PREDICTION AND PROVIDING MEDICATION FOR THYROID DISEASE USING MACHINE LEARNING TECHNIQUE (SVM)

[1]M. Ramya,[2]Dr. P V Siva Kumar

[1]Department of CSE, VNR VJIET, Hyderabad, Telangana, India, ramya.alice9@gmail.com
[2]Associate Professor, Department Of CSE, VNR VJIET, Hyderabad, Telangana, India,
sivakumar_pv@vnrvjiet.in

Abstract: Thyroid disorder leading cause of medical diagnosis and prediction development, which medical science is a complicated axiom. Thyroid gland is one of our body's main organs. Thyroid hormone secretions are responsible for regulating metabolism. Hyperthyroidism and hypothyroidism are the two prominent thyroid disorders that produce thyroid hormones for control of body metabolism. The machine learning is critical in the disease prediction process and in the study and classification models used for thyroid disease on the basis of data obtained from hospital datasets. A decent knowledge base must be ensured, built and used as a hybrid model to solve dynamic learning tasks like medical diagnosis and predictional tasks. Basic techniques of machine learning are used for the identification and inhibition of thyroid. The SVM is used to predict the approximate probability of a thyroid patient.If the patient has risk of getting thyroid our system has to give suggestions like recommending home remedies, precautions, medication etc…

Keyword: Machine Learning Algorithm, Thyroid disease, Support Vector Machine (SVM), K-NN, Decision Trees Prediction model system,

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 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. With T3 and T4 harmones, iodine is primary building block of thyroid glands and is prostrate in only some unique problems, which are exceedingly prevalent. Insufficient elements of these hormones to hypothyroidism and an inappropriate portion to hyperthyroidism. Hyperthyroidism and underactive thyroidism have multiple origins. There are a number of drugs. Thyroid surgery is weak to ionising radiation, continuous thyroid softness, iodine deficiency, and loss of enzyme to produce thyroid hormones.

2. RELATED WORKS

Diagnosis of Thyroid disorder using Artificial Neural Networks: A big problem in medical research is that body is responsive to other hormones that provide right diagnosis of the condition in the first place. Thyroid, by its therapy, participate in the cellular processes. This paper deals with the diagnosis of thyroid hormone-producing thyroid disorders (T4) and triiodothyronine, particularly thyroxine disorders (T3). These hormones fuel neural networks with the use of three algorithms to regulate the metabolism of artificial neural networks (ANNs); (BPA),efficiency of thyroid is often measured in terms of factors such as the accuracy of the calcitonin hormone, role in diagnosing calcium and training time. Thyroid is monitored by analogy to help find best model for hypothalamus and pituitary diagnosis.

A modern hybrid approach for thyroid diagnosis based on the artificial immune recognition system (AIRS) with blurry, weighted pre-treatment: Proper understanding of the functional data of thyroid glandis important concern diagnosis ofthyroid disease. Major function of thyroid gland is to help control the metabolism ofbody. This is provided by the thyroid hormone released by thyroid gland. Containing very little thyroid hormone (hypothyroidism) or producing more thyroid hormone (hyperthyroidism) determines kind of thyroid disorder. Artificial Immune Systems are recent yet powerful branch of artificial intelligence. Artificial Immune Recognition System (AIRS) as suggested has so far been one of the systems proposed in this area. Watkins has demonstrated an important and fascinating success on the topics that have been discussed. The purpose of this research is to diagnose thyroid using modern hybrid machine learning process, withclassification scheme. Via the hybridisation of AIRS with formed Fuzzy, the methods used to solve diagnostic problem by grouping are weighted.

The strength of the samples is checked using a cross-validation process. We used the data collection for thyroid disease taken from the UCI respiratory learning machine. We have also reached 85% classification accuracy, which is highest achieved to date. Accuracy of the classification was achieved by a 10-fold cross-validation.

An investigation of neural networks in thyroid function diagnosis: We are investigating the ability to classify thyroid disorders in artificial neural networks. Neural network robustness in relation to sampling differences is evaluated using a cross-validation process. We explain the connection between neural networks and conventional Bayesian classifications. Neural networks can have good estimations of posterior probabilities and thus provide improved classification efficiency than conventional statistical approaches such as logistic regression. Furthermore, neural network models have seen to be resilient to sampling variations. It is demonstrated for medical diagnosis issues, where data sometimes very unbalanced, neural networks may be a promising classification tool for practical use.

Proper diagnosis of thyroid dysfunction based on clinical and experimental testing is often complicated. One explanation is the non-specific existence of certain signs of thyroid disease. This is particularly true in hypothyroidism, where signs such as lethargy, confusion, weight gain, impaired memory are easily associated with psychological and medical disorders. The problem is sometimes compounded in elderly patients whose signs are often masked or due to other medical conditions. While lab-oratory tests have become more reliable and are effective in diagnosing thyroid disorders (positive predictive rates of certain tests have recently been reported to be over 90%. Findings are also not very satisfactory across all cases. Diagnostic problems emerge from variability in test results across patients and other considerations such breastfeeding, interactive medications.

A survey on applying machine learning techniques for management of diseases: In recent years, increasing research knowledge and large data output have contributed to exponential rises in databases and repositories. Biomedicine is one of rich data domains. Currently, a wide range of biomedical information available, ranging from explanations of clinical disorders to various types of biochemical data and image instrument outputs. Manually extracting and converting biomedical trends from data into machine-understood information is daunting task as the biomedical domain consists of vast, dynamic and nuanced expertise. Data mining can improve biomedical pattern extraction performance.

A description of the uses of data mining for disease control is provided in this report. The key emphasis analyse machine learning techniques (MLTs) that commonly used forecast, predict, and treat major common diseases such as cancer, hepatitis, and heart disease. The methods, including the Artificial Neural Network, K-Nearest Neighbour, Decision Tree and Associative Grouping, are explained and analysed. This survey offers general overview of current state of MLT disease control. The precision obtained for the different applications varied from 70% to 100% based on the disorder, the problem solved and the data and procedure used.

3. SYSTEM ANALYSIS

The Application Development Life Cycle (SDLC) or the Systems Architecture, Information Systems, and Network Engineering Software Development Life Cycle (SDLC) is process of developing or modifying systems and models and techniques people use to build these systems. SDLC concept supports all product development methodologies in software engineering. These methodologies form basis for planning and maintaining software development information system.

Existing System: Thyroid disorder significant cause of formation of medical diagnosis and estimation, which is a challenging axiom of medical science. The secretions of thyroid hormones are guilty of metabolism regulation. Hyperthyroidism and hypothyroidism are one of the two prevalent thyroid disorders that release thyroid hormones to control body metabolism. Data cleaning methods have been used to make data primitive enough to do analytics to demonstrate likelihood of patients having thyroid.**Disadvantages:** Energy level, Weakness, Breathing.

Proposed System: In the prediction process, machine learning plays a key role, and paper research and the classifications of models used in thyroid disease are based on information from UCI machine learning repositories. A decent knowledge base that can be centred 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, Decision Trees, K-NN were used to calculate an estimated probability of a patient having thyroid disease.

Advantages:

1. Avoids long-term risks of anti-thyroid and radioactive iodine.
2. Provides histology tissue, for childbearing instantly.

3.1. PROCESS MODEL USED WITH JUSTIFICATION

SDLC (Umbrella Model):

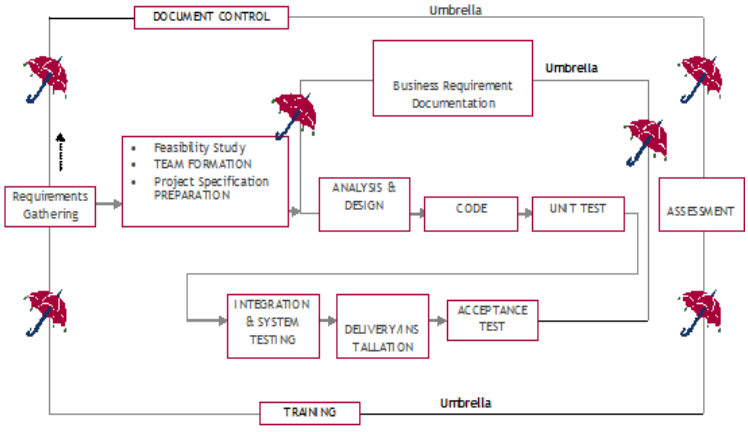


Fig 1: Block Diagram of SDLC

SDLC is Software Development Life Cycle. It's a tradition for software industry to build strong software.

Stages in SDLC:

• Necessity set
• Research

• Design

• Codification

• Testing

• Maintenance

3.2 External Interface Requirements

User Interface: It is a user friendly Java Graphical User Interface.

Hardware Interfaces: The user-console interaction is accomplished through Java capabilities.

4. IMPLEMENTATION

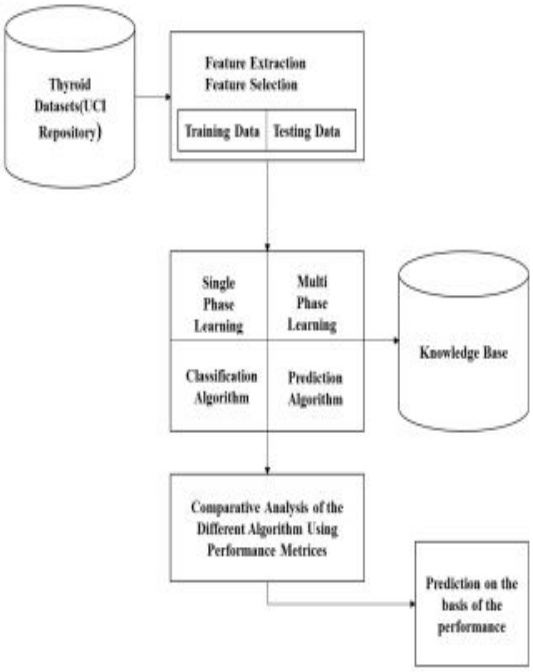
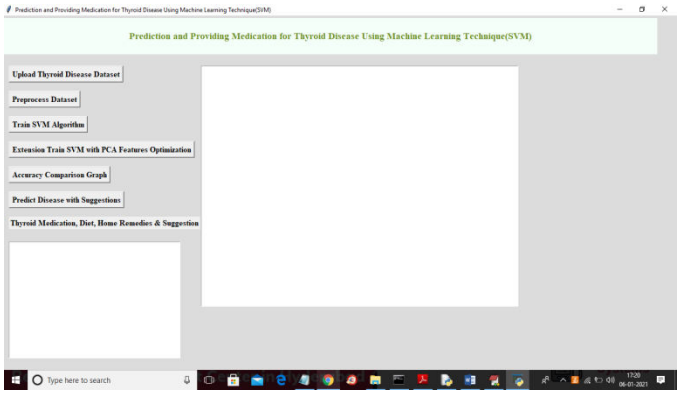


Fig 2: Architecture of Thyroid Prediction System

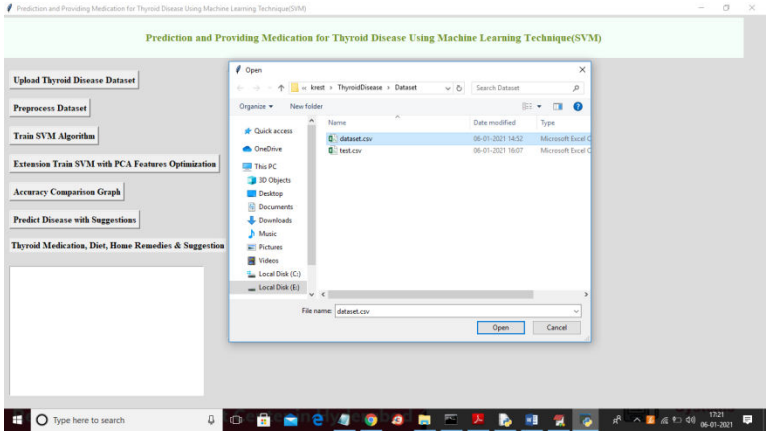
Support Vector Machine Algorithm: The Support Vector Machine SVM is one of the most commonly used classification and regression supervised learning algorithms. However, it is primarily used for problems of classification of machine learning. The purpose of SVM algorithm is to establish the best possible line or boundary to segment n-dimensional space into categories so that new data points can be located correctly in the future. The ideal boundary of selection is called hyperplane. To create a hyperplane, SVM chooses extreme points/vectors. These extreme cases are called vectors support, so the algorithm is called a vector support. Consider the diagram below, where a decision boundary or hyperplane classifies two distinct classes.

PCA Algorithm: PCA is a mathematical technique using an orthogonal transformation that transforms the number of associated variables into an array of uncorrelated variables.PCA is the most commonly used method for exploratory data processing and deep learning for predictive models. In addition, PCA is an unsupervised mathematical method used to analyse the interrelations between a number of variables. It is also known as a general factor analysis where regression defines the best fit rows.

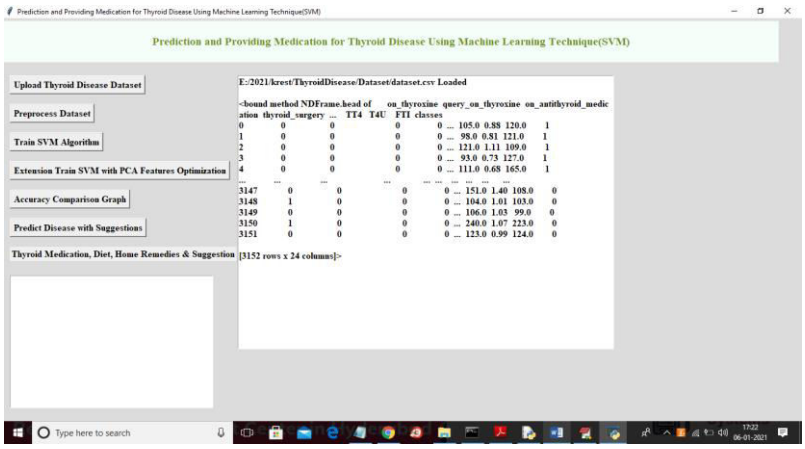
5. RESULTS



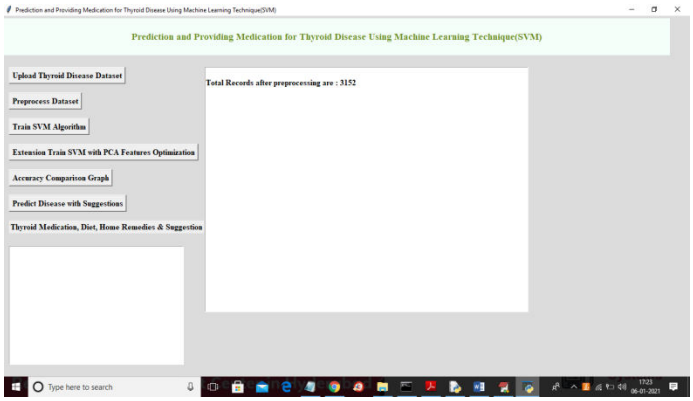
In above screen click on 'Upload Thyroid Disease Dataset' button to upload dataset and to get below screen



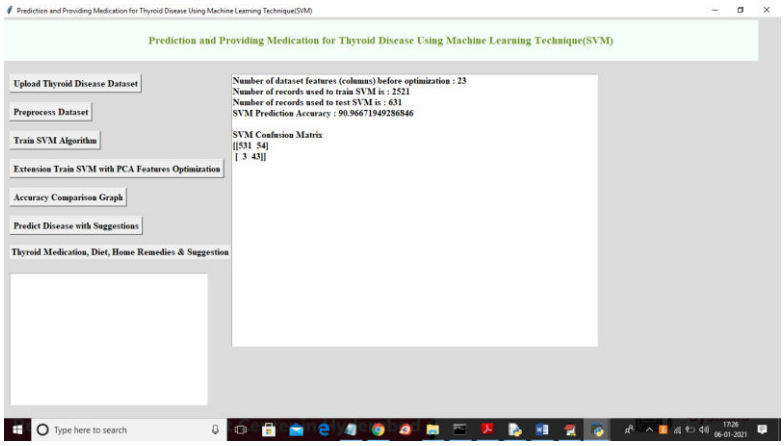
In above screen selecting and uploading 'dataset.csv' file and then click on 'Open' button to load dataset and to get below screen



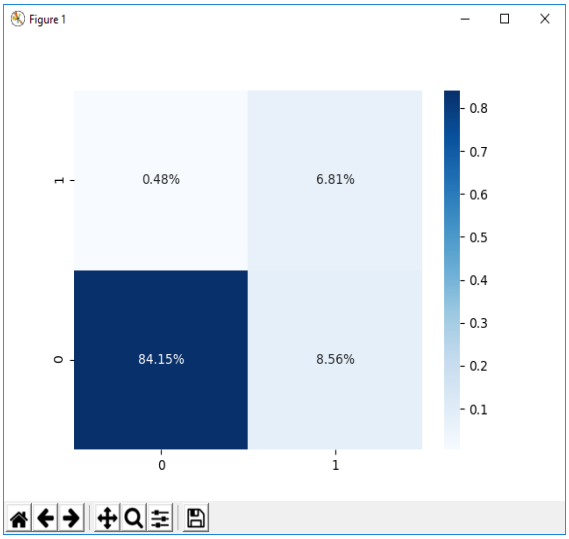
In above screen dataset loaded and displaying few records from dataset and then click on 'Pre-process Dataset' button to remove missing and NAN values from dataset and to separate X and Y values where X contains all dataset values and Y contains class label value.



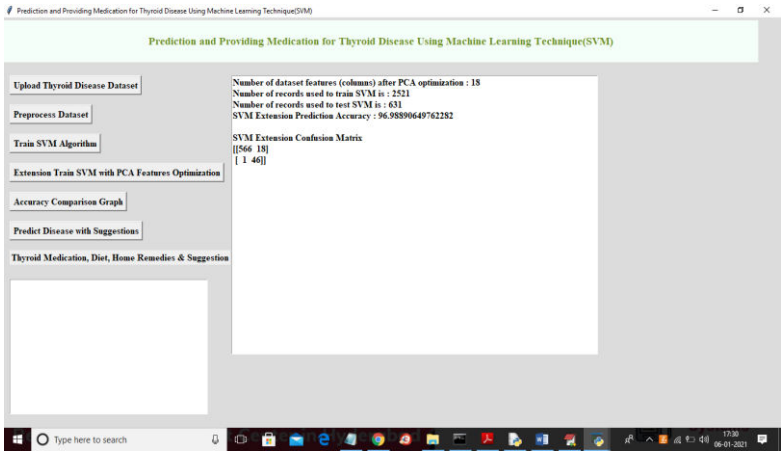
In above screen dataset showing 3152 pre-process records and now dataset is ready and now click on 'Train SVM Algorithm' button to split dataset into train and test and then apply SVM algorithm on train data to generate model and then model will be applied on test data to calculate prediction accuracy



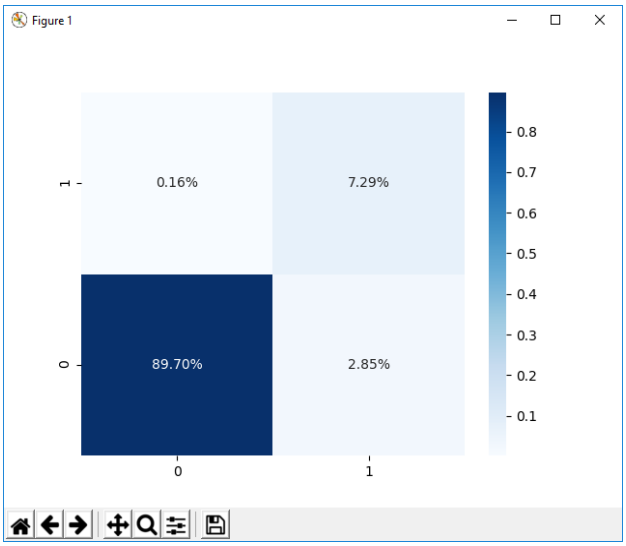
In above screen we can see dataset contains total 23 columns and using 2521 records to train SVM algorithm and using 631 test records to test SVM prediction accuracy and with normal SVM we got prediction accuracy as 90.96% and application showing confusion matrix of true and false prediction values where 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



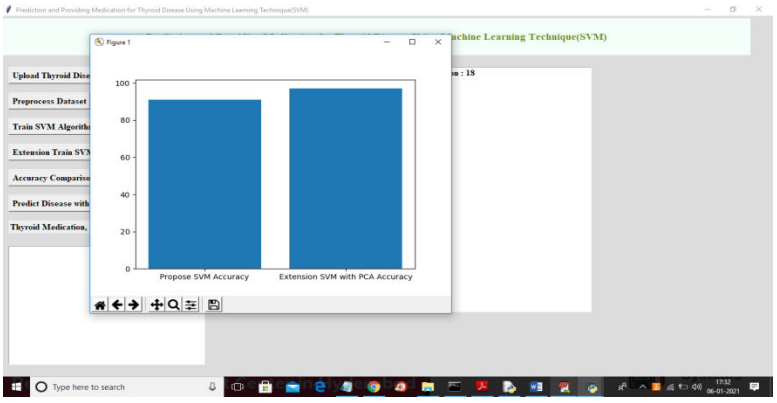
In above graph 84.15% and 6.81% is the true prediction and now click on 'Extension Train SVM with PCA Features Optimization' button to train SVM with PCA features optimization and to get below prediction accuracy



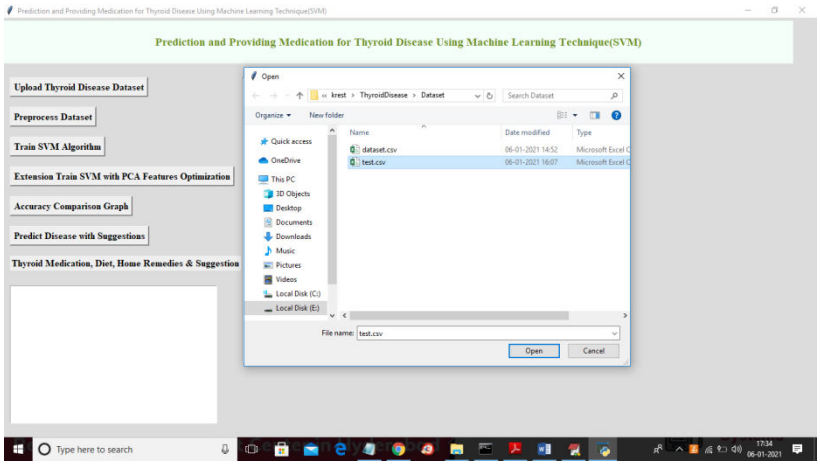
In above screen SVM with PCA extension got 96.98% prediction accuracy and confusion matrix values is also better compare to normal SVM and below is extension SVM confusion matrix graph



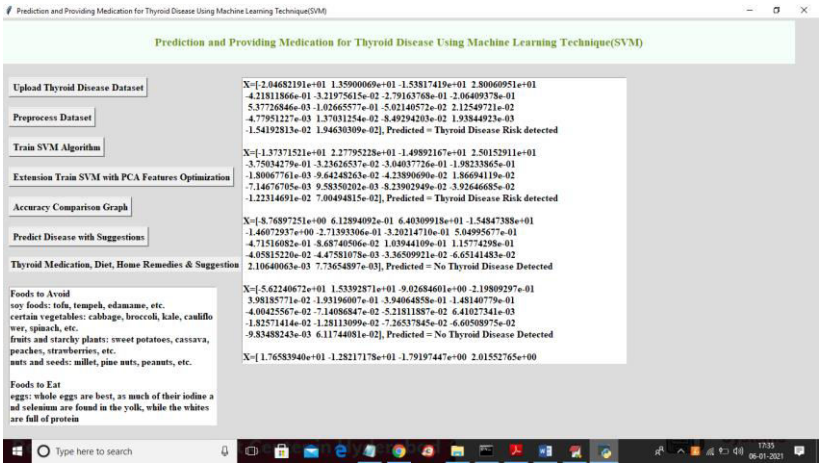
In above graph 89.70 and 7.29% is the correct prediction and other values are the false prediction. Now click on 'Accuracy Comparison Graph' button to get below accuracy comparison graph



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

CONCLUSION

The research work further studies the unusual machine learning strategies which can be mobilised in the diagnosis of thyroid diseases. In recent years, numerous approachable analyses for adequate and professional diagnosis of thyroid disease have been developed and used. Study reveals that various technologies used in both articles demonstrating different precision. Most academic papers indicate that the neural network outperforms other strategies. In other hand, also due to the fact that the help vector machine and decision tree have done well. There is no question that experts around the world have gained a great deal of improvement in diagnosing thyroid diseases, but it is recommended that number of criteria used by patients to diagnose thyroid disorders can be limited. More characteristics mean that a patient needs to perform a larger range of health evaluations that are both cost-effective and time-consuming. Thus, certain algorithms and predictive models of thyroid disease need to be developed, requiring a minimum number of criteria for person to diagnose thyroid disease and saving patient's time and money.

REFERENCES

1. L. Ozyılmaz and T. Yıldırım, "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, 2002) pp. 2033–2036.

2. K. Polat, S. Sahan and S. Gunes, "A novel hybrid method based on artificial immune recognition system (AIRS) with fuzzy weighted pre-processing for thyroid disease diagnosis," Expert Systems with Applications, vol. 32, 2007, pp. 1141-1147.
3. F. Saiti, A. A. Naini, M. A. Shoorehdeli, and M. Teshnehlab, "Thyroid Disease Diagnosis Based on Genetic Algorithms Using PNN and SVM," in 3rd International Conference on Bioinformatics and Biomedical Engineering, 2009. ICBBE 2009.
4. G. Zhang, L.V. Berardi, "An investigation of neural networks in thyroid function diagnosis," Health Care Management Science, 1998, pp. 29-37. Available: http://www.endocrineweb.com/ thyroid.html, (Accessed: 7 August 2007).
5. V. Vapnik, Estimation of Dependences Based on Empirical Data, Springer, New York, 2012.
6. Obermeyer Z, Emanuel EJ. Predicting the future— big data, machine learning, and clinical medicine. N Engl J Med. 2016; 375:12161219.
7. Breiman L. Statistical Modeling: the two cultures. Stat Sci. 2001; 16:199-231.
8. Ehrenstein V, Nielsen H, Pedersen AB, Johnsen SP, Pedersen L. Clinical epidemiology in the era of big data: new opportunities, familiar challenges. Clin Epidemiol. 2017; 9:245-250
9. Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature. 2015; 521: 452-459.
10. Azimi P, Mohammadi HR, Benzel EC, Shahzadi S, Azhari S, Montazeri A. Artificial neural networks in neurosurgery. J Neurol Neurosurg Psychiatry. 2015; 86:251-256.
11. Deo RC. Machine learning in medicine. Circulation .2015; 132: 19201930.
12. P.C. Austin, J.V. Tu, J.E. Ho, D. Levy, D.S. Lee, Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes, J. Clin. Epidemiol. 66 (4) (2013)398–407.
13. A.K. Pandey, P. Pandey, K.L. Jaiswal, A heart disease prediction model using Decision Tree, IUP J Comput. Sci. 7 (3) (2013) 43.
14. S. Ismaeel, A. Miri, D. Chourishi, in: Using the Extreme Learning Machine (ELM) technique for heart disease diagnosis, IEEE Canada International Humanitarian Technology Conference, 2015, pp. 1–3.
15. L. Verma, S. Srivastava, P.C. Negi, A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data, J. Med. Syst. 40 (7) (2016) 1–7.