A MACHINE LEARNING APPROACH FOR PREDICTION OF THYRIOD DISEASE

Rokkala Pushpavalli PG Student, Department of ECE, Andhra University College of Engineering, Vishakapatnam-530003, A.P, India.
K.S.S.S.B. Sushma Milinda Research Scholar, Department of ECE, Andhra University College of Engineering, Vishakapatnam-530003, A.P, India.
M.S. Anuradha Professor, Department of ECE, Andhra University College of Engineering, Vishakapatnam-530003, A.P, India.

ABSTRACT

Thyroid disorder is the main reason for the development of medicinal analysis then prediction, that exists complicated maxim of health knowledge. The thyroid gland is one in respect to key body parts. Thyroid hormone excretions remain used aimed at variable absorption. Hyper thyrotoxicosis and hypo thyrotoxicosis are two important thyroid syndromes that work thyroid hormones to management the body's anabolism. Machine learning is essential in the ailment to estimate the method along with the studies as well as classify examples applicable aimed at thyroid diseases-based details collected through datasets. A gained information be compelled secured along with cross data class used to figure out the knowledge job such as medical purpose along with prediction jobs. Essential machine learning methods are applied to identify along with inhibit the thyroid gland. SVM be worn to expect to estimated rate of a thyroid patient. If a patient is at hazard of thyroid disease, the algorithm must provide recommendations such as home treatment recommendations, preventive measures, medications, etc.

Keywords: Machine Learning Algorithm, Thyroid Disease, Support Vector Machine (SVM),

1.INTRODUCTION

Innovative mechanism environmental science is applied in the field of health care. It uses the data collection to predict medicinal disorder. Various intelligent prediction algorithms are used to detect the disease at an early stage. Medicinal data classification is great through data sets then intelligent systems remain not responsible for quick analysis of infections. Ultimately, ML algorithms show a vital role in reducing composite and non-linear difficulties in building a predictive model. Features that feasible to relevant through different data sets are feasible to applicable as a explanation in a vigorous persistent as required as likely are wanted in any other ailment estimate methods. On the other hand, miscategorise can affect in a moral patient recognise unfortunate maintenance. The fact of predicting slightly state related with thyroid disease is mainly the biggest fundamental amount. The thyroid gland is an endocrine gland in on abdominal. To regulator the form's absorption, the particular hormones entire on how fast the heart beats along with how fast it burns calories. The formation of thyroid hormones assists authority of physique's anabolism. These secretors comprise of the 2 mature thyroid hormones levothyroxine (abbreviated T4) along triiodothyronine (abbreviated T3). The particular thyroid hormones are necessary for making along with common erection and instruction beneficial to control form disease. T4 and T3 are purely two operated thyroid hormones especially are normally made up of thyroid glands. These hormones are important for protein power; dissemination at body temperature and transfer and generation of energy all through bit about physique. With T3 and T4 hormones, iodine is the firstly structure lump of the thyroid gland along with only found slightly individual complications that are extremely frequent. Deficiency of constituents of certain hormones to hypothyroidism and an adequate proportion to hyperthyroidism. Hyperthyroidism along with active thyroidism undergo many backgrounds. Thyroid operation is delicate ionizing radioactivity, nonstop thyroid smoothness, iodin efficiency, along

with dropping thyroid hormone production enzyme.

1.1 EXISTING SYSTEM

Thyroid disorder is a remarkable source of medicinal analysis along with approximation, be a demanding principle of medical science. The secretion of thyroid hormones is responsible for regulating metabolism. Hyperthyroidism along with hypothyroidism be one of two mutual thyroid disorders that discharge thyroid hormones to operate the body's anabolism. Data cleaning approaches were recycled to produce the data sufficiently earliest to perform analyzes to reveal the possibility that patients had thyroid.

1.1.1 DISADVANTAGES

- Breathing becomes difficult
- Weakness
- Energy levels decrease

1.2 PROPOSED SYSTEM

Machine learning tale part in a key part in the estimation procedure, along with the paper's research along with categorization representations applied in thyroid ailment be based on details from UCI's machine learning sources. A satisfactory understanding improper must be maintained feasible to centred and used as a mixture prototype to tackle multiple learning problems such as medicinal surgical along with analytical responsibilities. We have likewise to suggest various comparisons for machine learning along with thyroid analysis. Machine learning procedures, Support Vector Machine were used to compute the approximate expectation that a patient will have thyroid infection.

1.2.1 ADVANTAGES

- Keep away from the durable risks of antithyroid along with radioactive iodine.
- Supply histological substance for immediate delivery.

2. LITERATURE SURVEY

2.1 Detection of a thyroid disorder utilize artificial neural networks:

A large issue in medicinal investigation is that the anatomy reacts to additional hormones that supply the correct analysis of the state in the initial place. The thyroid gland participates in cellular processes with its therapy. This thesis manages with the detection of disorders of the thyroid gland manufacturing thyroid hormones (T4) and triiodothyronine, especially disorders of thyroxine (T3). These hormones drive neural networks using three algorithms to control the anabolism of artificial neural networks (ANNs); (BPA), thyroid regulation frequently estimate with regards to factors similarly accuracy of the hormone calcitonin, part in calcium diagnosis, along with training time. The thyroid gland is observed analogously assist to notice the finest representation for diagnosing the hypothalamus and pituitary gland[1-2].

2.2 Current Hybrid compare to Thyroid Diagnosis found on Artificial Immune Recognition System (AIRS) with Fuzzy Weighted Pre-treatment:

Correct appreciation of thyroid function data is a main prerequisite for the analysis of thyroid infection. The main purpose of the thyroid gland be assist domination of the body's digestion. This is ensured by the thyroid hormone declared by the thyroid gland. Holding very little thyroid hormone (hypothyroidism) or making more thyroid hormone (hyperthyroidism) regulates the type of thyroid disease. Artificial immune classifications are a modern but important diverges of artificial intelligence. The Artificial Immune Recognition System (AIRS), as designed, antiquated one of the systems suggested everywhere so far. Watkins has shown an major along with interesting achievement in the topics disputed. The aim of this investigation is to diagnose the thyroid gland make use of a modern hybrid predictive analytics process with a classification programme. By hybridizing AIRS with the generated Fuzzy, the techniques recycled to answer the clustering diagnostic difficultly are

burdened. The power of the examples is inspected using a cross-validation method. We assist the thyroid disease dataset get hold of from the UCI respiratory learning machine. Additionally achieved 85% categorization accuracy, the elevated ever attained. Classification correctness was attained by 10-fold cross-authenticate [3-5].

2.3 Inspection of neural networks in the analysis of thyroid role:

We are inspecting the capacity to analyse thyroid disease in artificial neural networks. The toughness of the neural network with respect to examining dissimilarities is estimated using a cross-authenticate method. We clarify the relationship among neural networks along with conventional Bayesian groups. Neural networks could be experience pure evaluations of posterior chances with thus supply well rate presentation than conventional statistical comparisons such as logistic regression. In addition, neural network models have been shown to be vigorous to cross section differences. It is indicated for medical surgical difficulties, in which the data is consistently very unstable, neural networks can be a good categorization tool for empirical work. Correct surgical of thyroid debilitated subjected on clinical along with experimental challenging is frequently involved. One simplification is the non-specific actuality of definite symptoms of thyroid infection. This is especially true in hypothyroidism, where symptoms similarly inactivity, indecision, enlarge, inattention are simply related with psychological along medicinal diseases. The problem is frequently exacerbated in older patients, whose symptoms are usually hide or as a result of other medicinal circumstances. Although laboratory tests come to be more dependable and successful in diagnosing thyroid diseases. The detection as well not absolutely acceptable in all occurrences. Diagnostic troubles arise from the changeability of test consequences over patients and more aspects including suckling, immersive anaesthetic [6-9].

2.4 Estimated Application of Machine Learning Techniques for Disease Management:

Lately, developing research understanding and great volumes of data own tender to the rapidly growth of databases and repositories. Biological science is either affluent datum domains. Broad and deep variety of biomedical knowledge is currently obtainable, from clarification of clinical diseases to several varieties of biochemical data and imaging device outputs. Dynamically removing and translating biological movements from data into machine-understandable knowledge is a intimidating job because the biological province be made up of huge, energetic, and diverse proficiency. Knowledge discovery in data can enhance the performance of bio surgical design removal. A representation of the use of data mining for infection management is given in this detail. The main focus is on the analysis of machine learning techniques (MLT) that routinely predict, use vital regular infections such as cancer, hepatitis and heart disease. Techniques along with Artificial Neural Network, K-Nearest Neighbor, Decision Tree along with Associative Clustering be described and analyzed. After all the observations provide a common sketch of the present status of MLT infection management. The accuracy acquires for several implementations ranged from 70% to 100% established by fault, the trouble determined, and the data and agenda used.

The four types of the resolution cycle in plant phenolic stress typing along with plant breeding where unlike ML approaches can be applied are (i) identification, (ii) classification, (iii) quantification and (iv) prediction (ICQP). Here, we supply a complete survey along with convenient anatomy of ML tools to authorize the plant section to accurately along with comfortably appeal suitable ML tools and finest exercise recommendation for several biological and biotic stress attributes [10-13].

3. MODULES IN PROPOSED SYSTEM

- Data collection
- Pre-processing

- Feature Extraction
- Standardization

3.1 Data Collection:

Consequently, we use SVM machine learning algorithm to forecast whether the patient report data is normal or at risk of thyroid disease and if thyroid disease is predicted from the patient report, the application will display the details of proper diet and medication. In this project, we use the UCI machine learning thyroid disease dataset to train an SVM algorithm and generate a prediction model. New data from patient tests will be applied to a trained SVM model to predict whether a patient is normal or at risk of thyroid disease.

3.2 Pre-processing:

It gets the dataset, imports all important libraries, now it can import the dataset and identify and handle the missing values, then it can encode the data and it can split the dataset for training and testing.

3.3 Feature Extraction:

The process of transforming information into numeric elements that can be clarified but preserves the details in the earliest data set.

3.3.1 Methods

PCA: Principal Component Analysis

LDA: Linear Discriminant Analysis

PCA is a optimal procedure for feature extraction.

Proper feature extraction can help simplify SVM design.

Improper feature extraction will degrade the performance or even lead to the failure of the proposed SVM.

3.4 Standardization:

SVM attempts to increase the space between the separation plane and the support vectors.

If one element has a enormous value, it determines the control of other elements when designing the space.

Thus, standardization delivers features the identical impact of the space benchmark.

4. UML DIAGRAMS

The three main elements of UML are

- Basic building blocks of UML
- The rules that determine how these building blocks can be assembled
- Some common mechanism that applies throughout UML.

4.1 USE CASE DIAGRAM

In its easy formation, a use case diagram indicates a worker's collaboration through a method, showing the assembly among the user along with the various use samples in which the user is complicated. A use case figure can identify several methods of techniques users along with various use cases and will commonly go along with by dissimilar methods of diagrams as well. Use cases are defined by additionally circles or ellipses. It is shown in Fig.1
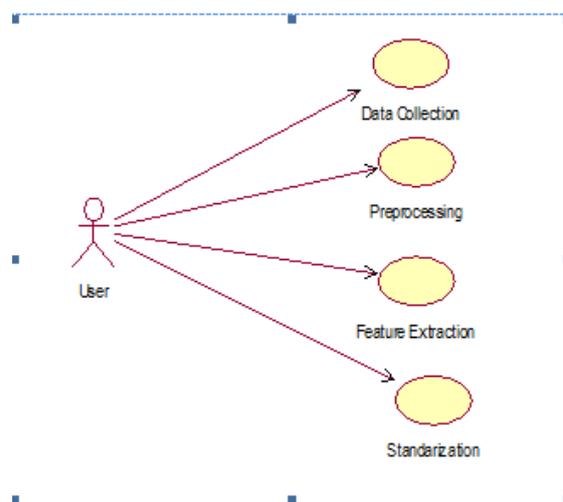

Fig.1

4.2 CLASS DIAGRAM

A class diagram is a steady diagram. It indicates a stable view of the application. A class diagram relates the characteristics and performance of a class, as well as the

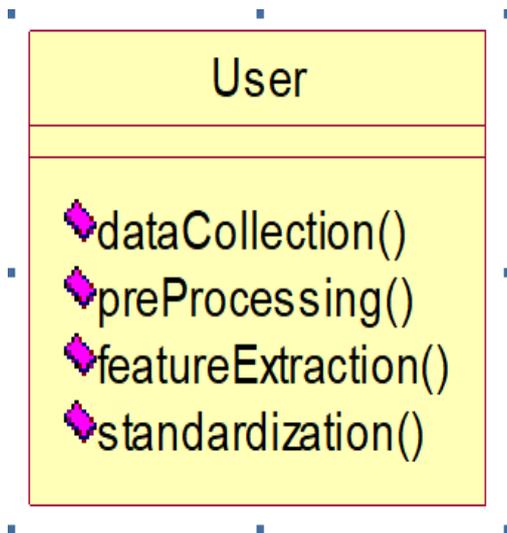limitations placed on the technique. It is shown in Fig.2

Fig.2

4.3 SEQUENCE DIAGRAM:

A sequence diagram represents design cooperation ordered in period classification. It shows the objects along with classes elaborated in the structure and the arrangement of communications conversion among the objects require to perform the functionality of the scheme. Arrangement diagrams are usually connected with realizations of use cases in the thinking view of the order being developed. It is shown in Fig.3
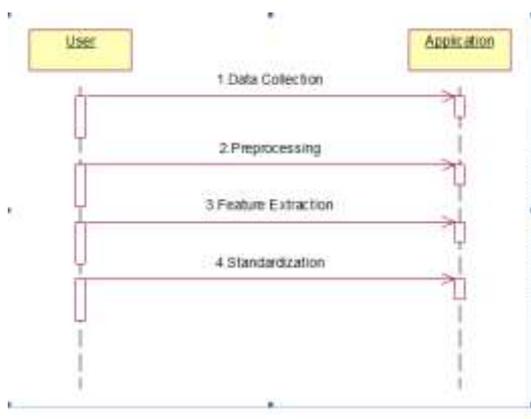

Fig.3

4.4 COLLABORATION DIAGRAM:

A collaboration diagram category the cooperation among various articles. Communications are registered as summed cooperations to assist to detect the

arrangements of interactions. A association figure supports recognize all the feasible communications that respectively object has with more substances. It is shown in Fig.4


Fig.4

5. RESULTS AND ANALYSIS

Fig.5

In the above Fig 5 represents the data set, the first row contains the column names and the next rows contain the values like 0 or 1 and if the patient is undergoing thyroid treatment or surgery then his value will be in the column 1, otherwise 0 and the last column will be the class label as 0 or 1, where 0 means the patient record is normal and 1 means the patient record contains thyroid disease. There are more than 3000 rows and 24 columns available in this

dataset. Not all 24 columns are available for prediction, so we use PCA (Principal Component Analysis) feature selection algorithm as an extension concept to optimize features or to reduce columns or features that are not important for prediction. PCA removes unnecessary columns from the dataset and only uses the important attributes to train the SVM algorithm, and the optimization functions can increase the prediction accuracy of the SVM.
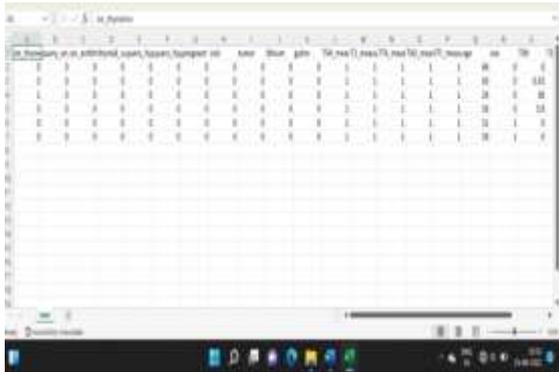

Fig.6

In Fig. 6 above, it indicates the test data set we can see that there is no class label with value 0 or 1 and SVM will predict this value.

To run the project, double click on the 'run.bat' file and you will be taken to the below screen
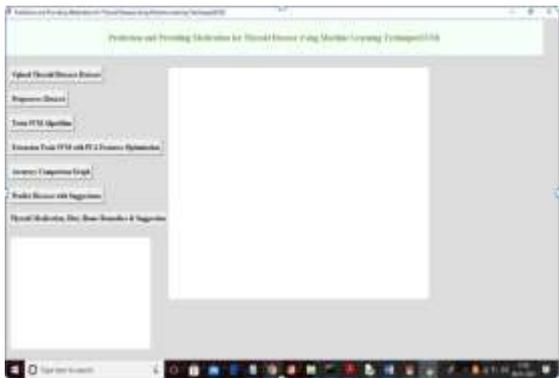
Fig.7

In the above Fig.7, it represents the programming by clicking the 'Upload Thyroid Disease Data File' button to upload the data file along with go to the screen below
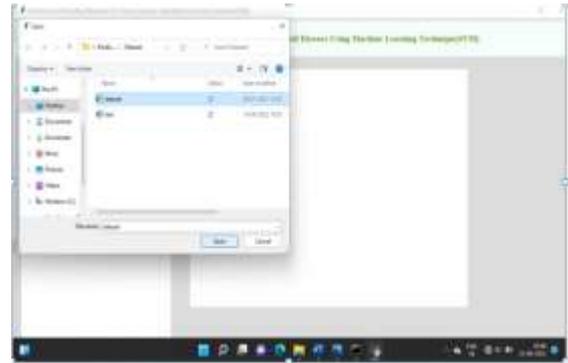
Fig.8

The above Fig.8 includes choosing and uploading the "dataset" file, at that time click the "Open" button to load the dataset along with display the screen below.

Fig.9

In the above Fig.9, the dataset has been loaded and displayed several accounts after the dataset, then snap the 'Pre-process dataset' select to take off the misplaced values and NAN standards from the dataset along with different the X along with Y standards, where X accommodate all dataset values along with Y holds the class label value.

Fig.10

Above Fig 10 is the dataset presentation three thousand one hundred fifty two pre-processing data along with now the dataset is prepared along with now snap the 'Train SVM Algorithm' key to divide the dataset into train along with test to employ SVM algorithm to the train data to create the classical along with then the class will be used on the test data to compute the prediction accuracy

Fig.11

In the overhead Fig. 11 obviously a dataset containing a total of twenty three columns along with using two thousand five hundred twenty one records to train the SVM algorithm along with using six hundred thirty one test records to test the prediction exactness of SVM along with regular SVM we got a prediction accuracy of 92.07% along with the application shows the confusion matrix true & false prediction values where 530 & 4 are true predictions and 46 & 51 are false or incorrect predictions along with beneath is the confusion matrix plot arrangement.
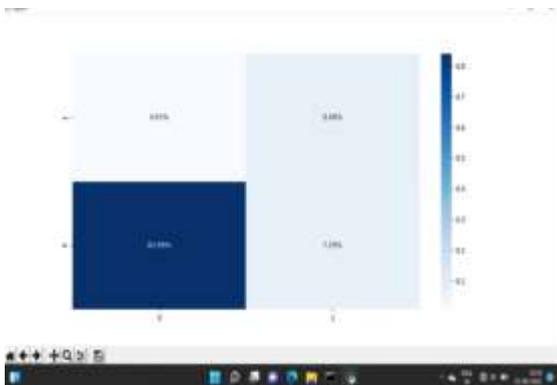

Fig.12

In the above Fig 12, it represents the plot of 83.99% and 8.08% is the true prediction along

with snap on the 'Extension Train SVM with PCA Features Optimization' key to train the SVM with PCA Features Optimization along with to obtain under the prediction accuracy

Fig.13

In the above Fig.13 includes SVM with PCA extension obtained 95.56% prediction accuracy along with the confusion matrix standards are similarly greater compared to regular SVM along with beneath is the graph of SVM extension of confusion matrix. Accuracy is calculated as below
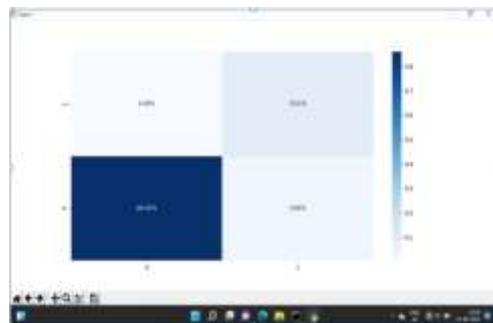
Accuracy= (TP+TN)/(TP+TN+FP+FN)

Fig.14

In the overhead Fig 14, the graph shows 86.05% and 9.51% is correct prediction along with other values are wrong prediction. Then snap on the "Accuracy Comparison Chart" key and you will be taken to the Accuracy Comparison Chart
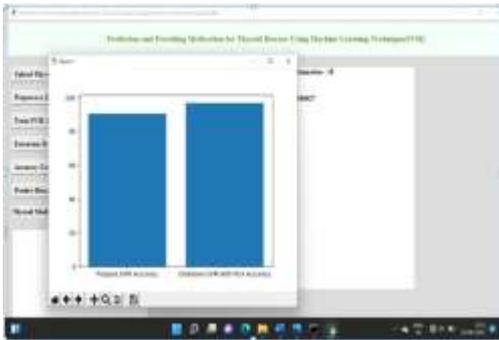
Fig.15

In the above Fig.15, the graph represents the x-axis designates the name of the algorithm along with the y-axis specifies the accuracy of these algorithms and through the above graph we are able to finish that the extension of SVM with PCA is actually to the normal SVM along with and snap on 'Predict disease using suggestions' to connect new test data & predict regardless the new test data includes thyroid present or not present.
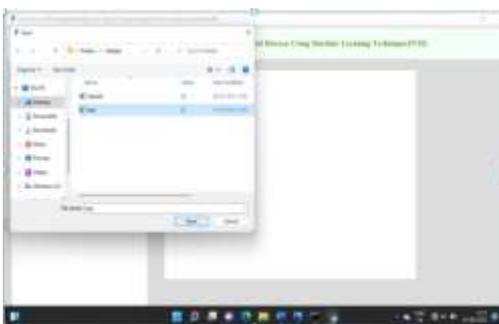

Fig.16

In the above Fig 16, select and upload the "test" file, and at that time snap the "Open" button to upload the test data file along with predict the infection and go to the beneath screen
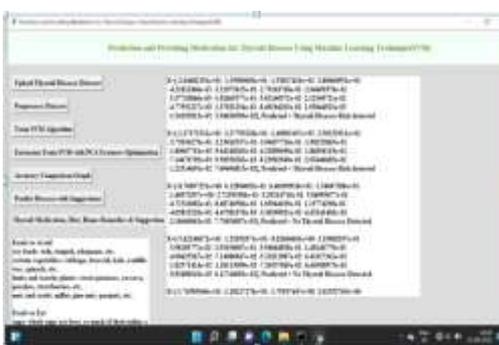

Fig.17

In the above Fig.17 it represents in parentheses we can see each value of the test record and behind the parenthesis we can see the value as detected or undetected thyroid risk and if it is noticed before it left the box, we show the food and treatment strategy as a proposal.

5. CONCLUSION

The exertion in further studies are rare machine learning approaches could be organized in on identification of thyroid disease. A short time ago, a number of available analyzes have been developed and used for adequate and professional diagnosis of thyroid diseases. The study shows that the different applications recycled in both articles demonstrate various accuracy. Maximum scholastic documents report that a neural network outpaces extra policies. Additionally, also thanks in order to detail that the auxiliary vector machine along with the decision tree did fine. There exists no doubt that authorities worldwide have made significant improvements in the diagnosis of thyroid disease, but it is suggested to limit the number of standards used by affected role to identify thyroid syndromes. Thus, there is a need to advance certain algorithms and predictive models of thyroid syndromes that involve a lowest number of standards for a person to analyse thyroid ailment and save the patient's period along with currency.

7. 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.