IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

"Thyroid Prediction Using Machine Learning"

[1]Rohit D. Mane, [2]Samruddhi P. Dixit, [3]Kranti A. Wani, [4] Anjali R. Mali, [5] Sonali B. Mohite

[1]Asst. Professor, [2]Student, [3]Student, [4]Student, [5]Student

[1, 2,3,4,5]Computer Science and Engineering,

[1,2,3,4,5]Dr. J. J. Magdm College of Engineering, Jaysingpur, Kolhapur, Maharashtra, India

---

Abstract: **The thyroid is an endocrine organ situated in the foremost locale of the neck: its principal task is to create thyroid chemicals, which are useful to our whole body. Its conceivable brokenness can prompt the development of an inadequate or extreme measure of thyroid chemical. Hence, the thyroid can become kindled or enlarged because of at least one swellings framing inside it. A portion of these modules can be the site of threatening growths. One of the most utilized medicines is sodium levothyroxine, otherwise called LT4, an engineered thyroid chemical utilized in the treatment of thyroid problems and sicknesses. Expectations about the treatment can be significant for supporting 'endocrinologists' exercises and work on the nature of the patients life. This work, in an unexpected way, plans to anticipate the LT4 treatment pattern for patients experiencing hypothyroidism. For every patient, the clinical history is accessible after some time, and hence based on the pattern of the hormonal boundaries and different traits thought of it as will be feasible to anticipate the course of every patient's treatment to comprehend on the off chance that this ought to be expanded or diminished. To direct this review, we will zero in on utilizing different AI calculations.**
`

IndexTerms – Introduction, Literature Survey, Proposed methodology and Discussion, Algorithms, Conclusion, References

---

I. INTRODUCTION

Thyroid complaint frequentness have been on the rise in recent times. A large number of people are diagnosed with thyroid conditions similar as hypothyroidism and hyperthyroidism yearly. The thyroid gland produces levothyroxine(T4) and triodothyronine(T3) and inadequate thyroid hormones may lead to hypothyroidism and hyperthyroidism. numerous approaches are proposed to descry thyroid complaint opinion in the literature. A visionary thyroid complaint vaticination is essential to duly treat the case at the right time and save mortal lives and medical charges. Due to the technological advancements in data processing and calculation, machine literacy and deep literacy ways can be applied to prognosticate the thyroid opinion in the early stages and classify the thyroid complaint types hypothyroidism, hyperthyroidism, etc.

II. LITERATURE SURVEY

According to YongFeng Wang (2020)[1] Thyroid module is diagnosed for benign type using ultrasound images of thyroid by image analysis. Limitations are Radiomics and deep learning based approaches.

According to Shiva Borzouei, Hossein Mahjub(Journal of Family medicine and primary care)(2020)[2] Comparison of Logistic Regression and Neural Network Model.

According to Ankita Tyagi And Rikita Nehara (2018)[3] They Used Different Classification Methods: SVM,KNN etc. Limitations are They have analysed accuracy of algorithms used and comparison is made to find best technique with high accuracy**.**

According to Sunila Godra(2018)[4] They have used logistic regression and SVM technique to analyze thyroid dataset . Limitations are Logistic regression turned out has best classifier.

According to S.Razia, N.Krishna, N.Suman (2018)[5] A Comparative study of machine learning algorithms on thyroid disease prediction

According to Hitesh Garg.(IEEE 2013)[6] Neural network is used for features extractions and segmentation for ultra sound images. Limitations are The accuracy and other factors were measured and all the average values were above 86%

III. PROPOSED METHODOLOGY AND DISCUSSION

- First we acquire the disease data set from UCI. Data set consist of several thyroid related disease record and many target classes.
- The samples for target classes are few , which are not enough to train models, so we select only those target classes whose samples are more than 500 as a result, we got 5 target classes.
- After selecting target classes for experiment we performed data balancing .
- Normal classes samples were 3771 in total which is more compared to other target class sample so we randomly select only 400 samples for normal class to make data balance.
- It is followed by future selection process where many feature selection techniques are applied.

Python:-

Python is a general purpose, dynamic, high-level, and interpreted programming language. It supports Object Oriented programming approach to develop applications. It is simple and easy to learn and provides lots of high-level data structures. Python is easy to learn yet powerful and versatile scripting language, which makes it attractive for Application Development. Python's syntax and dynamic typing with its interpreted nature make it an ideal language for scripting and rapid application development. Python supports multiple programming pattern, including object-oriented, imperative, and functional or procedural programming styles.

Pandas:-

Pandas is defined as an open-source library that provides high-performance data manipulation in Python. It is built on top of the NumPy package, which means Numpy is required for operating the Pandas. The name of Pandas is derived from the word Panel Data**,** which means an Econometrics from Multidimensional data. It is used for data analysis in Python and developed by Wes McKinney in 2008**.**

NumPy:-

NumPy is mostly written in C language, and it is an extension module of Python. It is defined as a Python package used for performing the various numerical computations and processing of the multidimensional and single-dimensional array elements. The calculations using Numpy arrays are faster than the normal Python array.

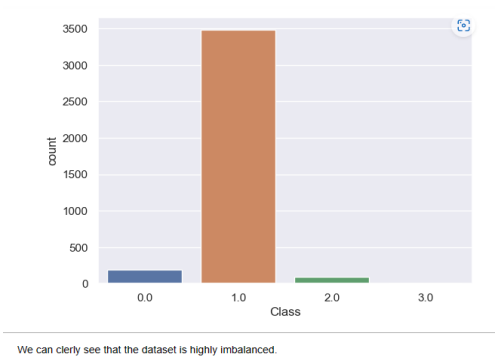


FIG 1:- DATA IMBALANCE

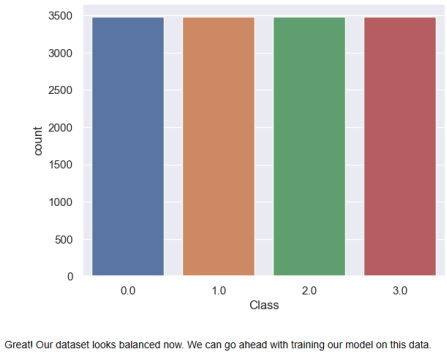


FIG 2:- DATA BALANCED GRAPH

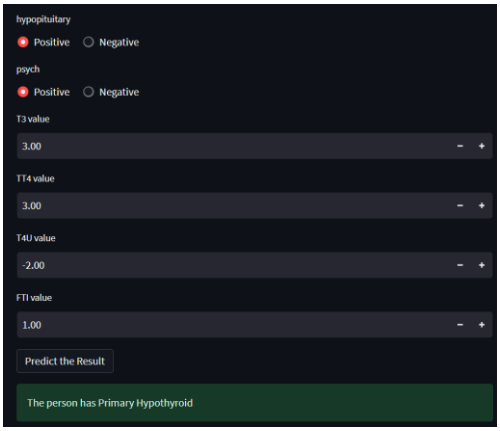


FIG 3:- THYROID PREDICTION

IV. ALGORITHMS

a) KNN ALGORITHM:-

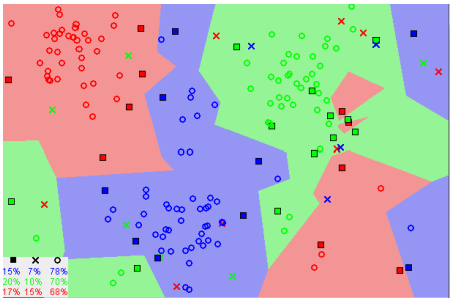
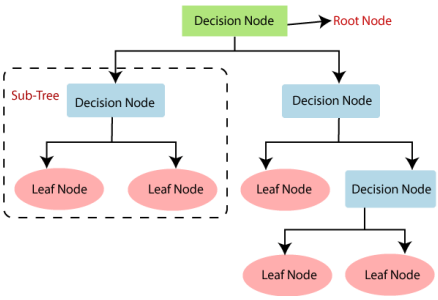


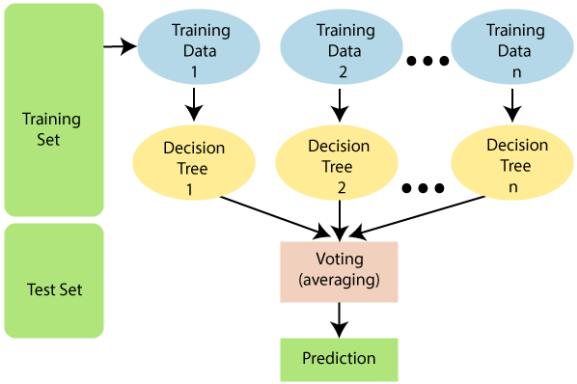
Image above that utmost of the time, analogous data points are close to each other. The KNN algorithm hinges on this supposition being true enough for the algorithm to be useful. KNN captures the idea of similarity with some mathematics we might have learned in our nonage — calculating the distance between points on a graph.

b) DECISION TREE ALGORITHM:-



Decision Tree is a Supervised literacy fashion that can be used for both bracket and Retrogression problems, but substantially it's preferred for working Bracket problems. It's a tree- structured classifier, where internal bumps represent the features of a dataset, branches represent the decision rules and each splint knot represents the outgrowth. In a Decision tree, there are two bumps, which are the Decision knot and Leaf Node. Decision bumps are used to make any decision and have multiple branches, whereas Leaf bumps are the affair of those opinions and don't contain any farther branches. The opinions or the test are performed on the base of features of the given dataset. It's a graphical representation for getting all the possible results to a problem/ decision grounded on given conditions.

c) RANDOM FOREST ALGORITHM:-



Random Forest is a popular machine literacy algorithm that belongs to the supervised literacy fashion. It can be used for both Bracket and Retrogression problems in ML. It's grounded on the conception of ensemble literacy, which is a process of combining multiple classifiers to break a complex problem and to ameliorate the performance of the model. As the name suggests," Random Forest is a classifier that contains a number of decision trees on colorful subsets of the given dataset and takes the average to ameliorate the prophetic delicacy of that dataset." rather of counting on one decision tree, the arbitrary timber takes the vaticination from each tree and grounded on the maturity votes of prognostications, and it predicts the final affair.

V. Conclusion

Thyroid Discovery using Machine literacy is a design idea that aims a smart and precise way to prognosticate thyroid complaint. Our ideal was to give society an effective and precise way of machine literacy which can be used in operations aiming to perform complaint discovery.

VI. Reference

1] Shiva Borzouei, Hossein Mahjub, NegarAsaad 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.

2] Ankita Tyagi and Ritika Mehra. Interactive thyroid disease prediction system using machine learning technique. 03 2018

3] S. Razia, P. SwathiPrathyusha, 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.

4] Mahdiyar O. Obeidavi M. R, Rafiee A. Diagnosing thyroid disease by neural networks. 2017.

5] Izdihar Al-muwaffaq and Zeki Bozkus. Mltdd : Use of machine learning techniques for diagnosis of thyroid gland disorder. 2016

Page size: A4 size only
Text Column: Single **texts align: justify**
Title: 24pt Times New Roman align: centre
Page Margins: Left – 0.51”, Right – 0.51”, Top – 0.75”, Bottom – 0.75”
Font: Use Only Times New Roman for whole paper
Figure caption: Font size- 12”, lower case and Write below the figure, position-center
Table Caption: Font- 12”, lower case and Top of the table, position-center
Paragraph: Paragraph Indentation by- 0.2”
Line Spacing: single
Before: 0” **After:** 0”
Header 0.3” **footer** 0”.