



International Journal for Innovative Engineering and Management Research

A Peer Reviewed Open Access International Journal

www.ijiemr.org

COPY RIGHT



ELSEVIER
SSRN

2022 IJIEMR. Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 25th Jun 2022. Link

[:http://www.ijiemr.org/downloads.php?vol=Volume-11&issue= Spl Issue 05](http://www.ijiemr.org/downloads.php?vol=Volume-11&issue= Spl Issue 05)

DOI: 10.48047/IJIEMR/V11/SPL ISSUE 05/24

Title Thyroid Disease Classification and Prediction System

Volume 11, SPL ISSUE 05, Pages: 157-162

Paper Authors

Mr.K.Venkatesh, Penumudi Pravanya Lakshmi, Kalimpudi Sandhya, Maganti Bhavani, Samudrapu Sree Lakshmi Sruthi



USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per **UGC Guidelines** We Are Providing A Electronic Bar Code

Thyroid Disease Classification and Prediction System

Mr.K.Venkatesh¹,Penumudi Pravanya Lakshmi², Kalimpudi Sandhya³,Maganti Bhavani⁴,Samudrapu Sree Lakshmi Sruthi⁵

¹Asst. Professor, Dept of CSE, ²18ME1A0579, ³19ME5A0502, ⁴18ME1A0561, ⁵18ME1A0597

Ramachandra College of Engineering,A.P., India

venkatesh.kondaveti@rcee.ac.in, pravanya222@gmail.com,

sandhyakalimpudi502@gmail.com, bhavanimaganti7979@gmail.com,

sssruthi13@gmail.com

Abstract:

Thyroid disease is one of the leading causes of the emergence of medical diagnosis and prognosis, the beginning of which is a problematic axiom of medical research. The thyroid gland is unique of the most significant organs in our body. Thyroid hormone secretion is involved in the regulation of metabolism. Hyperthyroidism and hypothyroidism are single of the two most common disorders of the thyroid gland, releasing thyroid hormones to regulate the body's metabolism. We applied cleansing techniques to data to generate sufficient primitive data to perform an analysis showing the risk of patient developing the thyroid gland. Machine learning plays an important role in the disease prediction process. This describes the analytical and Using data gathered from the UCI machine learning repository, classification models for thyroid illness are applied. It's crucial to have a solid knowledge base that can be fixed and used as a hybrid model while tackling challenging learning tasks like: This article also proposed various machine learning techniques and diagnoses to prevent the thyroid gland. Machine learning algorithms, support vector machines(SVMs), ANNs, and decision trees were used to predict the estimated risk of a patient developing thyroid disease.

Keywords:- Thyroid Disease, PredictionModel, Machine Learning Algorithms

Introduction

The thyroid gland generates three hormones – Tri - iodothyronine(T3), Tetraiodothyronine (thyroxine or T4) and Calcitonin. T3 and T4 are proper thyroid hormones. These are built as the follicular epithelial cells of the thyroid. Iodine is essential for these hormones.

As Our body cannot generate this trace element, so we must try to get it our diet. Iodine is absorbed into the bloodstream from the diet, and is taken to the thyroid gland to generate thyroid hormones. The pituitary gland instructs the thyroid gland either to release more or to release less hormones in the bloodstream. Few thyroid hormones are included to send proteins into the blood. T3 and T4 are released from the proteins in the blood when the body requires adequate

hormones. Calcitonin, the third hormone is made by C-cells. It involves in calcium and metabolism of Bone. Hyperthyroidism are included to send proteins into the blood. T3 and T4 are released from the proteins in the blood when the body requires adequate hormones.

Calcitonin, the third hormone is made by C-cells. It involves in calcium and metabolism of Bone. Hyperthyroidism (An overactive thyroid) occurs if the thyroid gland makes more hormones. Hypothyroidism (An underactive thyroid) occurs when the gland does not have enough hormones. These imbalances lead to high count of symptoms. Thyroid diseases can be classified based on Euthyroidism, hyperthyroidism and

hypothyroidism indicating Thyroid hormone concentrations that are healthy, too much, or wrong. The country When a person has euthyroidism, their thyroid gland produces thyroid hormones normally and their cellular levels are normal. The clinical manifestation of hyperthyroidism is increased circulation and intracellular thyroid hormones.

The absence of thyroid hormone production is primarily to blame for the condition hypothyroidism. and poor alternate therapy. Many people would like to know about How thyroid affects the certain age group or search for treatment to cure the disease, etc. we can try to get answers by implementing some Machine Learning Methods and process the data with certain methodologies. Machine learning is being used in the medical field and is also helpful for diagnoses of different diseases. Many Researchers have been done to detect and cure the Diseases. We can detect Thyroid Diseases by applying Classifiers like Decision Tree, Random Forest, XGBoost, etc on the Dataset. These Classifiers provide results more accurately for the Thyroid Diseases.

Related Work:

In literature, numerous research cope with the identity of thyroid illnesses way to using hormonal parameters and private records of the affected person, which include age and sex. Notably, a few research use device getting to know category and prediction fashions, even as different techniques use deep neural community fashions. In the primary group, we discover the paintings of Izdihar and Bozkus [2] who used a dataset from the UCI repository to classify thyroid disorder the usage of the selection tree algorithm. In particular, they have got advanced a device getting to know device for the analysis of thyroid illnesses, known as MLTDD able to making an wise forecast of thyroid gland illnesses. This examine

indicates an typical accuracy of 98.7% and 99.8% for testing.

Authors in [4,5], additionally recognition on device getting to know strategies which include Support Vector Machine (SVM), Multiple Linear Regression, Nave Bayes, Decision Trees, to carry out a comparative analysis of thyroid disorder. Their consequences (precision is same to 99.23%) display that selection timber have the great overall performance and may be used efficaciously as an resource with inside the detection of thyroid disorder. In [16] and [5], the authors carried out a examine, the aim of which become to expect thyroid disorder the usage of one-of- a-kind records mining strategies and discover the correlation among TSH, T3, T4 traits with hyperthyroidism or hypothyroidism. In particular, the authorstaken into consideration KNN, Nave Bayes, Support vector device, ID3 as records mining algorithms, making use of them to the general public dataset of the UCI records archive. On the alternative hand, with inside the examine [7], the authors used neural networks (MLP, PNN, GRNN, FTDNN, CFNN) to diagnose sorts of thyroid disorder. More mainly they carried out a examine on 244 topics affected by one-of- a-kind pathologies to analyze the kingdom in their thyroid, taking into consideration a few hormonal parameters and the affected person`s age. The consequences of this studies display that the neural community gives very particular answers, classifying accurate thyroid pathologies primarily based totally on hormonal parameters. Authors in [6] carried out a examine geared toward diagnosing hyperthyroidism and hypothyroidism, the 2 maximum common thyroid disorders. . Many people would like to know about How thyroid affects the certain age group or search for treatment to cure the disease, etc. we can try to get answers by implementing some Machine Learning Methods and process the data with certain methodologies. Machine learning is being used in the medical field and is also helpful for diagnoses of different diseases. Many Researchers have been done to detect and cure the Diseases. We can detect Thyroid Diseases by applying Classifiers like Decision Tree, Random Forest, XGBoost, etc on the Dataset. These Classifiers provide results more

accurately for the Thyroid Diseases.

The category become done the usage of strategies, multinomial logistic regression fashions and neural networks. The studies become carried out on 310 sufferers, or even on this case, the fashions took as input demographics and hormonal parameters. The consequences confirmed higher overall performance of the neural community model (with a median accuracy of 96.3%) than multinomial logistic regression (with a median accuracy of 91.4%) in all cases. This paper differs from the above mentioned techniques due to the fact it's miles the primary (to the great of our knowledge) to recognition at the thyroid disorder remedy prediction. Therefore, the principle goal of this paintings is to apply all of the records collected time beyond regulation on a affected person to expect whether or not the LT4-primarily based totally remedy wishes to be multiplied or decreased. According to this aim, on this examine, we use and examine one-of-a-kind device getting to know strategies to expect the direction of care of sufferers affected by thyroid disorder. Moreover, this examine proposes a brand new set of functions diagnosed on the idea of the endocrinologist's revel in withinside the affected person's remedy. Finally, any other contribution of our paintings is represented via way of means of using a dataset received via way of means of extracting actual records. This dataset is constructed via way of means of integrating sub-datasets including.

Problem Statement

A . Existing System

In the Existing System, the reports of the Patient are diagnosed by the Doctor and many doctors should observe all the reports which might take more time, also some of the details in the Report might be missed

or neglected and might increase Patient at a High Risk. Some of the models are also done based on Multilayer Classifier, Linear Discriminant Analysis (LDA), KNN (K-nearestneighbours), etc. are used to Detect Thyroid Diseases.

B. Proposed System

On studying the thyroid problems, we trained and prepared classifiers that can depict the kind of thyroid problem patient suffers. We will use six datasets from

<https://archive.ics.uci.edu/ml/datasets/Thyroid+Disease>. The Proposed System is to collect all the data required for detecting Thyroid Disease from various datasets and merge them together and perform compare with various other Classifiers like Decision Tree, Random Forest were used to get more accurate results.

Architecture:

Architecture of Thyroid Prediction System

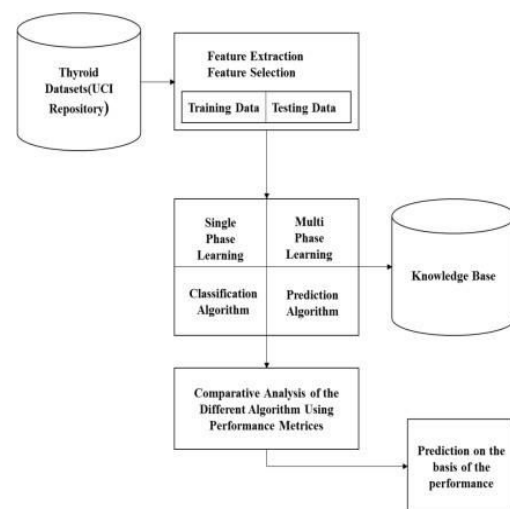


Fig 1: Architecture of thyroid prediction system

Methodology:

Supervised Learning is an

information mining company for deriving features from named training information. The training information consisted of a series of preparatory illustrations. In managed customization, each case consists of an information input object (usually a vector) and a desired output value (also known as a monitor tag). You can use the to create a new illustration[9]. The ideal extension takes calculations into account and effectively sets the class name when hidden. This should be included in the calculation to "wisely" summarize the hidden situation from the training information. On studying the thyroid problems, we trained and prepared classifiers that can depict the kind of thyroid problem patient suffers. We will use six datasets from.

A. Attributes Used to diagnose the thyroid disease:

We found that frequently used medical attributes to perform experimental work for the diagnosis of thyroid disease. The below table describes that necessary attributes for every researcher to perform work for thyroid disease diagnosis.

Table1: Commonly needed data to perform diagnosis

Attributes	Description
Age	In years
Sex	Male or Female
TSH	Thyroid-stimulating Hormone
T3	Triiodothyronine
T4U	Thyroxine utilization rate
TT4	Total Thyroxin

FTI	Free Thyroxin Index
TBG	Thyroid binding globulin

Table2: Shows the features contained in the dataset

Attributes	Value type	Clarification
Age	Number	1,12,8...
Gender	1,0	1= yes, 0= No
query_thyroxine	1,0	1= yes, 0= No
on_antithyroid_medication	1,0	1= yes, 0= No
on_thyroid_medication	1,0	1= yes, 0= No
thyroid_surgery	1,0	1= yes, 0= No
sick	1,0	1= yes, 0= No
pregnant	1,0	1= yes, 0= No
query_hypothyroid	1,0	1= yes, 0= No
query_hyperthyroid	1,0	1= yes, 0= No
TSH measured	1,0	1= yes, 0= No
TSH	Analysis ratio	Numeric value
T3 measured	1,0	1= yes, 0= No
T3	Analysis ratio	Numeric value
T4 measured		1= yes, 0= No
T4	Analysis ratio	Numeric value

category	0,1,2,3	0=normal,1=primary,2=secondary,3=compensated
----------	---------	--

B. Data Machine Learning techniques

Decision Tree: It is a decision support tool and uses a tree structure model of decisions and it predicts the value of a target variable using some simple decision rules inferred from the 2 data features. It requires some data preparation. It helps to handle both categorical and numerical data. On studying the thyroid problems, we trained and prepared classifiers that can depict the kind of thyroid problem patient suffers. We will use six datasets from In the Existing System, the reports of the Patient are diagnosed by the Doctor and many doctors should observe all the reports which might take more time, also some of the details in the Report might be missed or neglected and might increase Patient at a High Risk. Some of the models are also done based on Multilayer Classifier, Linear Discriminant Analysis (LDA), KNN (K-nearestneighbours), etc. are used to Detect Thyroid Diseases.

Random Forest: It is a classification of algorithm which has multiple decisions trees. While building each individual tree Random Forest uses bagging and feature randomness to create an uncorrelated forest of trees which helps to get more accurate results. In the random forest, all the individual trees will divide the class prediction and the class with huge number of votes will be our model's

prediction.

Support Vector Machine: Support vector machine is considered as an assorted research algorithm that helps in performing the analysis in a precise way. Support vector machine is an approach that is commenced with a concept of an ace of separating hyper plane to aid in the distribution for sampling of data. A hyper plane or multiple planes are created by the support vector machine classifier in high dimensional space. The training data samples are being separated as a positive and negative data samples by the hyper plane. 3. Decision Tree: Tree-like graph is used in decision tree classifier. A decision tree is classified by its 3 nodes i.e. internal nodes, leaf nodes, and the root nodes. The internal node connotes as the test on an attribute, the leaf node connotes as the distribution of the class and the root node connotes as the tree that has the top most node. The two most extensive algorithms that are used in the as semblances of a decision tree for diagnostic and prognostic model of thyroid diseases are C4.5 and ID3. Researchers use Decision Tree

Algorithm Used	Accuracy
DT	90.13
SVM	92.53
Random Forest	91.2

Conclusion:

In this thyroid disease detection and classification, we got some very good and accurate classifiers. We have used Decision Tree, Random Forest with these we achieved a good F-Score, and also, we have got good accuracy. which means the age group anyway between 60 to 90 is most likely in being attacked by thyroid disease. If we consider gender women are more likely to suffer from hypothyroidism when compared with men. If we consider young aged people, men are at

higher risk of this disease than women.

This conclusion shows that age and gender are two important considerations. From our observations of , the risk of thyroid disease increases with age, causing thyroid disease. In our future work, we need to list age and gender as important factors influencing thyroid disease. In general, different patients have different factors that contribute to thyroid disease. The dataset used has some restrictions, like Geography climate, food, etc. Another problem is the small data set, as only a small part of the population suffered from thyroid disease.

References:

- [1] Izdihar Al-muwaffaq and Zeki Bozkus. MLtd : Use of machine learning techniques for diagnosis of thyroid gland disorder. 2016
- [2] A. Begum and A. Parkavi. Prediction of thyroid disease using data mining techniques. In 2019 5th International Conference on Advanced Computing Communication Systems (ICACCS), pages 342–345, 2019.
- [3] Irina Ionit and Liviu Ionit. Prediction of thyroid disease using data mining techniques. BRAIN. Broad Research in Artificial Intelligence and Neuroscience, 7(3), 2016.
- [4] S. Razia, P. Swathi Prathyusha, N. Krishna, and N. Sumana. A comparative study of machine learning algorithms on thyroid disease prediction. International journal of engineering and technology, 7:315, 2018.
- [5] Ankita Tyagi and Ritika Mehra. Interactive thyroid disease prediction system using machine learning technique.

03 2019.

- [6] Shiva Borzouei, Hossein Mahjub, Negar Asaad Sajadi, and Maryam Farhadian. Diagnosing thyroid disorders: Comparison of logistic regression and neural network models. Journal of Family Medicine and Primary Care, 9:1470, 06 2020.
- [7] Mahdiyari O. Obeidavi M. R, Rafiee A. Diagnosing thyroid disease by neural networks. 2017.
<https://www.medicalnewstoday.com/articles/248679>
- [8] Tyagi, Ankita, Ritika Mehra, and Aditya Saxena. "Interactive thyroid disease prediction system using machine learning technique." 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC). IEEE, 2018
- [9] <https://www.medicalnewstoday.com/articles/248679>
- [10] Y.T. Lo, H. Fujita, T.W. Pai, Prediction of coronary artery disease based on ensemble learning approaches and co-expressed observations, J. Mech. Med. Biol. 16 (01) (2016) 164001
- [11] Tyagi, Ankita, Ritika Mehra, and Aditya Saxena. "Interactive thyroid disease prediction system using machine learning technique." 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC). IEEE, 2018
- [12] Y.T. Lo, H. Fujita, T.W. Pai, Prediction of coronary artery disease based on ensemble learning approaches and co-expressed observations, J. Mech. Med. Biol. 16 (01) (2016) 164001