International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER
SSRN

IJIEMR Transactions, online available on 13th Aug 2022. Link :http://www.ijiemr.org/downloads.php?vol=Volume-11&issue=Issue 08

DOI: 10.48047/IJIEMR/V11/ISSUE 08/06

Title **DETECTION OF THYROID DISEASE USING MACHINE LEARNING TECHNIQUE**

Volume 11, ISSUE 08, Pages: 38-47

Paper Authors
M.Vidyavathi, Dr.P.Chandra Kanth

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

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

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

DETECTION OF THYROID DISEASE USING MACHINE LEARNING TECHNIQUE

M.Vidyavathi,PG Scholar,ASCET,Gudur
E-mail :vidyavathimannarapu@gmail.com
Dr.P.Chandra Kanth,Assoc Prof,Dept of CSE,ASCET,Gudur
E-mail : chandrakanthc4u@gmail.com

ABSTRACT: A complex premise in the field of medicine is that thyroid disorders are the major cause of medical diagnosis and prediction. The thyroid gland is one of our body's most important organs. Metabolism is regulated by the release of thyroid hormones. Both overproduction and underproduction of thyroid hormones affect the body's ability to regulate its metabolism. The application of machine learning in illness prediction and in the study of classification models for thyroid disease based on data from hospital datasets is crucial. To deal with dynamic learning activities like medical diagnosis and predication, it is necessary to ensure, build, and apply a decent knowledge base as a hybrid model. Thyroid may be detected and inhibited using simple machine learning approaches. Predicting the likelihood of a thyroid patient using an SVM model is common practise. Whenever a patient is at risk for developing thyroid disease, our system must propose home remedies, warnings and medicine.

Keywords: Machine learning, classification model, Thyroid diseases, Support vector machines,

1. INTRODUCTION

Advanced machine biology is used in the area of healthcare. It required data to be collected for medical disease prediction. For early-stage disease detection, various intelligent prediction algorithms are used. The Medical Information System is good with data sets, but intelligent systems are not available for the fast diagnosis of diseases. Eventually, machine learning algorithms play a key position in solving complex and non-linear problems during the creation of prediction model. The characteristics that can be selected from the various data sets that can be used as description in a healthy patient as specifically as possible are needed in any disease prediction models. Otherwise, misclassification can result in a good patient receiving inappropriate care. The reality of forecasting any condition associated with thyroid illness is also of the greatest cardinal number. Thyroid gland is endocrine in stomach. It is erected in lowered portion of human neck, under apple of Adam, and assists in secretion of thyroid hormones and which ultimately affects metabolism rate and protein synthesis. To control body metabolism, these hormones count on how quickly heart beats and how quickly calories burn. The composition of thyroid hormones helps to control the body's metabolism. These glands consist of two mature levothyroxine (abbreviated T4) and triiodothyronine thyroid hormones (abbreviated T3). These thyroid hormones are essential for manufacturing and general

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

construction and regulation in order to regulate body temperature. T4 and T3 are exclusively two activated thyroid hormones that usually compose of thyroid glands. These hormones are vital to the control of proteins; distribution at body temperature and energy-bearing and propagation in every part of the body. Iodine, along with T3 and T4 harmones, is the fundamental building block of thyroid glands and is only present in a few uncommon yet extremely common diseases. Inadequate amounts of these hormones cause hypothyroidism, whereas an excessive amount causes hyperthyroidism. There are several causes of hyperthyroidism and underactive thyroidism. There are numerous medications. Thyroid surgery is susceptible to ionising radiation, ongoing thyroid softening, iodine shortage, and loss of thyroid hormone-producing enzyme..

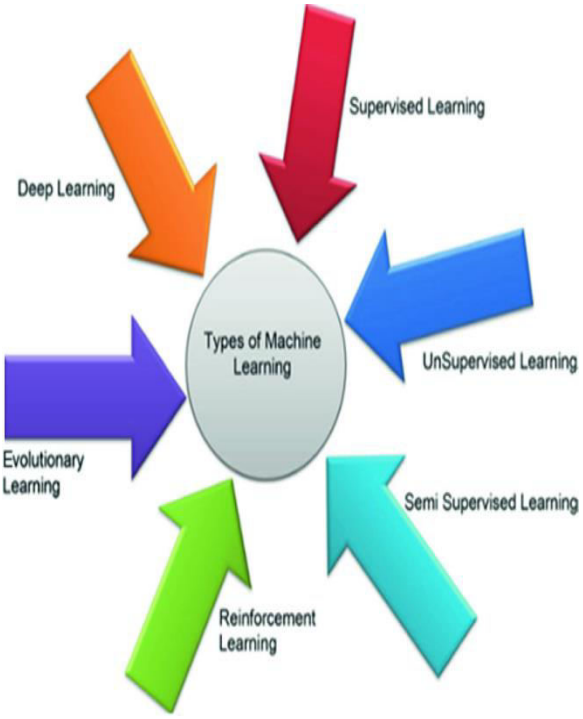


Fig.1: Machine learning techniques for Thyroid detection

2. LITERATURE REVIEW

2.1 Interactive Thyroid Disease Prediction System using Machine Learning Techniques:

Thyroid disease is a primary goal of development in clinical prognosis and prediction, and its start is a difficult axiom in clinical research. The thyroid gland is one of the most important organs in our bodies. Thyroid hormone releases play a role in metabolic regulation. Hyperthyroidism and hypothyroidism are two common thyroid disorders in which thyroid hormones are released to regulate the charge of the body's metabolism.. Data cleaning strategies have been utilized to make the information primitive adequate for performing analytics to exhibit the danger of sufferers acquiring thyroid. The laptop studying performs a decisive position in the procedure of disorder prediction and this paper handles the evaluation and classification fashions that are being used in the thyroid ailment based totally on the data gathered from the dataset taken from UCI laptop studying repository. It is vital to make certain a respectable information base that can be entrenched and used as a hybrid mannequin in fixing complicated mastering task, such as in clinical prognosis and prognostic tasks. In this paper, we additionally proposed one-of-a-kind computing device mastering methods and analysis for the prevention of thyroid. Machine Learning Algorithms, guide vector computer (SVM), K-NN, Decision Trees have been used to predict the estimated threat on a patient's threat of acquiring thyroid disease.

2.2 Comparison Study of Radiomics and Deep-Learning Based Methods for Thyroid Nodules Classification using Ultrasound Images:

Thyroid nodules have a excessive occurrence and a small share is malignant. Many non-invasive strategies have been developed with the assist of the Internet of Things to enhance the detection price of malignant nodules. These techniques can be roughly labeled into two classes: radiomics based totally and deep getting to know primarily based approaches. In general, convolutional neural networks primarily based deep getting to know strategies have executed promising overall performance in many clinical picture evaluation and classification applications; however, no present contrast has been accomplished between radiomics primarily based and deep studying based totally approaches. As a result, the purpose of this work is to assess the overall performance of radiomics and deep studying-based algorithms for the categorization of thyroid nodules from ultrasound pictures. On the one hand, we created a radiomics-based approach for collecting high-throughput 302-dimensional statistical components from pre-processed photos. The dimension discount was then performed using mutual facts and linear discriminant evaluation to reach the final classification. A deep studying-based method, on the other hand, was additionally built and tested by pre-training a VGG16 mannequin with fine-tuning. Ultrasound pics along with 3120 snap shots (1841 benign nodules and 1393 malignant nodules) from 1040 instances had been retrospectively collected. The dataset was once divided into 80% coaching and 20% checking out data. The absolute best accuracies yielded on the trying out information for radiomics and deep studying based totally strategies had been 66.81% and 74.69%, respectively. A evaluation end result tested that the deep studying based totally technique can gain a higher overall performance than the use of radiomics..

2.3 Prediction of Thyroid Disease Using Machine Learning Techniques:

The paper gives a number of techniques of function determination and classification for thyroid ailment diagnosis, associated to the computer studying classification problems. Two frequent illnesses of the thyroid gland, which releases thyroid hormones for regulating the fee of body's metabolism, are hyperthyroidism and hypothyroidism. Classification of these thyroid ailments is a giant task. An necessary hassle of sample cognizance is to extract or pick out function set, which is covered in the pre-processing stage. The proposed techniques of characteristic resolution are Univariate Selection, Recursive Feature Elimination and Tree Based Feature Selection. Three classification methods have been used particularly Naïve Bayes, Support vector machines and Random Forest. Results suggests that the Support Vector Machines are the most correct method and therefore this used to be used as a classifier to separate the signs of thyroid illnesses into four instructions particularly Hypothyroid, Hyperthyroid, Sick Euthyroid and Euthyroid (negative).

2.4 Segmentation of Thyroid Gland in Ultrasound image using Neural Network:

The thyroid gland is exceedingly vascular organ, and lies in the anterior section of the neck simply under the thyroid cartilage. Ultrasound imaging is most typically used to realize and classify abnormalities of the thyroid gland. Other modalities (CT/MRI) are additionally used. There is a assignment to phase ultrasound scientific photo which is regularly blurred and consists of noise as different modalities like CT carries ionizing radiations and expensive. Thus, there is a want to follow a technique to computerized section properly the objects for future evaluation except any assumptions about the object's topology are made. Various techniques or methods are used for automated segmention of thyroid gand however the software of neural community in photo processing offers a higher answer to segmentation problem. In this paper we use Feedforward neural community to classify the area the use of function extraction and then section it. Experiment and outcomes are proven.

2.5 Diagnosis of thyroid disease using artificial neural network methods," in: Proceedings of ICONIP

The proper interpretation of thyroid gland functioning data is critical in the diagnosis of thyroid illness. The thyroid gland's principal function is to aid in the regulation of the body's metabolism. This is provided by thyroid hormone, which is generated by the thyroid gland. The kind of thyroid disease is determined by the production of too little thyroid hormone (hypothyroidism) or too much thyroid hormone (hyperthyroidism). Various neural network methods were utilised in this study to aid in the detection of thyroid illness.

2.6 A novel hybrid method based on artificial immune recognition system (AIRS) with fuzzy weighted preprocessing for thyroid disease diagnosis,"

The assessment of thyroid gland functioning data is critical in the diagnosis of thyroid illness. The thyroid gland's principal function is to aid in the regulation of the body's metabolism. This is provided by thyroid hormone, which is generated by the thyroid gland. The kind of thyroid disease is defined by the production of too little thyroid hormone (hypothyroidism) or too much thyroid hormone (hyperthyroidism). Artificial immune systems (AISs) are a relatively new but highly successful branch of artificial intelligence.

2.7 Thyroid Disease Diagnosis Based on Genetic Algorithms Using PNN and SVM

Thyroid gland produces thyroid hormones to help the regulation of the body's metabolism. The abnormalities of producing thyroid hormones are divided into two categories. Hypothyroidism which is related to production of insufficient thyroid hormone and hyperthyroidism related to production of excessive thyroid hormone. Separating these two diseases is very important for thyroid diagnosis. Therefore support vector machines and probabilistic neural network are proposed to classification.

2.8 An investigation of neural networks in thyroid function diagnosis

We are investigating the utility of artificial neural networks in the diagnosis of thyroid illnesses. A cross-validation method is used to test the robustness of neural networks to sample fluctuations. We show how neural networks and classical Bayesian classifiers are related. Neural networks can produce strong estimates of posterior probability and so outperform standard statistical methods such as logistic regression in classification. The neural network models are also demonstrated to be resistant to fluctuations in sampling. It is proved that neural networks can be a promising classification method for practical usage in medical diagnosis situations where the data is frequently severely imbalanced.

2.9 Clinical epidemiology in the era of big data: new opportunities, familiar challenges.

Routinely collected health data have progressed from being merely byproducts of health care delivery or billing to a significant research tool for analysing and improving patient care via clinical epidemiologic research. In the context of epidemiologic research, big data refers to enormous interconnected data sets within a single country or networks of transnational databases. Several Nordic, European, and other transnational collaborations have now established themselves. Big data benefits clinical epidemiology by improving estimate precision, which is especially significant for comforting ("null") findings; the ability to conduct meaningful analysis in subgroups of patients; and the early discovery of safety warnings.

2.10 Evaluation of Predictive Machine Learning Techniques as Expert Systems in Medical Diagnosis

The medical research business possesses a massive amount of data; unfortunately, a major chunk of this data is not mined. Machine Learning pushes analytics to its limits by uncovering hidden knowledge in data. The primary goal of a medical decision support system is to aid clinicians in making informed decisions. Statistical Analysis/Methods: Machine Learning approaches such as K-Nearest Neighbors, Decision Tree, Artificial neural networks, Radial Basis Function neural networks, and Support Vector Machine are examined in this research effort.

2.11 A Novel Weighted Class based Clustering for Medical Diagnostic Interface

Medical Decision Support System (MDSS) is a diagnostic interface which provides computer assisted information retrieval as well as may support excellence decision making, to stay away from human error. Even if human decision-making is frequently most advantageous, but it is poor when there are vast amounts of data to be classified. Also capability and accuracy of decisions will decrease when humans are set into pressure and massive work. Forever there is a need and scope for a better MDSS.

3. IMPLEMENTATION

In the prediction process, machine learning plays a key role, and paper research and

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

the classifications of models used in thyroid disease are based on information from UCI machine learning repositories. A decent knowledge base that can be centered and used as a hybrid paradigm must be preserved in order to address complex learning issues, such as medical diagnostics and statistical tasks. We also proposed different approaches for machine learning and thyroid diagnosis. Machine Learning Algorithms, Vector Support Machine, were used to calculate an estimated probability of a patient having thyroid disease.

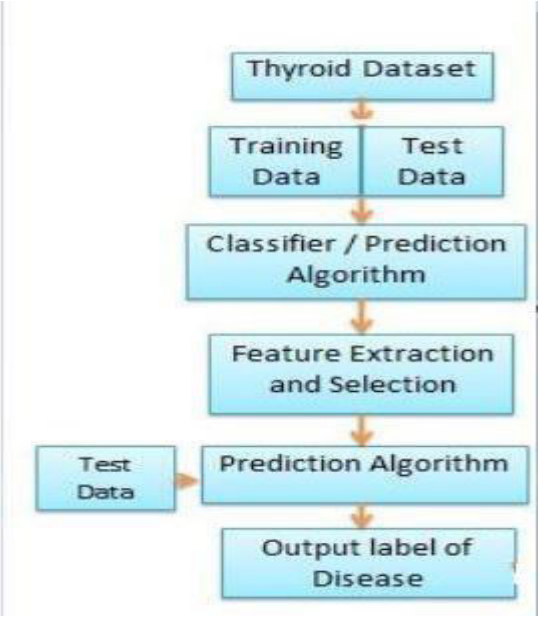


Fig.2: Workflow diagram

4. ALGORITHMS

SVM:

A **support vector machine** (SVM) is machine learning algorithm that analyzes data for classification and regression analysis. SVM is a supervised learning method that examines data and categorises it into one of two groups. An SVM generates a map of the sorted data with the margins between them as wide as possible.

Support Vector Machine, or SVM, is a prominent Supervised Learning technique that is used for both classification and regression issues. However, it is mostly utilised in Machine Learning for Classification difficulties. The SVM algorithm's purpose is to find the best line or decision boundary that can divide n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

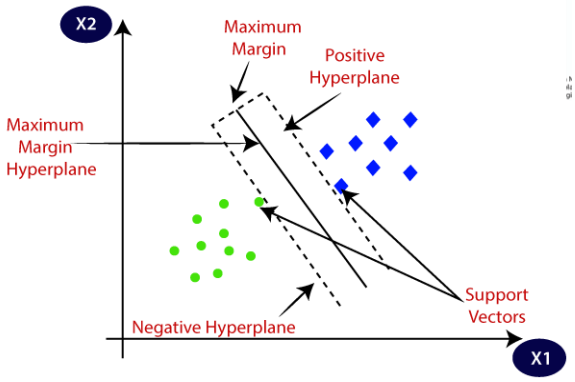


Fig.3: SVM model

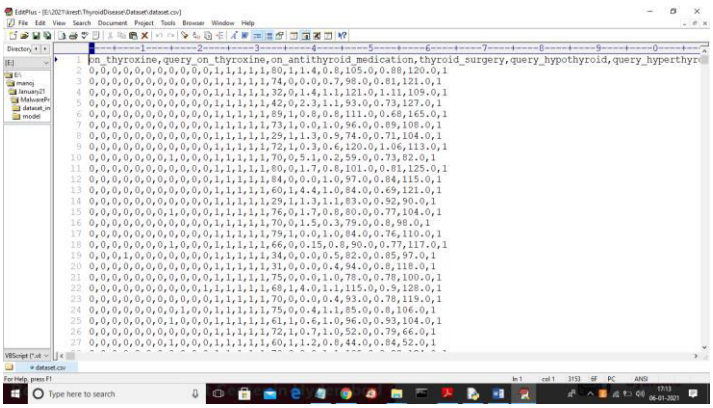
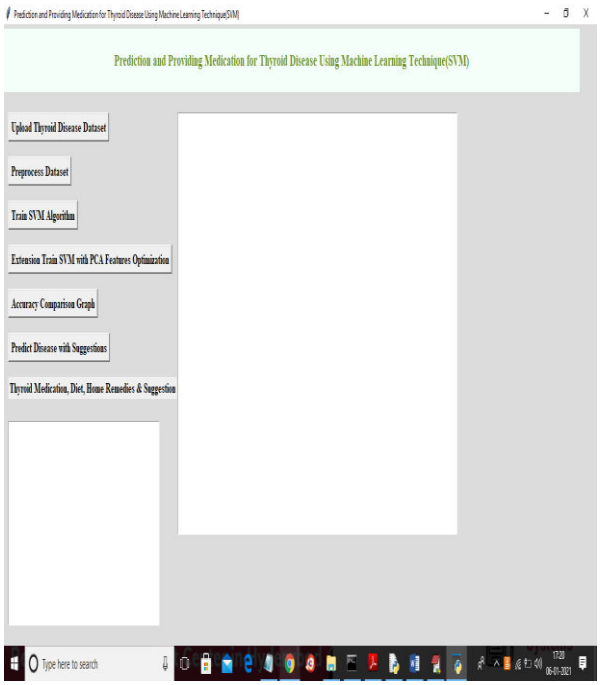


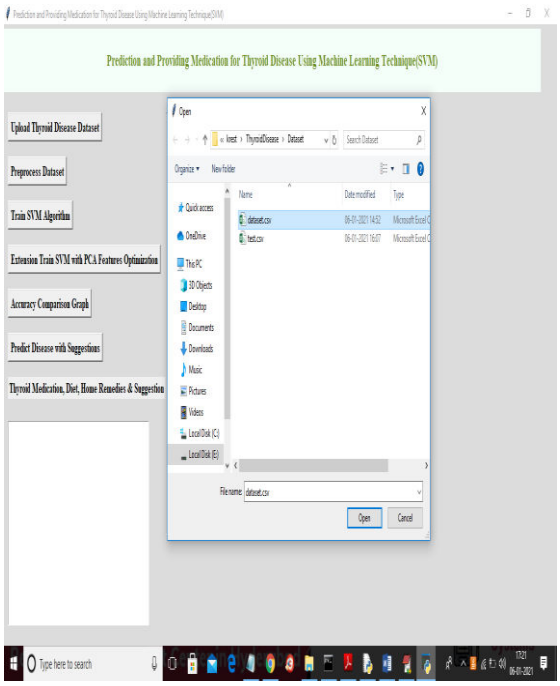
Fig 4:Data Set Values

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

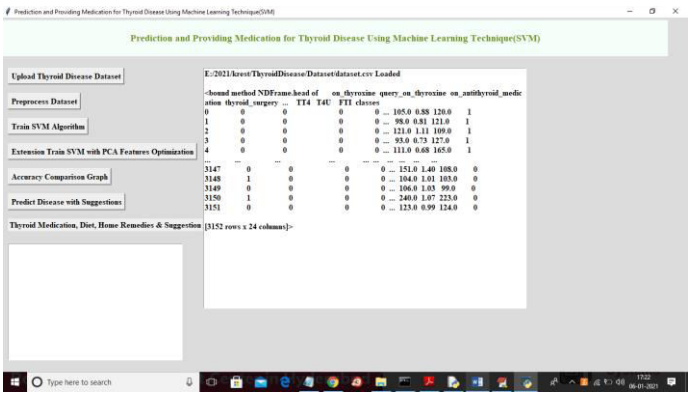
5. EXPERIMENTAL RESULTS



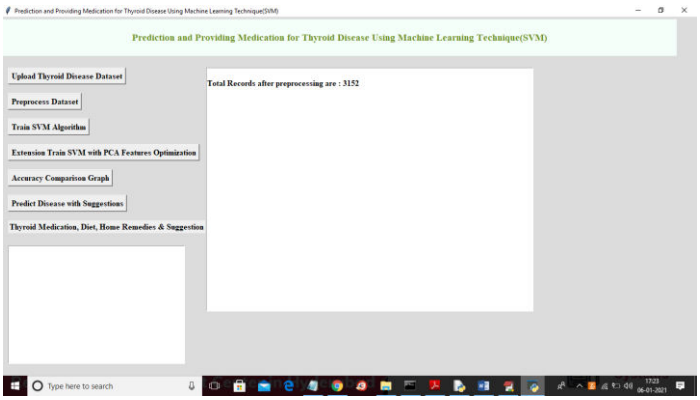
In above screen click on 'Upload Thyroid Disease Dataset' button to upload dataset and to get below screen



In above screen selecting and uploading 'dataset.csv' file and then click on 'Open' button to load dataset and to get below screen

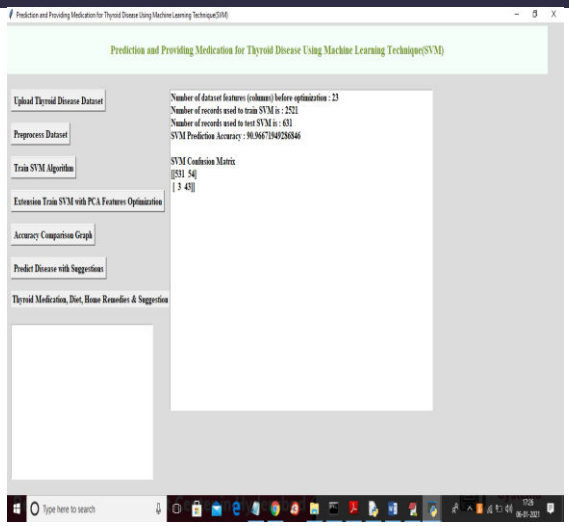


In above screen dataset loaded and displaying few records from dataset and then click on 'Preprocess Dataset' button to remove missing and NAN values from dataset and to separate X and Y values where X contains all dataset values and Y contains class label value.

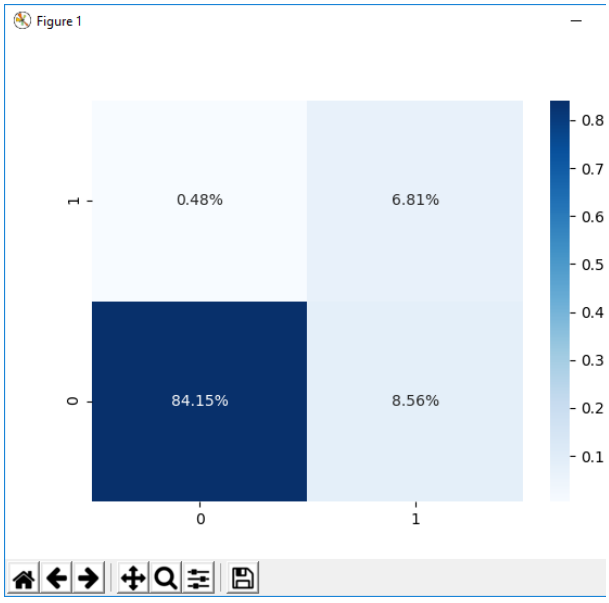


In above screen dataset showing 3152 preprocess records and now dataset is ready and now click on 'Train SVM Algorithm' button to split dataset into train and test and then apply SVM algorithm on train data to generate model and then model will be applied on test data to calculate prediction accuracy

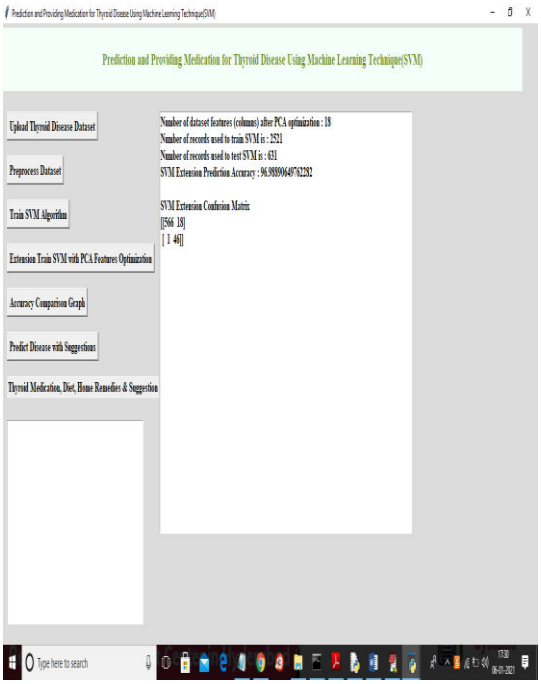
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org



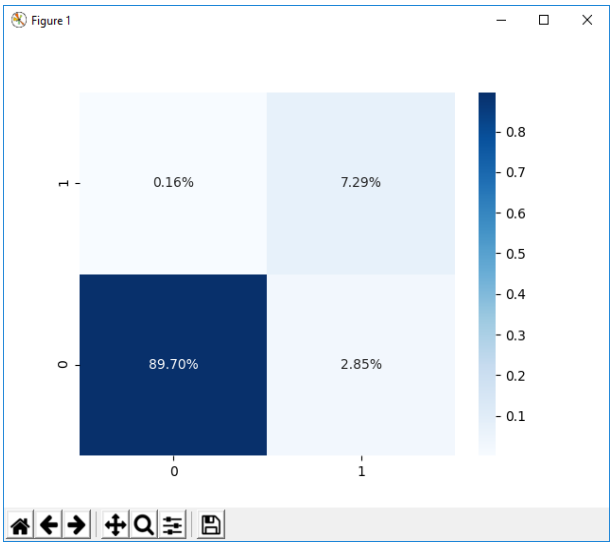
In above screen we can see dataset contains total 23 columns and using 2521 records to train SVM algorithm and using 631 test records to test SVM prediction accuracy and with normal SVM we got prediction accuracy as 90.96% and application showing confusion matrix of true and false prediction values where 531 and 3 are the true prediction and 54 and 43 are the false or incorrect prediction and below is the graph format of confusion matrix



In above graph 84.15% and 6.81% is the true prediction and now click on 'Extension Train SVM with PCA Features Optimization' button to train SVM with PCA features optimization and to get below prediction accuracy



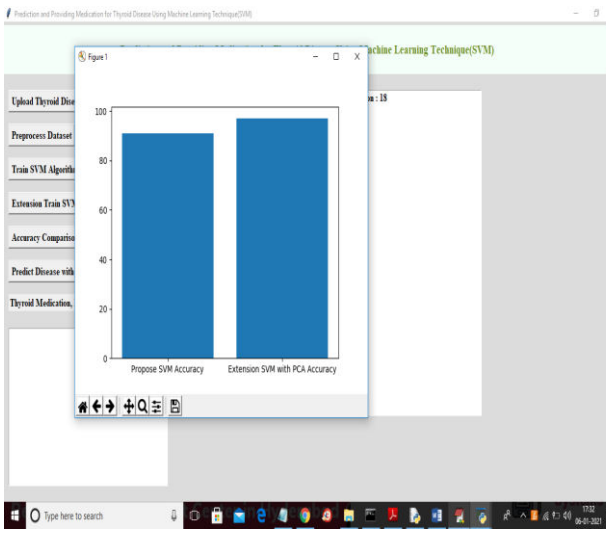
In above screen SVM with PCA extension got 96.98% prediction accuracy and confusion matrix values is also better compare to normal SVM and below is extension SVM confusion matrix graph



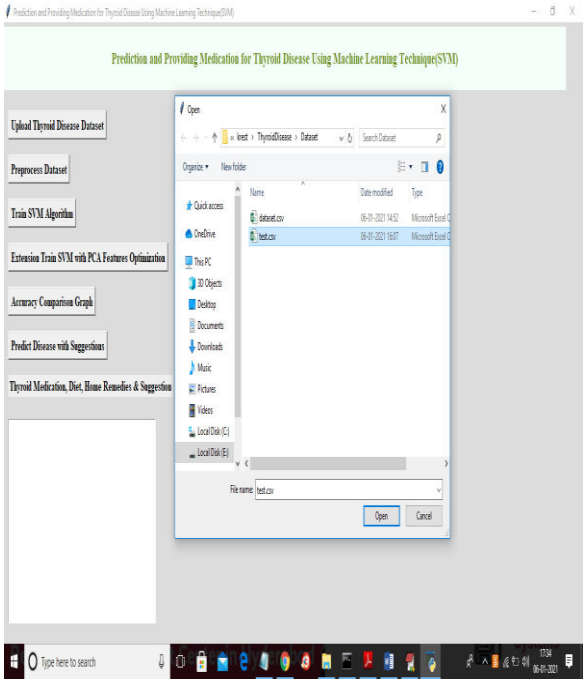
In above graph 89.70 and 7.29% is the correct prediction and other values are the

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

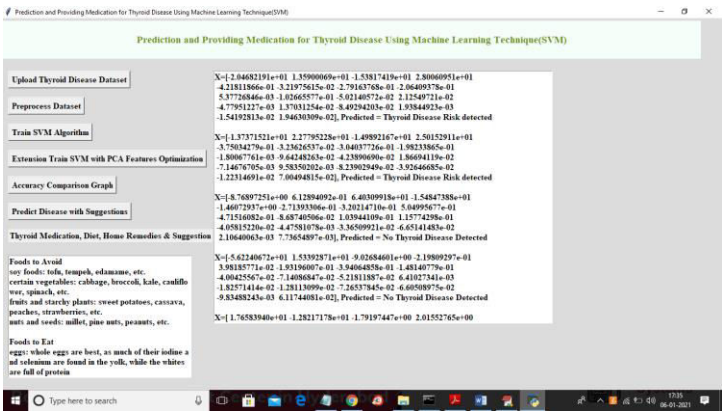
false prediction. Now click on 'Accuracy Comparison Graph' button to get below accuracy comparison graph



In above graph x-axis represents algorithm name and y-axis represents accuracy of those algorithms and from above graph we can conclude that extension SVM with PCA is better than normal SVM and now click on 'Predict Disease with Suggestions' button to upload new test data and predict whether new test data contains thyroid or not



In above screen selecting and uploading 'test.csv' file and then click on 'Open' button to upload test dataset and to predict disease and to get below screen



In above screen in brackets we can see each record test value and after bracket we can see value as thyroid risk detected or not and if detected then it left box we are showing diet and medication plan as suggestion

6. CONCLUSION

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 logistic regression algorithm to train our dataset and to predict thyroid disease with more accuracy. Here the machine is trained to detect whether the person normal, hyperhypothyroidism based on the user's input. So when user enters data in web app the data will be processed in backend (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..

7. FUTURE SCOPE

Further development can be do by using image processing of ultrasonic scanning of thyroid images to predict thyroid nodules and cancer, which cannot be recognized in blood test report. By combining both the results, thyroid disease prediction can cover all thyroid related diseases..

REFERENCES

[1]Ankita Tyagi and Ritika Mehra. (2018).“Interactive Thyroid Disease Prediction System using Machine Learning Techniques” published on ResearchGate.

[2] YongFeng Wang,(2020). “Comparison Study of Radiomics and Deep-Learning Based Methods for Thyroid Nodules Classification using Ultrasound Images” published on IEEEAccess.

[3] Sunila Godara,(2018). “Prediction of Thyroid Disease Using Machine Learning Techniques” published on IJEE.

[4] Hitesh Garg,(2013). “Segmentation of Thyroid Gland in Ultrasound image using Neural Network” published on IEEE.

[5] L. Ozyılmaz and T. Yıldırım,(2002). “Diagnosis of thyroid disease using artificial neural network methods,” in: Proceedings of ICONIP’02 9th international conference on neural information processing (Singapore: Orchid Country Club, pp. 2033–2036).

[6] K. Polat, S. Sahan and S. Gunes,(2007) “A novel hybrid method based on artificial immune recognition system (AIRS) with fuzzy weighted preprocessing for thyroid disease diagnosis,” Expert Systems with Applications,(vol. 32, pp. 1141-1147).

[7] F. Saiti, A. A. Naini, M. A. Shoorehdeli, and M. Teshnehlab,(2009) “Thyroid Disease Diagnosis Based on Genetic Algorithms Using PNN and SVM,” in 3rd International Conference on Bioinformatics and BiomedicalEngineering. ICBBE 2009.

[8] G. Zhang, L.V. Berardi,(2007) “An investigation of neural networks in thyroid function diagnosis,” Health Care Management Science,1998, (pp. 29-37.)

[9] Ehrenstein V, Nielsen H, Pedersen AB, Johnsen SP, Pedersen L. (2017) Clinical epidemiology in the era of big data: new opportunities, familiar challenges. Clin Epidemiol. ; 9:245-250

[10] S. Godara and R. Singh,(2016) "Evaluation of Predictive Machine Learning Techniques as Expert Systems in Medical Diagnosis", Indian Journal of Science and Technology, (Vol. 910).

[11] Sunila, Rishipal Singh and Sanjeev Kumar.(2016) "A Novel Weighted Class based Clustering for Medical Diagnostic Interface." Indian Journal of Science and Technology (Vol 9).