

Prediction of Thyroid Disease (Hypothyroid) In Early Stage Using Feature Selection and Classification Techniques

Dr.Syeda Asra, Sonu V Bukkan

Dr.Syeda Asra, Dept. computer science & Technology Sharnabasava University Kalburgi

Sonu V Bukkan Dept. computer science & Technology Sharnabasava University Kalburgi

Abstract - There is a high rate of thyroid disease among Bangladesh, which is one of the most common diseases among females in the country. Hypothyroidism is a common variation of thyroid disease. It is clear that hypothyroidism is more common in women. Most people don't know about the disease that is rapidly turning it into a critical health problem. It's important to detect it in its early stages so that doctors can provide better medication to help the problem stay in a manageable state. Predicting thyroid disease is important because it can help doctors determine the best course of treatment

Key Words: Prediction Of Thyroid Disease, Thyroid disease, early stage of detecting thyroid,

1.INTRODUCTION :

In the modern stage, thyroid is the most integral part of the frequent disease among women. In Bangladesh, according to experts, 50 million humans suffer from thyroid disease.

Of these, women are 10 times more likely to suffer from thyroid disease. Although about 50 million humans are affected by thyroid disease, about 30 million of them are not fully aware of the Society's (BES) depiction that about 20-30% of women are suffering from thyroid disease.

Thyroid is a gland located in the middle of the neck in our body. It is butterfly shaped and small in size. It controls many activities throughout the body. Thyroid hormone is responsible for metabolism, growth, sexual function and the preservation of mood. Depending on the secretion of thyroid hormones, we may feel tired or restless and may also lose weight.

There are two fundamental thyroid hormones: triiodothyronine (T3) and thyroxine (T4). Broadly speaking, both these

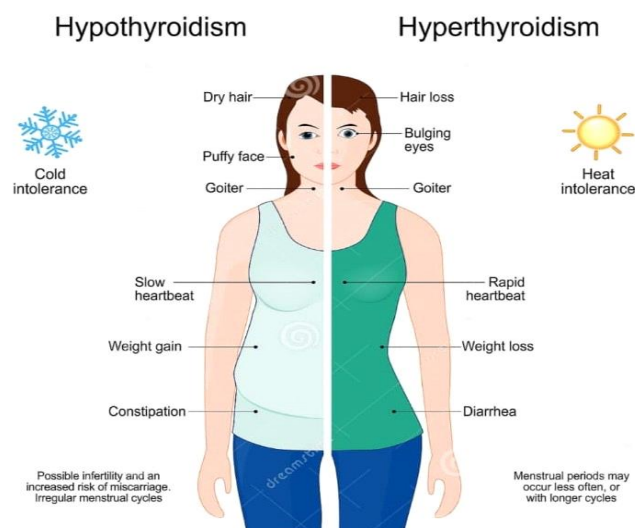
hormones are responsible for maintaining strength in our body. Thyroid stimulating hormone (TSH) is produced through the pituitary gland which helps the thyroid gland to release T3 and T4. There are two common thyroid diseases

1) Hypothyroid

2) Hyperthyroid.

Hypothyroid: When the thyroid gland does not produce enough thyroid hormone, the degree of T3 and T4 decreases and the level of TSH becomes elevated. The symptoms it provides are weight loss, fatigue, talent fog etc.

Hyperthyroid: When the thyroid gland produces more thyroid hormones than our body needs, T3 and T4 levels will become too high and TSH levels will drop. Its symptoms are hair loss, anxiety, sweating etc., In our research, we have focused on hypothyroidism when you consider that this is what occurs most often in women in Bangladesh. Therefore, our lookup is typically targeted at detection, so our research is often not focused on detecting hypothyroid in the early stage.



Nowadays, computing devices have become a very popular means of detecting more than a few diseases. It is very easy and wonderful to predict diseases using laptop having knowledge of techniques. Here, we have used specific determination and classification methods to predict hypothyroid at major stage. We have collected the data from a registered analysis center in Dhaka, Bangladesh. In total, we have collected the correct number of statistics with a total of 9 features. Of these figures, 77% are women while the rest are men. We basically use three function determination techniques called Recursive Feature Elimination (RFE), Univariate Feature Selection (UFS) and Principal Component Analysis (PCA), as well as Support Vector Machine (SVM), Decision Tree, Logistic Regression (LR)) use terrible classification algorithms. Random Forest (RF) and Naive Bayes (NB). We eventually concluded that the RFE feature selection technique helps us achieve better accuracy with any classification approach used.

Therefore, our research is generally aimed at detecting hypothyroid in the seminal stage. Nowadays, desktop mastering has become an extremely popular means of detecting many diseases. It is very easy and good to assume diseases using desktop to get knowledge of techniques. Here, we have used specific judgment and classification methods to predict hypothyroid at critical stage. We have collected data from a registered disease diagnostic center in Dhaka, Bangladesh. In total, we have collected real detailed information with a total of 9 features. Of these figures, 77% are women while the discounters are men. We typically use three feature choice methods called Recursive Feature Elimination (RFE), Univariate Feature Selection (UFS) and Principal Component Analysis (PCA), as well as Support Vector Machine (SVM), Decision Tree, Logistic Regression (LR)) use terrible classification algorithms. , Random Forest (RF) and Naive Bayes (NB). We ultimately concluded that the RFE characteristic resolution approach helps us to achieve high accuracy with any classification technique used.

II. LITERATURE REVIEW

[1] A. K. Aswathi, and A. Antony, "An Intelligent System for Thyroid Disease Classification and Diagnosis", 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT), pp. 1261-1264, 27

Sept,2018.

Abstract:

Data mining technology performs an essential function in healthcare companies such as making choice, diagnosing disease and providing higher treatment to victims. The thyroid gland performs the primary function of maintaining the metabolism of the human body. Data mining in the fitness care enterprise gives systematic access to clinical data. Thyroid diseases are the most common today. Initial changes in the thyroid gland will not affect the desired function of the gland. By early detection of thyroid disorders, better treatment can be supplied at an early stage, resulting in an alternative treatment of thyroid and avoidance of thyroid removal to some amount disease. This paper proposes an approach to the classification and analysis of thyroid disease whereby the user is struggling with disease description and healthy advice. Support vector machine is used for classification. Particle swarm optimization is applied to optimize the SVM parameters. A window is provided to the user to enter the values of small print like TSH, T3, T4 etc. Some values may also be missing while the user comes in. The K-nearest neighbor algorithm is used to estimate the deficient values in the individual inputs.

[2] A. Begum, and A. Parkavi, "Prediction of thyroid disease using data mining techniques", 5th International Conference on Advanced Computing & Communication Systems (ICACCS)PP. 342-345, June 06, 2019.

Abstract:

Classification based data mining plays an important role in many health services. In the field of health care, it is an important and difficult task to diagnose health conditions at an early stage and provide desirable treatment of the disorder. There are many diseases that can be diagnosed early and treated at an early stage. For example, thyroid disease. General approaches to diagnosing thyroid diseases depend on clinical examination and several blood tests. The main task is to detect the disease diagnosis at early stages with greater accuracy. Data mining strategies serve an essential function in the field of healthcare for decision making, disease diagnosing and offering high medical attention to victims at low cost. Thyroid

disease classification is an important task. The reason for this finding is to predict thyroid disease using specific classification methods as well as to find the TSH, T3, T4 correlation in the direction of hyperthyroidism and hypothyroidism and additionally, to explore the correlation towards hyperthyroidism with gender. Hypothyroidism.

[3] Pavya.K, and Srinivasan.B, "Feature selection algorithms to improve thyroid disease diagnosis", IEEE International Conference on Innovations in Green Energy and Health Care Technologies (ICIGEHT'17). 1-5, November 2, 2017.

Abstract:

Correct and early diagnosis of diseases is important and essential in the healthcare industry for correct and timely treatment. This reality is more important in diseases like thyroid, which is very difficult to notice because its symptoms match those of many diseases. The use of computers gaining knowledge of algorithms for thyroid disease diagnosis is key. A standard thyroid disease diagnosis gadget uses three major steps, namely, task extraction, feature judgment, and classification. The primary objective of this paper is to analyze the use of filter-based (F-score) and wrapper-based (recursive feature elimination) attribute selection algorithms on its impact on disease identification and examinations the evaluation is additionally done with the principle component dimensionality reduction algorithm. The performance was compared against three metrics, accuracy, sensitivity and specificity. Four classifiers, namely, multilayer perceptron, back propagation neural network, support vector machine and extreme learning machine, were used to analyze the chosen algorithm. Experimental effects confirmed that each F-score and recursive feature elimination expanded the overall performance of thyroid disease diagnosis, with the wrapper-based algorithm producing the most effectiveness and a maximum accuracy of 98.14% with the ELM classifier.

[4] F. Saiti, A.A. Naini, M.A. Shurehdeli, and M. Teshnehlab, "Thyroid disease Identification based on genetic algorithms using PNN and SVM", Third International Conference on Bioinformatics and Biomedical Engineering, paper published. 1-4, July 14,09.

Abstract:

The thyroid gland produces thyroid hormones to help regulate the body's metabolism. Abnormalities of thyroid hormone production are divided into two categories. Hypothyroidism which is associated with insufficient thyroid hormone production and hyperthyroidism related to excessive thyroid hormone production. It is very important to differentiate these two diseases for the diagnosis of thyroid. Therefore, aid vector machines and probabilistic neural networks have been proposed for classification. These techniques rely on powerful classification algorithms to deal with redundant and irrelevant features in most cases. This paper argues feature choice as an essential hassle with forecasting and demonstrates that GA presents a simple, standard and effective framework for making decisions on the exact subset of elements leading to better forecast rates. We do. Thyroid disease datasets are taken from the UCI machine learning dataset.

[5] Q. Pan, Y. Zhang, M. Zuo, L. Jiang, and D. Chen, "Improved Ensemble Classification Method of Thyroid Disease Based on Random Forests", 8th International Conference on Information Technology in Medicine and Education. 567-571, July 13, 2017.

Abstract:

Thyroid disease is already the second largest in the field of endocrine, and the classification of the disease is the major problem in scientific treatment. In Computer Aided Prognosis (CAD), laptops have been used to assist scientific experts in a wide selection of learning strategies. This paper proposed a new approach to thyroid disease classification based primarily on random forest. Principal component analysis is used to maintain the variability in the data. The rotation transformation can increase the discrepancy of the base classifier and increase the accuracy of the ensemble classifier. Our technology performs much better than Bagging, Random Forest, and AdaBoost, and can overcome the accuracy-variety dilemma. Experimental effects show that the classification accuracy of this approach can reach as high as 95.63% on datasets from UCI laptops to know repositories. In addition to confirming the effectiveness of the method, this paper additionally selects the actual clinical medical data set. It is more complex than the UCI

popular dataset in record extent and dimension. Compared with different methods, the classification accuracy of our approach reaches 96.16%.

[6] A. Tyagi, R. Mehra, and A. Saxena, "Dialogue of Thyroid Disease Prediction System Using Machine Learning Techniques", 5th IEEE International Conference on Parallel, Distributed and Grid Computing (PDGC-2018). 689-693, 27 June 2019.

Abstract:

Thyroid disease is one of the most important objects of formation in clinical analysis and prediction, the introduction of which is a challenging axiom in scientific reason. Thyroid gland is play the important roles of our body. The secretions of thyroid hormones are responsible for controlling the metabolism. Hyperthyroidism and hypothyroidism are two of the common diseases of the thyroid that release thyroid hormones to help regulate the body's metabolic charges. This method where we applied to make the data primitive for analysis to show the risk of patients receiving thyroid. Computer learning plays a pivotal role in disorder prediction techniques and this paper handles the assessment and classification models being used in thyroid disease solely based on data collected from datasets taken from the UCI Desktop data. It is important to ensure a first-class knowledge base that can be used as a hybrid model in solving complex mastering tasks, such as in scientific analysis and predictive. In this paper, we also proposed a kind of laptop mastering technique for thyroid prevention and prognosis. The machine learning algorithm, Support Vector Laptop (SVM), K-NN, Decision Trees was used to estimate the patient's estimated risk of developing thyroid disease.

[7] S. Dash, MN. Das, and B.K. Mishra, "Implementation and classification of model prediction of hypothyroid disease risk", International Conference on Inventive Computing Technologies (ICICT), pp. 1-4, January 19, 2017.

Abstract:

Data mining mainly based classification is one of the important role in the field of healthcare. Diagnosis of health prerequisites

is a very important and challenging undertaking in clinical science. There are many types of diseases in scientific science. Thyroid disorder is one of those essential diseases which is a very serious problem and affects the health of human beings. Thyroid disease classification is one of the essential problems in clinical science because it is immediately related to the health status of the human body, this type of disease can be find out by ideal discovery and careful solved. This paper focuses on the survey of diagnoses of thyroid. There are more than a few authors who have worked concerning the classification of thyroid diseases and have supplied classification accuracy with robust models. This research also focuses on the many strategies that can be applied for the classification of thyroid data.

[8] VS. Vairale, SUKALA.S"Classification of Hypothyroid Disorder Using Optimized SVM Method", Second International Conference on Smart Systems and Inventive Technology (ICSSIT 2019), pp. 258-263, February 10, 2020.

Abstract

Hypothyroidism is an endocrine disease where the thyroid organ does not supply enough thyroid hormone. It is one of the frequent diseases prescribed in women. Detection of hypothyroidism requires prompt evaluation and appropriate clinical examinations to encourage medication. Accurate and early detection of disease is additional important and imperative in the healthcare sector to facilitate accurate and prompt diagnosis and timely treatment. The facts generated in the healthcare sector are massive, important and complex with few parameters. Interpreting and identifying such large data and obtaining accurate and applicable data from it is a difficult and difficult task. However, there is a need and importance to facilitate the victims with highly scientific solutions. This will help in reducing the cost, time and providing more relief to the users through applying advanced and advanced knowledge. This will also help prevent similar complications. Proposed learning of how to derive information from hypothyroid datasets to predict the degree of disease. To be aware of the level of hypothyroid disorder, we used 4 classification computing device to learn techniques, specifically KNN (K-nearest neighbour), SVM (Support Vector Machine), LR (Logistic Regression) and NN (artificial neural network). The

experimental results contrast with the classification accuracy of the four methods. The logistic regression approach yielded 96.08% accuracy among the different three classifiers. But, SVM has been found to be great classifier after standardization of facts and parameter tuning with an accuracy of 99.08%.

[9] K. Shankar, S.K. Laxmanprabu, D. Gupta, A. Masaleno, V.H.C.D. Albuquerque, "The optimal feature-based multi-kernel SVM method for thyroid disorder classification", Springer Science + Business Media, LLC, Phase of Springer Nature 2018, pp. 1128-1143, July 2, 2018.

Abstract

Thyroid diseases are spread all over the world. In India too, an integral problem has arisen due to the fact of this disease. Various research studies estimate that around four million people in India suffer from the ill consequences of "thyroid diseases". Diagnosing fitness conditions in medical science is life and full of challenge. Our proposed model is to classify this thyroid record using ideal feature selection and a kernel-based classifier process. We will build the classification model and its crew performance for information classification using the "Multi Kernel Support Vector Machine". The novelty and goal of this proposed efficacies as trait selection, it is used to beautify the performance of the classification process with the aid of advanced gray wolf adaptation. The reason for this most important function resolution as unimportant aspects from the unique dataset and increase the performance of the model computationally. The proposed thyroid classification results in accuracy, sensitivity and specificity of 97.49, 99.05 and 94.5% compared to the current model. This overall performance measure is calculated from the confusion matrix, which consists of a number of measures in contrast to male or female fashion and in addition to existing classifiers and optimization techniques.

[10].p. Duggal, and s.shukla, "Prediction of Thyroid Disorders Using Advanced Machine Learning Techniques", 10th International Conference on Cloud Computing, Data Science & Engineering (Sangam), pp. 670-675, 09 April 2020

Abstract:

The paper offers several methods of task judgment and classification for thyroid disorder diagnosis, related to gaining knowledge of classification problems from a computing device. Two frequent diseases of the thyroid gland, which releases thyroid hormones to regulate the body's metabolic charges, are hyperthyroidism and hypothyroidism. The classification of these thyroid diseases is an important task. An important perturbation of pattern cognition is extracting or selecting a function set, which is preserved in the pre-processing stage. The proposed methods of feature determination are univariate selection, recursive feature elimination and tree-based feature selection. Three classification methods have been used notably Naive Bayes, Support Vector Machine and Random Forest. The results indicate that support vector machines are the most accurate method and were therefore used as a classifier to separate the signs and symptoms of thyroid diseases into four classes named hypothyroid, hyperthyroid, sick euthyroid and euthyroid (negative).

III.OBJECTIVES

- The basic aim of improving a gadget that can predict what type of thyroid disease a patient is affected by.
- Predicting thyroid disease with the use of a minimal number of parameters.
- This tool will lead to a huge reduction in misdiagnosis as it is able to differentiate between thyroid gland problems and other diseases in the body.
- Also giving the ability to detect disease before it turns into a more unfavorable anomaly.
- Provide an efficient answer for healthcare practitioners using support vector machines to diagnose/classify thyroid disease that may additionally occur in a character

IV.REQUIREMENT

SOFTWARE REQUIREMENTS

- Technology: python
- IDE: anaconda
- Server-Side Technologies: python
- Operating System: Microsoft Windows/Linux

HARDWARE REQUIREMENTS

- System: Pentium i3 Processor.
- Hard Disk :500 GB.
- Monitor: 15'' LED
- Input Devices: Keyboard, Mouse
- Ram: 4G

V. EXISTING SYSTEM

Our strategy basically proposes a feature determination method for the prediction of hypothyroidism and a manifold to realize hypothyroidism at the most critical stage. Various associated methods have been observed over the years and some of them are discussed here. A texture-based segmentation i.e., two parallel versions of a code for fine needle aspiration cytology thyroid snap shots is the most essential first step in locating a fully automated computer aided design solution. The code is developed in MPI version to take advantage of computer assets like PC cluster

VI. PROPOSED SYSTEM

Their ultimate aim in this proposed work is to implement a coronary heart diagnostic system to reduce the number of wasted echocardiograms and to prevent the pregnancies of newborns who are actually affected by heart disease. This paper used the optimized SVM method for classification of hypothyroid disorders. In this work, he proposed a way to use classification desktop study techniques to detect the degree of hypothyroid disease, specifically KNN (KNearest Neighbor), SVM (Support Vector Machine), LR (Logistic Regression), and NN (artificial neural network). The logistic regression approach achieved 96.08% accuracy among the various three classifiers, although SVM gives the highest accuracy of 99.08% **after standardization of records and parameter tuning.**

VII.METHODOLY

Data set and process: data set is collecting all sicked people data and data will process and send to next level of the architecture

Splitting Data: In this mode data splitting valid and invalid data it taken only used and process them, this level of data will split into types training data and test data

Trained data test data: The trained data will stored the information of standard values of the thyroid disease and test data will comparing collected data into trained data that comparison values will be depend on the level of thyroid.

Classification: The classification is KNN (knearest neighbor) and SVM (suport vector machine) and logistic regression methods are achieved

evolution and result: when all stages are compared it give the thyroid negative or positive result.

VIII. SYSTEM ARCHITECTURE

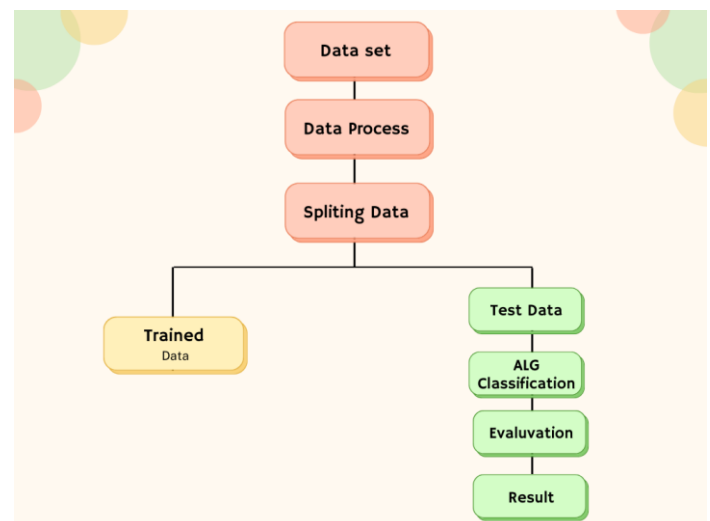
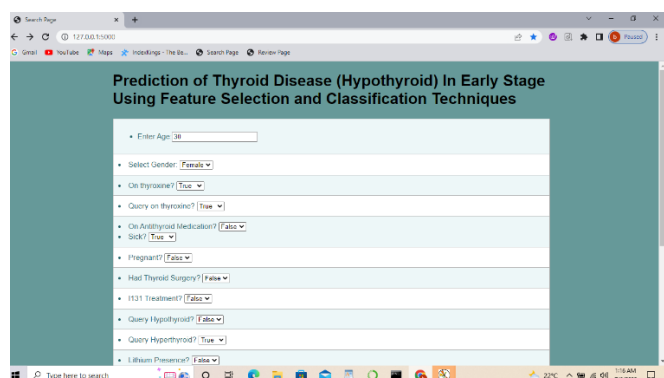


Fig: system architecture

SCREENSHOT:



PREDICTION
Your Thyroid Prediction : Negative

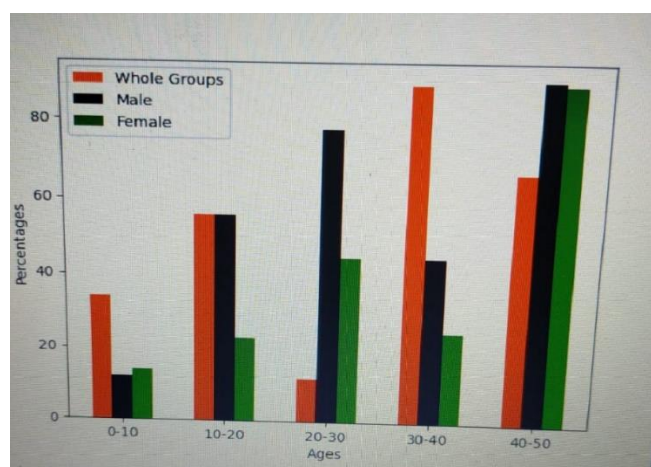
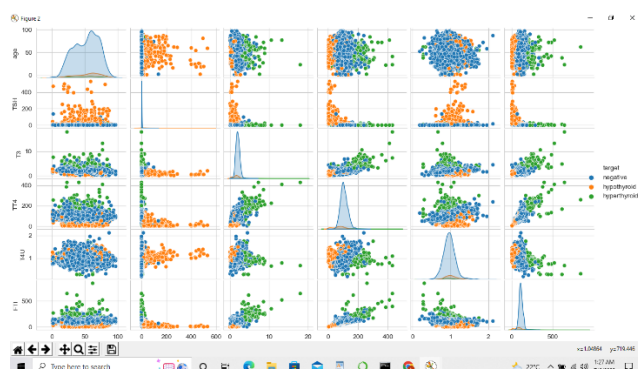


Fig. age chart

IX. TESTING

1) Unit Testing

Unit testing includes testing of each unit or male or female aspect of the software program application. This is the first degree of functional testing. The goal behind the unit effort is to validate the unit components along with its performance. A unit is a testable part of a software program machine and the software is tested in the development section of the software. The purpose of Unit Check Out is to take a look at the correctness of the remote code. An entity component is an individual function or code of the application. White field try out method used for unit check out and usually accomplished through developers. Whenever the application is equipped and delivered to the test engineer, he shall independently or by means of one begin to test each element of the module or modules of the application, and this method is called Unit Testing or are accepted as trying components. Out

2) Integration Testing

Integration try out is the 2D degree of software checking out technique which comes after unit testing. In this test, individual components of the equipment or software are tested in a group. The focus of the integration check out level is to uncover defects at the time of interaction between built-in elements or units. Unit uses modules for testing purpose, and these modules are combined and tested in integration tests. Software is developed with many software modules which are coded by a kind of coder or programmer. The purpose of integration checking is to take a look at the correctness of communication between all modules.

3) Software Testing

Black Box Testing (Nature of Input -> System -> Nature of Output)

Black box testing is a technique of software testing

that checks the functionality of the software without looking at its internal structure or coding. The primary source of black box testing is a specification of the requirements stated by the customer.

In this method, the tester selects a function and gives input values to check its functionality, and checks whether the function is giving the expected output. If the function produces correct output, it is passed to the test, otherwise failed. The test team reports the results to the development team and then tests the next task. After all functions are tested, if there are serious problems, it is returned to the development team for improvement.

4) White Box Testing (Non-Functional: Non-Parametric Details)

Box testing approaches to software testing include black box testing and white box testing. We are discussing here White Box Testing, also known as Glass Box Testing, Structural Testing, Clear Box Testing, Open Box Testing and Transparent Box Testing. It tests the internal coding and checking predefined inputs and desired outputs. It is based on the inner workings of an application and revolves around testing the internal structure. This type of testing requires programming skills to design the test cases. The primary goal of white box testing is to focus on the flow of inputs and outputs through the software and to strengthen the security of the software.

- The term 'white box' is used because to inner working of the system The name clear box or white box or transparent box refers to the software's ability to see through its outer shell into its inner workings.
- Developers do white box testing. In this, the developer will test every line of code in the program. Developers do white-box testing and then send the application or software to the testing team, where they will do black box testing and verify the application with the requirements and identify the bug and send it to the developer.

- The developer fixes the bug and does a round of white box testing and sends it to the testing team. Here, fixing the bug means that the bug has been removed, and the particular feature is working fine on the application.

X.ADVANTAGE

- Separation in time of image acquisition and analysis,
- Exam field time saving
- Less operator dependent
- increased confidence in identity
- Possibility of remote consultation
- Convenient presentation for practitioner and for teaching
- Possibility of inclusion in the information base of patients with future analysis of recent options

XI.DISADVANTAGE

- Experiencing Associate in Nursing intolerance to cold temperatures
- Experiencing anxiety, irritability and nervousness.
- Having bother sleeping.
- Gain weight
- Having muscle weakness and tremors

XII. APPLICATION

Thyroid Prediction may be a project concept aimed at providing a good and accurate method for predicting thyroid malignancy. We have used the supplied regression formula to train our dataset and further development will be done by the abused image process of supersonic scanning of thyroid images to predict thyroid malignancy with added accuracy.

XIII.FUTURE ENHANCEMENT

We have a data problem to work with. Therefore, in the future, we would like to work with a larger dataset and we hope that additional humans from our United States will show interest in

working on this disorder in a position to make us find a better answer that prediction of disease. In key stage with better accuracy. Hope it will give us to maintain a healthy society.

XIV. CONCLUSION

We see that the specific decision approach RFE helps us to achieve higher accuracy with all other classifiers. In our findings, we have observed that RFE helps us in predicting hypothyroid in the seminal stage using real-time datasets. It is very difficult for us to collect data in this current pandemic situation. As a result, we have collected only 519 data. Therefore, we could not learn on a large dataset considering the conditions and constraints. In our study, we have seen that no work has been done before in the thyroid based in Bangladesh. We have a data constraint to work with. Therefore, in the future, we want to work with a larger dataset and we hope that more and more people in our United States will show interest to work on this disease which will help us find a higher answer and major The stage will be able to predict the disease. with high accuracy. Hope it will help the people of our country to maintain a healthy society.

REFERENCES:

- [1] A. M. Amiri, and G. Armano, "Early Diagnosis of Heart Disease Using Classification And Regression Trees", In The 2013 International Joint Conference on Neural Networks, pp. 1-4, 09 //127.0.0.1:5000/
- [2] A. K. Aswathi, and A. Antony, "An Intelligent System for Thyroid Disease Classification and Diagnosis", 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT2018), . 1261-1264, 27 September, 2018.
- [3] A. Begum, and A. Parkavi, "Prediction of thyroid Disease Using Data Mining Techniques", fifth International Conference on Advanced Computing & Communication Systems ICACCS), pp. 342-345, 06 June, 2019.
- [4] K. Pavya, and B. Srinivasan, "FEATURE SELECTION ALGORITHMS TO IMPROVE THYROID DISEASE DIAGNOSIS", IEEE International Conference on Innovations in Green Energy and Healthcare Technologies(ICIGEHT'17), . 1-5, two November, 2017.
- [5] F. Saiti, A. A. Naini, M. A. ShOOREHDLIr, and M. Teshnehlab, "Thyroid Disease Diagnosis Based on Genetic Algorithms the usage of PNN and SVM", 3rd International Conference on Bioinformatics and Biomedical Engineering, pp. 1-4, 14 July, 2009.
- [6] image
https://stock.adobe.com/in/images/Hyperthyroidism-and-Hypothyroidism.-Female-with-Signs-and-symptoms-of-different-/369674920?as_campaign=TinEye&as_content=tineye_matc h&epi1=369674920&tduid=3bd8d8ac8e1bb4787bed533aec4a5612&as_channel=affiliate&as_campclass=redirect&as_source=arvato