IOSR Journal Of Humanities And Social Science (IOSR-JHSS)
Volume 28, Issue 6, Series 1 (June, 2023) 01-05
e-ISSN: 2279-0837, p-ISSN: 2279-0845.
www.iosrjournals.org

Thyroid Disease Prediction With Machine Learning Approaches

Prof. Manisha D. Rohakale[1] Prof. Komal K. Kawale[2]
[1]*Assistant Professor, New Arts, Commerce and Science College, Ahmednagar.*
[2]*Assistant Professor, New Arts, Commerce and Science College, Ahmednagar.*

Abstract:
Thyroid is one of the most common diseases found in human being which cause many other side effects. It is of two types: Hyperthyroid and Hypothyroid. Thyroid disorders (i. e. Hyperthyroidism, Hypothyroidism, Goitre and thyroid cancer) occur due to dysfunction of the thyroid gland or pituitary gland, iodine deficiency, cancer or due to side-effects from other medications. Some other reasons like pregnancy, or illnesses may also show abnormal levels of thyroid hormones. An early diagnosis and detection of thyroid disease can help human being to fight against this disease. In this study, various machine learning techniques like K-NN, Random Forest, Logistic Classifier and Decision Tree are used to develop model for diagnosis of hypothyroid disease.
Keywords: *Pituitary gland, pleural effusion, flaky, appetite, lousy, impotency, ascites.*

Date of Submission: 22-05-2023 Date of Acceptance: 02-06-2023

I. Introduction:

Knowledge acquisition and management has become the foremost activities of all organizations. Valuable information hides in historical data. We should use the right tools to analyse the past data and to find new facts that could help in taking better decisions to improve the present status or to use the new information to take new steps. Machine learning is a technique to extract useful and interesting facts or knowledge from existing large datasets.

Thyroid is one of the most popular disease across the World is not exception for the India. Various studies indicated 42 million people in India undergoes through thyroid disease. Thyroid is a gland located below the Adam's apple in our neck. It releases hormones, thyroxine (T4) and triiodothyronine (T3), which increase the amount of oxygen and stimulate our cells to produce new proteins. By controlling the release of these hormones, the thyroid determines the metabolic rate of most of our body's organs. The thyroid gland is regulated by thyroid stimulating hormone (TSH), which is produce by the pituitary gland. Normally, when thyroid hormone levels in the body are high, they will "switch off" the production of TSH, which result in stop the production of T4 and T3. Problem occur when the thyroid gland becomes either underactive (hypothyroidism) or overactive (hyperthyroidism). Thyroid problems are more common in women than men.

Hypothyroidism:

The most common thyroid disorder is hypothyroidism. Hypo- means deficient or under(active), so it is a condition in which the thyroid gland is underperforming or producing too little thyroid hormones (i. e.T4 and T3) in the blood. It results in gaining weight, losing appetite and in general, feeling lousy!. It is a disease encompassing all the socio-economical and cultural groups and is more commonly seen in women and people over the age 50.

Symptoms:

General • tiredness, weight gain • hoarseness • high blood cholesterol level.	Skin • dry, flaky skin • hair fall & brittle hair • purplish lips
Cardiorespiratory • slow heart rate • hypertension • pericardial & pleural effusion	Reproductive • scanty periods • infertility(in both male & females) • impotency

Neuromuscular • aches, muscle stiffness	Gastrointestinal • constipation • ascites
Haematological • anaemia	Psychological • depression

Literature Survey:

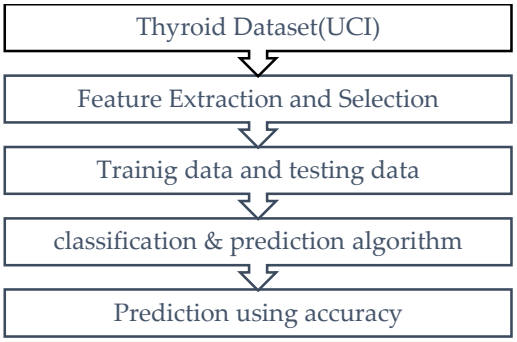

Fig: Machine Learning Algorithms on Data

II. Methodology:

Data Source: We have taken a thyroid disease database from the UCI Repository left by Ross Quinlan at the University of California, Irvine. There are 30 attributes and 3775 records.

The Proposed Feature Model: Of the initial feature set, we select 27 attributes that are related to patient information and thyroid parameters. This selection performed by an expert on the basis of criteria usually used to evaluate the patient's treatment.

Variables:

Sr. no.	Variables	Description
1.	Age	Patient's age on the date of medical check-up.
2.	Sex	Gender of the patient.
3.	Pregnant	Female patient was pregnant or not.
4.	On thyroxine	Patients are given external supplements of thyroxine or not.
5.	Thyroid surgery	Patient has undergone the required surgery or not.
6.	Goitre	Suffer through goitre or not.
7.	TSH	Thyroid stimulating hormone.(Numeric)
8.	T3	Informs about levels of T3 in the patient.
9.	TT4	The levels of TT4 in patients.
10.	T4U	The levels of T4U in patients.
11.	Binary class	Patient is suffering from hypothyroidism or not.

Classifier: To predict the course of treatment to which a patient with a thyroid problem is subjected, we used more than one machine learning classifier. The classifier suggests to the doctor if, based on the historical and current patient's state, the stage of hypothyroidism.

Validation: The metrics we used to validate the model are accuracy, precision, recall, and F-score.

Confusion Matrix:

	Actual Positive	**Actual Negative**
Predicted Positive	True Positive(TP)	False Positive(FP)
Predicted Negative	False Negative(FN)	True Negative(TN)

a) Accuracy = $\frac{(TP+TN)}{(TP+FP+TN+FN)}$

b) Precision = $\frac{TP}{(TP+FP)}$

c) Recall = $\frac{TP}{(TP+FN)}$

d) Fl-score = $\frac{2*(Precision*Recall)}{(Precision+Recall)}$

III. Results:

Risk Associated with Various Factors:

Risk Ratio Table		
Pregnant\Prediction class	Negative	Positive
No Pregnancy	100	0
Pregnancy	97.7953	2.20466

Table. a) Risk ratio for female participants

Risk Ratio Table		
Goitre \ Prediction class	Negative	Positive
No Goitre	8.66564	91.3344
Goitre	0	100

Table. b) Risk ratio for female participants

From the above table a) we can conclude that there is 2.20% chances that after pregnancy mother will suffer through hypothyroidism. From the above table b) we can conclude that if female patient has goitre then she will definitely suffer through hypothyroidism and also if female patient is goitre free then there is 91.33% chances that she will suffer through hypothyroidism.

Risk Ratio Table		
Goitre \ Prediction class	Negative	Positive
No Goitre	5.752212389	94.2478
Goitre	0	100

Table. c) Risk ratio for male participants

From the above table c) we can conclude that if male patient has goitre then he will definitely suffer through hypothyroidism and also if he is goitre free then there is 94.25% chances that he will suffer through hypothyroidism.

Chi Square Independence test:

Feature 1	Feature 2	P-value
Predicted class	Gender	0.022
Predicted class	On thyroxine	0.000
Predicted class	Pregnancy	0.034
Predicted class	Thyroid surgery	0.2788

From the above table:

i. Person had hypothyroidism or not is closely related with the gender of the given patient.
ii. Person had hypothyroidism or not is closely related with the external supplement of thyroxine given to the patient.
iii. Person had hypothyroidism or not is dependent on the pregnancy.
iv. Person had hypothyroidism or not is independent of the thyroid surgery.

Model:

binaryClass =2.280 + (-1.432*TSH) + 1.1172*T3 + 4.178*on.thyroxine + 4.192*thyroid.surgery + 0.2474*TT4

IV. Discussion:

Sr. no.	Classifier	Prediction Class	Accuracy	Precision	Recall	F1-Score
1	Logistic Classifier	0(Negative)	97.35 %	0.95	0.69	0.8
		1(Positive)		0.97	1	0.99
2	Random Forest	0(Negative)	99.74%	0.98	0.98	0.98
		1(Positive)		0.99	0.99	0.99
3	KNN	0(Negative)	95.63%	0.93	0.47	0.62
		1(Positive)		0.96	0.99	0.98
4	Decision Tree	0(Negative)	99.60%	0.97	0.98	0.97
		1(Positive)		0.99	0.99	0.99

Table: Classification report for different machine learning classifiers.

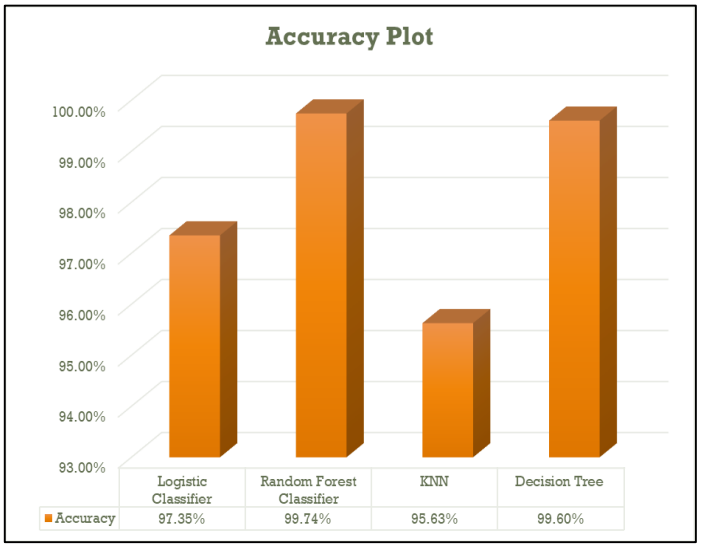

Fig: Accuracy plot for various classifiers.

We can understand that Logistic classifier performed with an accuracy of 97.35%, Random forest classifier has an accuracy of 99.74% and KNN and Decision tree Shows accuracy of 95.63% and 99.60% respectively.

V. Conclusion:

The early diagnosis of a possible malfunction plays a fundamental role, as well as the prediction of a patient with hypothyroidism, which can be of great help for doctors who have patients under treatment. In this study we proposed an approach to predict the thyroid disease. In particular, the proposed model is able to predict the hypothyroidism on the basis of other parameters related to the person being treated, therefore the doctor is facilitated in choosing the treatment to prescribe. The models are tested on an overall dataset containing 755 instances referring to the patients.

Acknowledgements:

I would also like to acknowledge the help and support of my acquaintances and friends.

References

[1]. Shivanee Pandey, Rohit Miri, S. R. Tandan, Diagnosis and Classification of Hypothyroid Disease Using Data Mining Techniques, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY, IJERT, 29-06-2013

[2]. N. Vijayalakshmi, P. Nithya, Information Retrieval from Thyroid Database Through Data Mining, INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING, July 2018

[3]. Suwarna Gothane, Data Mining Classification on Hypo Thyroids Detection: Association Women Outnumber Men, International Journal of Recent Technology and Engineering Volume 8, March 2020

[4]. Agnes Stephanie Harahap,1,2 Desty Gusti Sari,3 Marini Stephanie,1 Alvita Dewi Siswoyo,4 Litta Septina Mahmelia Zaid,1 Diani Kartini,5 Maria Francisca Ham,1,2 and Tri Juli Edi Tarigan6, Clinicopathological Profile of Thyroid Carcinoma in Young Patients: An Indonesian Single-Center Study, 2022 Jan 11
[5]. H. Bhasin, Python Basics Aself teaching introduction, Mercury Learning and Information, New Delhi