JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Thyroid Disease Detection using SVM Algorithm

Sakshi Math[1] ,Anikesh Giri[2],Arati Ugale[3] , Pallavi Sawant[4]
Department of ENTC , SKNCOE , Pune.

***Abstract*- Thyroid disease is clinical disease of high prevalence in the world of population leading cause of medical diagnosis and prediction development which medical science is complicated axiom. One of most used treatments is sodium levothyroxine, also known as LT4, a synthesis thyroid hormone used in the treatment of thyroid disorders and diseases. Prediction about the treatment can be important for supporting endocrinologists' activities and improve the quality of the patients' life. To date, there are numerous studies in the literature that focus on the prediction of thyroid diseases on the trend of hormonal parameters of people. A person can develop problems if their thyroid overproduces or underproduces hormones. These states are known as hyperthyroidism and hypothyroidism. The proposed system predicts the thyroid treatment using different Machine Learning algorithms. In this paper our objective is to predict thyroid disease to aware the patient. Thyroid is predicted in three parameters that is Normal, hypothyroidism, hyperthyroidism In Machine learning various algorithm are used to predict the thyroid disease. SVM is also used to predict the thyroid. In this paper we proposed SVM algorithm that gives the maximum accuracy as compared to other algorithms. SVM gives 95.98 percent of accuracy. The original dataset contains unwanted data with null csv so the new dataset has been created which removes the unwanted data. SVM algorithm which gives maximum accuracy that is used to build the model. Then proposed system uses thyroid disease dataset. We need 30 percent data for testing and 70 percent data is been used for training for training.**

***Keywords* – Thyroid, Hypothyroidism, Hyperthyroidism, Prediction, SVM, Machine Learning.**

I. INTRODUCTION

For thyroid detection there are various techniques used deep learning, image processing, machine learning. This paper proposed machine learning model to detect thyroid. It allows collection of stored patient data for the prediction of the disease. There are prediction algorithms which are available for the diagnosis of the disease. The medical information systems are rich of datasets but there are only few intelligent systems which can easily analysis the disease. Over a period, the machine learning algorithms have started playing a crucial role in resolving the complex and non-linear problems in the developing model. In any disease prediction models are used to override the features that can be selected from different datasets which can be use classification in healthy patient as accurate as possible. If this is not done, misclassification can lead to a healthy patient getting unnecessary treatment.

II. LITERATURE SURVEY

For forecasting Thyroid diseases there are different machine learning algorithms are used.Chandel Khusboo et.al [1] proposed KNN and Naïve Bayes algorithm which worked on 45 patients and prove KNN algorithm gives better accuracy than Naive Bayes. Banu, G. Rasitha et.al [2] proposed J48 algorithm got accuracy 91.85. AKGÜL, Göksu, et.al [3] proposed Nave bayes J48 and random forest and got J48 91 , Random Forest =93.3, Nave bayes =95 . Sindhya, Mrs K et.al [4] proposed K nearest neighbours and SVM and got k nearest neighbours accuracy is 92, SVM accuracy is 95.98 .Vijayakumar, K., et al proposed K Nearest Neighbours and SVM and K Nearest Neighbours got 92, accuracy and SVM got accuracy 95.98 Chaurasia, Vikas et.al [5] proposed Nave Bayes, RBF Network, and J48, J48 algorithm got accuracy 93.41, Naive Bayes algorithm got 92.36 accuracy , RBF Network got 94.77 accuracy. Begum, Amina, and A. Parkavi et.al [6] proposed Naive Bayes, Decision Tree, MLP, and RBF Network Naive Bayes got 91.63 accuracy, Decision Tree got 93.91 accuracy, MLP got 95.15 accuracy, and RBF Network got accuracy 93.03 . Table 1 shows the comparison of different algorithm for prediction model. For these comparison it shows the SVM gives better accuracy than other algorithm.

Sr.NO	Authors	No of Patient	Algorithms	Accuracy
1	Chandel et.al	45	KNN	93.44,
			Naive Bayes	22.56
2	Banu, G. Rasitha	55	J48	91.85
3	AKGÜL, Göksu, et al	66	Nave bayes	95
			J48	91
			Naive bayes	95
			Random Forest	93.3
4	Sindhya, Mrs K et.al	49	KNN	92
			SVM	95.98

Table 1 Related work

III. BLOCK DIAGRAM

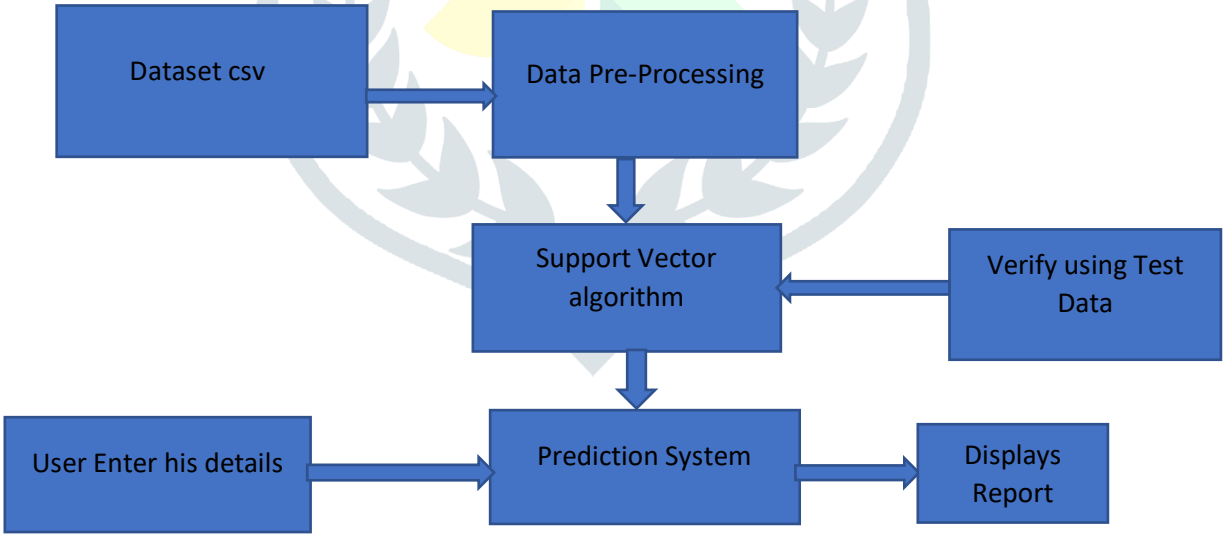

Fig 1. Block Diagram

Dataset has been taken from Kaggle website https://www.kaggle.com/datasets/yasserhessein/thyroid-disease-data-set and operations are performed on the dataset. The pre-processing is done on the data by removing empty elements, question marks, all the un wanted data is removed and then SVM algorithm is applied and data and data has been verified by entering the details of patients prediction system displays the report that whether the patient has hypothyroidism, hyperthyroidism or normal. Dataset has been used from the Kaggle website and we have these following features id, age ,gender etc has been showed in table 2 by considering these parameters we can predict the thyroid disease.

No	Attribute Name	Value Type	Clarification
1	id	number	1,2,3…….,
12	age	number	1,10,20,50,…….,
3	gender	1,0	1=m,0=f
4	query_thyroxine	1,0	1=yes,0=no
5	on_antithyroid_medication	1,0	1=yes,0=no
6	sick	1,0	1=yes,0=no
7	pregnant	1,0	1=yes,0=no
8	thyroid_surgery	1,0	1=yes,0=no
9	query_hypothyorid	1,0	1=yes,0=no
10	query_hyperthyroid	1,0	1=yes,0=no
11	TSH measured	1,0	1=yes,0=no
12	TSH	Analysis ratio	Numeric value
13	T3 measured	1,0	1=yes,0=no
14	T3	Analysis ratio	Numeric value
15	T4 measured	1,0	1=yes,0=no
16	T4	Analysis ratio	Numeric value
17	category	0,1,2	0=normal,1=hypothyroid,2=hyperthyroid

Table 2 Data Features

IV. RESULTS

We have applied our data to a range of machine learning algorithms (Decision Tree, SVM, Random Forest, Naive Bayes, Logistic Regression, Linear Discriminant Analysis, k-Nearest neighbours, Multi-Layer Perceptron) We divided the existing data into two parts, 70% for training and 30% 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 a group of algorithms shown in the table below, and after the application process these results appeared to us. This practical part has been implemented on the python platform and is considered a complete and integrated platform.

NO	Algorithm	Accuracy
1	Decision Tree	90.13
2	SVM	95.38
3	ANN	75.38
4	KNN	93.84
5	Logistic Algorithm	92

Table 3 Accuracy of Algorithms

Table 3 shows the our proposed system output hence we can observe that SVM is better than the other algorithm which gives 95.38% accuracy.

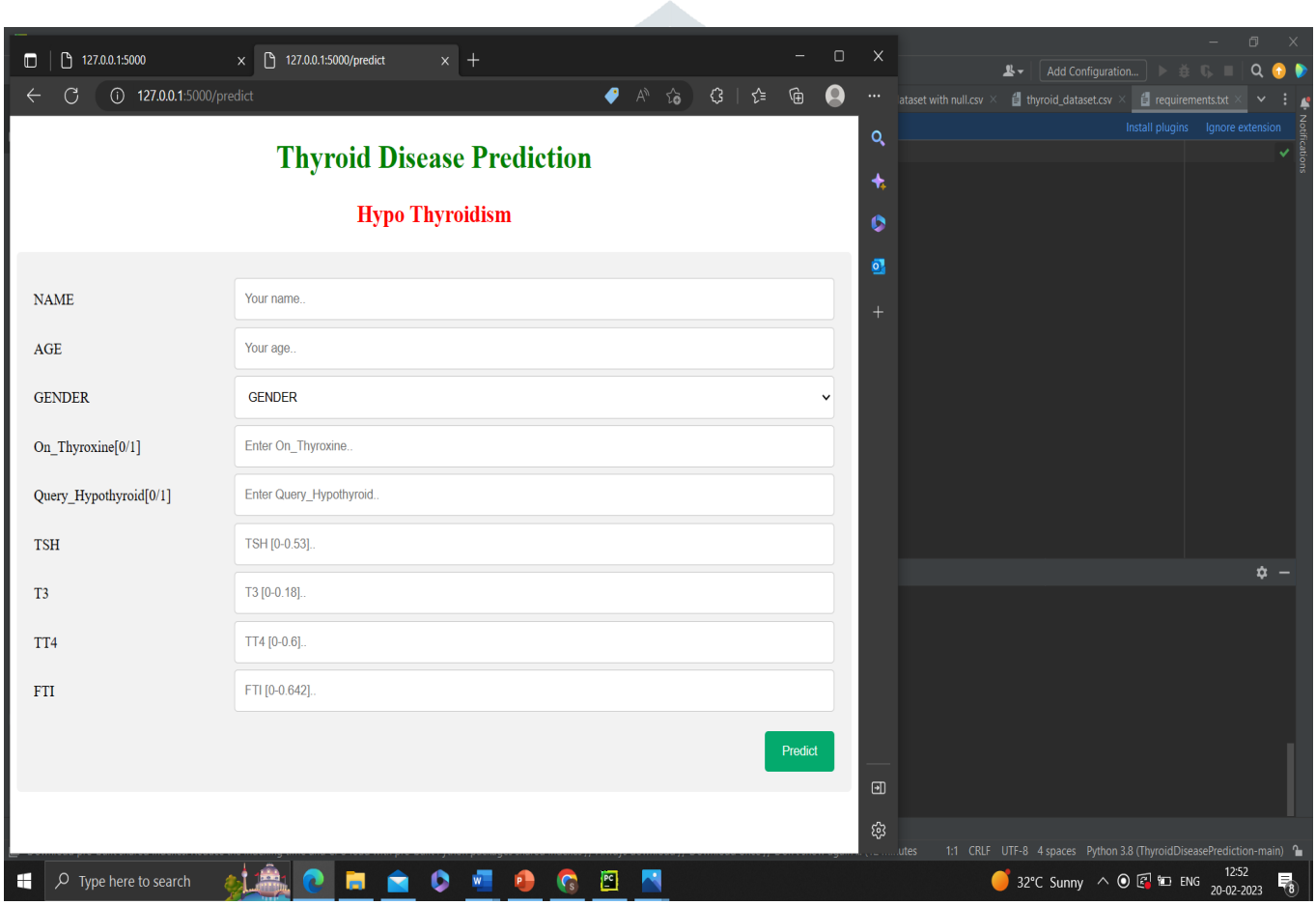

Fig.2 Frontend of the system

We use PyCharm and input the file and python programming after that we can apply different algorithms KNN, ANN, SVM etc among these we got SVM of maximum accuracy. Hence the figure shows the frontend of our system which takes the detail from the patient and gives the output that is hypo, hyper and normal thyroid.

V. CONCULSION

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 Support Vector Machine 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-hypothyroidism based on the user's input. So when user enters data in software the data will be processed in frontend (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.

REFERNCES

[1] Chandel, Khushboo et al 2016 A comparative study on thyroid disease detection using K-nearest neighbor and Naive Bayes classification techniques CSI transactions on ICT 4 313-319 Google Scholar Mining Technique in Diagnosing Thyroid disease International Journal of Computer Sciences and Engineering 4 64-70

Author: Khalid salman, Emrullah Sonuç Publish Year: 2021

[2] Banu G and Rasitha 2016 A Role of decision Tree classification data Mining Technique in Diagnosing Thyroid disease International Journal of Computer Sciences and Engineering **4** 64-70b

[3] AKGÜL, Göksu, et al. "Hipotiroidi Hastalığı Teşhisinde Sınıflandırma Algoritmalarının Kullanımı." Bilişim Teknolojileri Dergisi 13.3 (2020): 255-268.

[4] Sindhya, Mrs K. "EFFECTIVE PREDICTION OF HYPOTHYROID USING VARIOUS DATA MINING TECHNIQUES."

[5]VijiyaKumar, K., et al. "Random Forest Algorithm for the Prediction of Diabetes." 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN). IEEE, 2019.

[6]Chaurasia, Vikas, Saurabh Pal, and B. B. Tiwari. "Prediction of benign and malignant breast cancer using data mining techniques." Journal of Algorithms & Computational Technology 12.2 (2018): 119-126

[7] Begum, Amina, and A. Parkavi. "Prediction of thyroid disease using data mining techniques." 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS). IEEE, 2019.