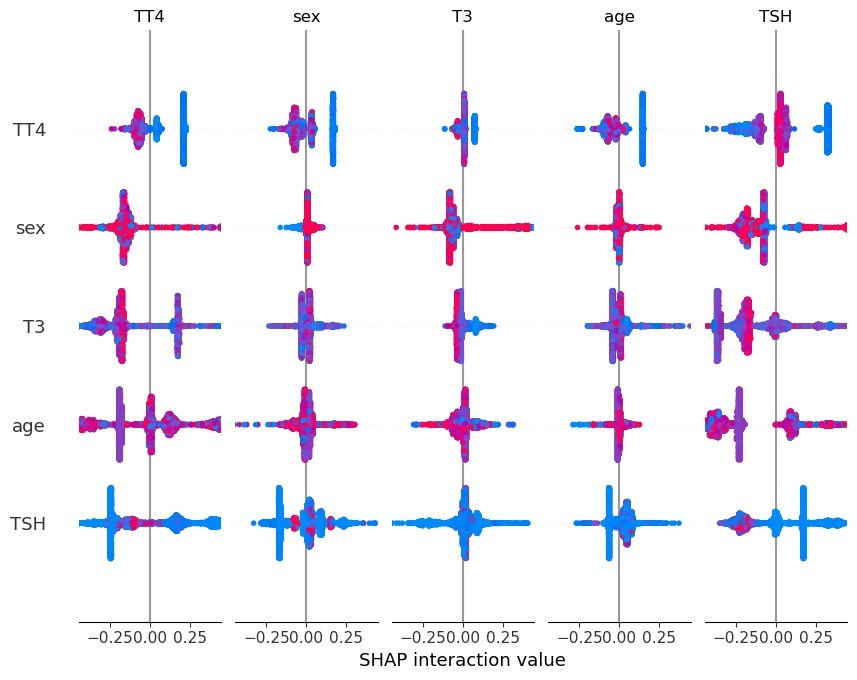


Fig. 3 Interpretation of XRF model using SHAP.

Fig. 3 explains that the interaction between the TT4 and sex is highly significant, with a strong negative prediction. The interaction between T3 and age is also significant with positive prediction. The interaction between TSH and age is less significant, with negative prediction. The red dots indicate the negative prediction and the blue dots indicate the positive prediction.

1. Comparative Analysis: Over the dataset, different machine learning models are used for analysis. The training data contains 21195 rows and the test data contains 5299 rows. Among these models, XRF shows the most prominent results. The Accuracy of the XRF model is 98.74%. Table I shows the results of other models and comparison with the XRF model. Meanwhile, the accuracy of Decision Tree, Logistic Regression, Naive Bayes, Support Vector Machine, and K-nearest Neighbour is 98.58%, 87.39%, 65.21%, 31.49%, and

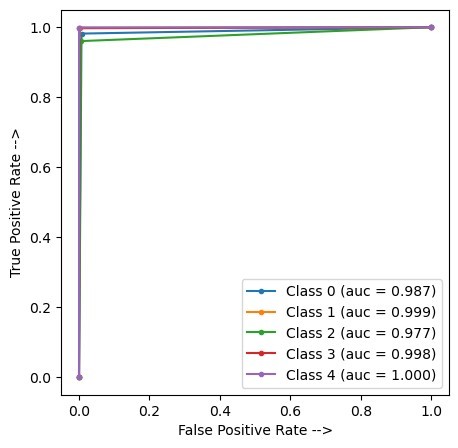
98.09%, these five accuracy values of the other algorithms are lower than the XRF, hence XRF is producing better accuracy and it is a better technique for the prediction.

Table I

Comparison of Different Machine learning models.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Method | Training | Test | Accuracy | Precision | Recall | F1-score |
| XRF | 21195 | 5299 | 98.74 | 97 99 98  100  100 | 98  100  96  100  100 | 98  100  97  100  100 |
| Decision Tree | 21195 | 5299 | 98.58 | 97 99 97  100  100 | 98  100  96  100  100 | 97  100  97  100  100 |
| Logistic Regression | 21195 | 5299 | 87.39 | 75 87 91  94  93 | 87  100  71  81  98 | 81 93 80  87  95 |
| Naive Bayes | 21195 | 5299 | 65.21 | 68 66 67  48  96 | 28  99  2  98  100 | 40  79  4  65  98 |
| Support  Vector  Machine | 21195 | 5299 | 31.49 | 6 3 0  99  97 | 7 7 0  66  78 | 6 4 0  79  87 |
| K-nearest Neighbour | 21195 | 5299 | 98.09 | 95 98 99  100  100 | 100  99 92  100  100 | 97 98 95  100  100 |

Receiver Operating Curve [ROC] is a graph, that is the comparison between the True Positive [TP] rate and the False Positive [FP] rate. Area Under Curve [AUC] of ROC is the representation of the performance of the classifier. In Fig. 4, there are five classes named from 0 to 4, they are Compensated Hypothyroid, Hyperthyroid, Negative, Primary Hypothyroid, and Secondary Hypothyroid respectively. The AUC is in the range of 0.98 to 1.00 units. This area represents the success rate of the classification of classes. The success rates for the classification of Compensated Hypothyroid, Hyperthyroid, Negative, Primary Hypothyroid, and Secondary

Hypothyroid

are 98.7%, 99.9%, 97.7%,

99.8%, and 100%

respectively.

Fig. 4 ROC and AUC for the XRF model.

As the AUC values are higher, nearly equal to 1 and equal to 1. XRF can be considered as the one of the best machine learning model for classification.