**THYROID DISEASE CLASSIFICATION**

**USING ML**

**SUBMITTED BY**

**TEAM LEADER**

**G.SINDHU**

**TEAM MEMBERS**

**G.SIVASAKTHI**

**R.SNEGA**

**S.SNEKA**

**A.PRIYANKA**

**ABSTRACT**

ABSTRACT:

* INTRODUCTION

1.1 Overview

1.2 Purpose

The use of Thyroid Disease Classification Using ML.

What can be achieved using this.

* Problem Definition & Design Thinking

2.1 Empathy Map

2.2 Ideation & Brainstorming Map

* RESULT

Final Outputs of the thyroid disease of classification using ML

* ADVANTAGES & DISADVANTAGES

List of advantages and disadvantages of the proposed solution

* APPLICATIONS

The areas where this solution can be applied

* CONCLUSION

Conclusion summarizing the entire work and findings.

* FUTURE SCOPE

Enhancements that can be made in the future.

* APPENDIX

Source Code

**INTRODUCTION**

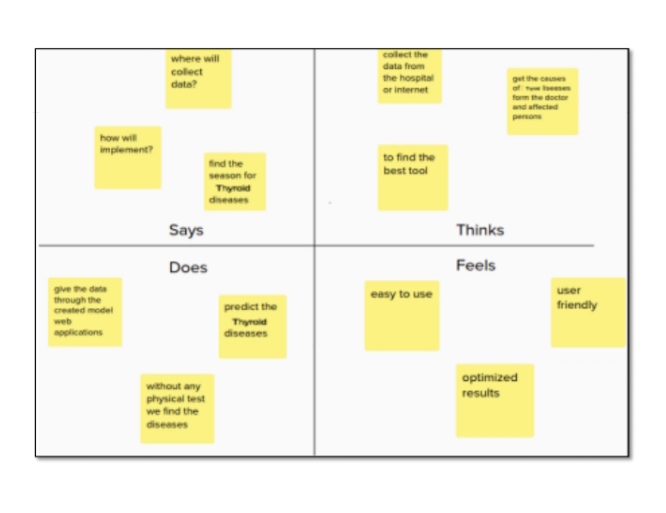
* Overview
* The Thyroid gland is a vascular gland and one of the most important organs of the human body. This gland secretes two hormones which help in controlling the metabolism of the body.
* The two types of Thyroid disorders are Hyperthyroidism and Hypothyroidism. When this disorder occurs in the body, they release certain types of hormones into the body which imbalances the body’s metabolism.
* A thyroid-related Blood test is used to detect this disease but it is often blurred and noise will be present. Data cleansing methods were used to make the data primitive enough for the analytics to show the risk of patients getting this disease.
* Machine Learning plays a very deciding role in disease prediction. Machine Learning algorithms, SVM - support vector machine, Random Forest Classifier, XGB Classifier and ANN - Artificial Neural Networks are used to predict the patient’s risk of getting thyroid disease. The web app is created to get data from users to predict the type of disease.
* Purpose

* The primary diagnosis of thyroid tumors based on histopathological patterns can be ambiguous in some cases, so the use of machine learning techniques might improve the classification of thyroid diseases.
* Moreover, the lack of high sampled datasets makes the classification issue even more complex. This paper proposes a comparative evaluation of two classical machine learning techniques and one Bayesian network framework
* . We use Exploratory Data Analysis techniques and oversampling methods for data preprocessing and overfitting reduction. Results show that the use of Bayesian network frameworks can help in integrating prior expertise knowledge in the classification problem and build new hypotheses about features interaction.

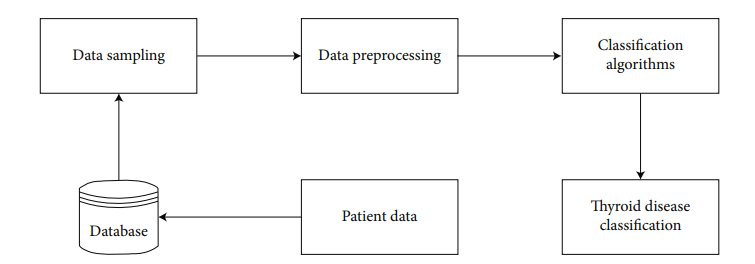
**PROBLEM DEFINITION & DESIGN THINKING**

* Problem Definition
* At present, diseases have become dangerous and rapidly spread, and their exploration and diagnosis require a great deal of time and effort. The correct and accurate diagnosis of the disease early has become one of the problems that the health system suffers from it.

* The critical role of early and correct diagnosis of the disease, including thyroid disease, is vital because it increases patient treatment opportunities and reduces mortality. Among the vast amount of clinical data, early diagnosis is a challenging task.
* Today the machine learning has had impressive and good results in many sciences. Hence, it had a prominent and valuable role in diseases, so this study used machine learning algorithms with thyroid disease. To detect and classify thyroid disease into three types hyperthyroidisms, hypothyroidism, and normal.
* Design Thinking
* Empathy Map

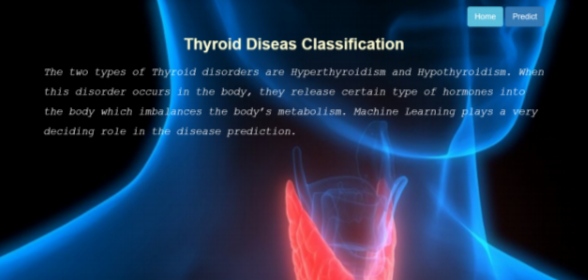


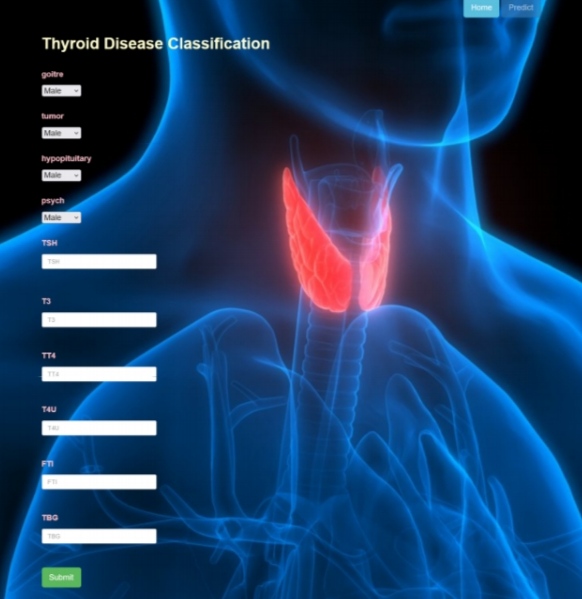
* Ideation & Brainstorming Map

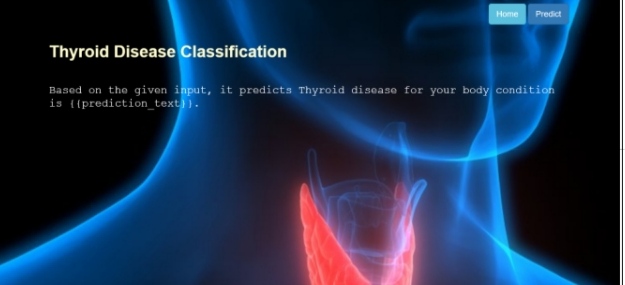


**RESULT**

* RESULT
* Thyroid illness is estimated to affect more than 200 million individuals worldwide. It is also widespread among Iraqis, particularly women. This condition has the potential to have a significant impact on adult bodily processes as well as infant development.
* Thyroid diseases are generally treatable, but they may be dangerous if they progress to the advanced stage, and they may even result in death. The successful classification was performed using various techniques using the thyroid disease data set obtained from private laboratories and hospitals.
* We used eight algorithms in this study (Decision Tree (DT), Supporting Vector Machine (SVM), Random Forest (RF), Naive Bayes (NB), Logistic Regression (LR), Linear Discriminant Analysis (LDA), KNearest Neighbor (K-NN), Multilayer (MLP)). We divided the existing data into two parts, 30% for training and 70% for testing, as this training is the first training on this data. In the first step, we took all the properties in our data and applied them to the eight algorithms.
* We take all the features in our data in the second step except for query\_thyroxine, query\_hypothyorid, and query\_hyperthyroid and they ared used to predict using eight algorithms. Our data contains 16 attributes as one input and one output. That is, the sum of the attributes in this data is 17 attributes.







**ADVANTAGES**

**AND**

**DISADVANTAGES**

* Advantages and Disadvantages
* Machine learning algorithms are one of the best solutions to many problems that are difficult to solve Classification is a data extraction technique (machine learning) used to predict and identify many diseases, such as thyroid disease, which we researched and classified here because machine learning algorithms play a significant role in classifying thyroid disease and because these algorithms are high performing and efficient and aid in classification .
* Although the application of computer learning and artificial intelligence in medicine dates back to the early days of the field , there has been a new movement to consider the need for machine learning-driven healthcare solutions. As a result, analysts predict that machine learning will become commonplace in healthcare in the near future .
* Hyperthyroidism is a disorder in which the thyroid gland releases so many thyroid hormones. Hyperthyroidism is caused by an increase in thyroid hormone levels . Dry skin, elevated temperature sensitivity, hair thinning, weight loss, increased heart rate, high blood pressure, heavy sweating, neck enlargement, nervousness, menstrual cycles shortening, irregular stomach movements, and hands shaking are some of the signs .
* Hypothyroidism is a condition in which the thyroid gland is underactive Hypothyroidism is caused by a decline in thyroid hormone production. Hypo means deficient or less in medical terms. Inflammation and thyroid gland injury are the two primary causes of hypothyroidism.
* Obesity, low heart rate, increased temperature sensitivity, neck swelling, dry skin, hand numbness, hair issues, heavy menstrual cycles, and intestinal problems are some of the symptoms. If not treated, these symptoms can escalate over time.

**APPLICATION**

* The proposed framework will take input in the form of dataset and then forward to the preprocessing module. In the preprocessing module, the normalization of images is performed in this module.
* After preprocessing the images, augmentation is performed. In augmentation, the dataset is divided into two parts: the training dataset and the testing dataset. After the augmentation process, import AlexNet and compare it with the customized AlexNet, and meet the criteria and store it in a trained model .
* The missing values will be checked in the preprocessing steps. If we detect a missing value, the mean value will replace the value in that column. As the missing value had a data loss of about 91%, that parameter is removed from the dataset.
* We have adapted the dataset to be better processed with the chosen models. Initially, only two columns are removed. In the second step of the methodology, we performed dataset manipulation by undersampling the classes.
* Classes 0 and 1 are highly different in size: class 0 has 2870 samples, while class 1 only contains 293 values. The uneven class representation will cause the accuracy to be very high as machine learning algorithms are sensitive to skewed values. The results will contain many false-positive values, and accuracies will be high compared to the more balanced dataset.
* The last step was to divide the dataset into training and testing datasets. We have kept the traditional spill which is 70 percent of the data used for the training and 30 percent of the dataset used for testing purposes. While keeping in mind the dataset distribution, we have to oversample the majority class (class 0) and undersample the minority class.
* Extracting accurate information for medical purposes is an essential task, and it defiantly helps future medical decisions. Feature selection is made to reduce the dimensionality in the dataset. It removes the irrelevant and redundant entries in the dataset. Hence, it increases the accuracy and improves the results.
* The feature selection identifies the most relevant features for the classification in the classification problems. When raw data is extracted many times, there are missing values in the dataset.
* The primary demographics contain information regarding the diseased patient’s age, gender, medication, patient condition, and hormone levels like TSH, T3, and TT4 and category. The classification will contain two classes. Class 0 is negative, and class 1 is positive. Normal means that the patient is not suffering from thyroid disease.

##### Preprocessing Steps

(1)Data cleansing

(2)Data processing

(3)Data elimination

(4)Data wrangling

* Preprocessing is arranged to overcome the different processing issues involving noisy data, redundant information, and missing values. The high quality of data will produce high-quality results according to the measuring metrics. The cost of computations will also reduce.

**CONCLUSION**

* Thyroid Detection using Machine Learning is a project idea that aims a smart and precise way to predict thyroid disease. We have made use of logistic regression algorithm to train our dataset and to predict thyroid disease with more accuracy.
* Here the machine is trained to detect whether the person normal, hyper- user enters data in web app the data will be processed

hypothyroidism based on the user’s input. So whenuser enters data in web app the data will be processed

in backend (model) and the result will be displayed on the screen.

* Our objective was to give society an efficient and precise way of machine learning which can be used in applications aiming to perform disease detection.
* Further development can be do by using image processing of ultrasonic scanning of thyroid images to predict thyroid nodules and cancer, which cannot be recognized in blood test report.
* By combining both the results, thyroid disease prediction can cover all thyroid related diseases.

**FUTURE SCOPE**

* The study showed that they could diagnose thyroid functional disease with we use eight machine learning techniques. We were represented by (Decision Tree (DT), Support Vector Machine (SVM), Random Forest (RF), Naive Bayes (NB), Logistic Regression (LR), Linear Discriminate Analysis (LDA), K-Nearest Neighbor (K-NN), and multi-layer perceptron (MLP)).
* This study assists to doctors and medical staff in the healthcare field. We were also able to compare the eight algorithms and which one could reach the best accuracy. We predicted and classified thyroid disease by applying machine learning techniques to a data set consisting of 1250 actual samples.
* We divided the dataset as follows: 30% of the data were used for training, and 70% were used for testing. After applying these techniques to dataset one that consists of all the characteristics, the random forest algorithm obtained an accuracy rate of 98.93%.
* In the second step, and based on a previous study, we deleted a set of features which are 1- query\_thyroxine 2- query\_hypothyorid 3-query\_hyperthyroid. We applied machine learning techniques to this data, and the MLP algorithm got the highest accuracy of 95.73%.
* The results obtained in this study help us in the rapid prediction of thyroid disease. And the classification of the disease (Hyperthyroidism or Hypothyroidism).

**APPENDIX**

