ISSN: 2366-1313

PREDICTION OF HYPOENDOCRINE GLAND USING MACHINE LEARNING ALGORITHMS

[1]E.Shirisha, [2]RagiddiPoojitha ,[3]KanukuntlaSreeja,[4]Jella Anil Kumar,

[1]Assistant Professor, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad, sirishaedikoju@gmail.com

[2]BTech student, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad, poojitharagiddi10@gmail.com

[3]BTech student, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad,kanukuntlasreeja12@gmail.com

[4]BTech student, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad,anilkumarjella25@gmail.com

Abstract: *The thyroid gland is a vascular gland and one of the most vital organs in the human body. This gland secretes two hormones that help control the body's metabolism. The types of thyroid disorders are hyperthyroidism and hypothyroidism. When this disease occurs in the body, they secrete certain hormones that lead to an imbalance in the body's metabolism. A thyroid-related blood test is used to screen for this disorder, but it's often blurry, and the noise can be a gift. Data cleaning techniques were used to make the information primitive enough to be analysed to detect the likelihood of patients having the disorder. MachineLearning plays a very deciding role in the diseaseprediction. Machine Learning algorithms like, Decision Tree(DT), Support vector machine (SVM), K-Nearest Neighbour (KNN), and Artificial neural network (ANN)Algorithms are used to predict thepatient's risk of getting thyroid disease. Web app iscreated to get data from users to predict the type ofdisease. The proposed machine learning algorithms outperformed all other classifiers and achieved the highest accuracy with 97.5%.*

Keywords: *Machine learning, Random Forest, Logic regression, thyroid gland,endocrine gland.*

I. INTRODUCTION

Computational biology advances are used in healthcare organizations. It allows the

ISSN: 2366-1313

collection of patient data for the scientific prediction of disease. Various sensitive prediction algorithms are available for early-grade disease analysis. The medical data system is rich in information sets. However, there are no intelligent structures that can easily analyze the disease. Over time, machine control algorithms play an important role in solving complex and nonlinear problems in developing a prediction version. In all disease prediction models, there is a need for overriding functions that can be selected from unique data sets that can be easily used as a species in a healthy patient with the highest possible accuracy. Otherwise, misclassification can result in a healthy patient enduring an unhelpful treatment. Therefore, waiting for many diseases besides endocrine diseases is fundamental [1].

The endocrine gland is an endocrine gland in the neck. Standing on the miniature part of the human neck, just below Adam's apple, it helps in the secretion of endocrine hormones. It mainly affects the metabolic rate and protein synthesis. Endocrine hormones are useful in many ways, by calculating how fast the heart beats and how quickly calories are burned to control the metabolism within the body. Facilitates the formation of endocrine hormones through endocrine glands to control the body's metabolism. The endocrine glands consist of the active endocrine hormones levothyroxine (T4 abbreviated) and triiodothyronine (T3 abbreviated).) and triiodothyronine (T3) are two energy hormones generally made by the endocrine glands. These hormones are crucial in controlling protein synthesis. It diffuses at body temperature, with electricity transmitted to every body part [2].

For these two types of endocrine hormones, i.e. (T3 and T4), iodine is the main endocrine gluttonous component and is associated with some specific problems, some of which may be distinctly systemic. Inadequate elements of thyroid hormones for endocrine deficiency and details of hyperthyroidism for hyperthyroidism. There are many causes related to hyperthyroidism and hypothyroidism. There are several types of medication. Surgical treatment of the thyroid gland is susceptible to ionizing radiation, persistent thyroid sensitivity, iodine deficiency, and the absence of enzymes to produce endocrine hormones [3].

Computational biology is used for development in healthcare companies. It allows for saving the infected person's data series to predict the disease. Existed prediction algorithms that may be available for early disease diagnosis.

ISSN: 2366-1313

Medical statistics structures are rich in data sets, but only a few sensitive systems can easily assess the disease. Over time, machine learning algorithms have become important in solving complex and non-linear problems in increasing versions. In any disorder, prediction models are used to override the capabilities that can be selected from different data sets that can be used to separate healthy patients as correctly as possible. If this is not always achieved, misclassification may result in a healthy patient receiving unhelpful treatment.

While many parts of the body make hormones, the major glands that make up the endocrine system are the

- Hypothalamus
- Pituitary
- Thyroid
- Parathyroids
- Adrenals
- Pineal body
- The ovaries
- The testes

II. PROBLEM DESCRIPTION

More than half of the Indian population suffers from undiagnosed or misdiagnosed endocrinediseases. Women are seven times more likely to contract endocrineproblems than men and nearlyhalf of all women and a quarter of all men in India will die with evidence of an inflamed endocrine. The symptoms of this disease often vary from person to person and are nonspecific, so a correct diagnosis can easily be missed or misdiagnosed for irrelevant issues.

Finding an accurate solution to this problem for healthcare practitioners via Classification techniques for diagnosing/classifying a particular endocrinedisease that a person may have will cause an immense decrease in misdiagnoses as it is capable of distinguishing between problems of the endocrinegland and other illnesses in the body as well as providing the ability to detect the disease before it forms into a more destructive anomaly. A. Dataset Description the Dataset was extracted from The UCI Machine Learning Repository. This dataset was used for research, development and experimental purposes. The number of instances is 7200 instances and 27 attributes.

III. LITERATURE SURVEY

Parveen Sehgal and Khushboo Taneja [4], proposed a technique fordetecting the thyroid by utilizing the back propagation algorithm. ANN(ArtificialNeural Networks) is developed using the back propagation of error to identifypreliminary endocrine prediction , it is trained using

ISSN: 2366-1313

different training datasets. MATLAB was used to provide experimental results.

Quan Z.Sheng and Wen ChihPeng [5] proposed an expert system comprising of 3 stages usingthe support vector machines model for endocrine disease diagnosis. Ammulu&Venugopalutilized the random forest approach to predict the hypothyroid disorder bycollecting the dataset from UCI repository. The performance measure is calculatedfrom the confusion matrix with the accuracy Shankar and Lakshman proposed amulti kernel support vector machine model and optimal feature selection to classifythyroid patients.

SunilaGodara. [6] They have used Logistics Regression and SVM machine learning Technique to analyze Thyroid Dataset. Comparison was made between these two algorithms based on Precision, Recall, F measure, ROC, RMS error. Logistic Regression turned out has best classifier.

YongFeng Wang. [7] Thyroid Nodule is diagnosed for benign or malignant type using Ultrasound images of thyroid by image analysis - radiomics and deep learning-based approaches. Comparison is made between these two approaches. The classification accuracy, sensitivity, and specificity of applying radiomics based method are 66.81%, 51.19% and 75.77%, respectively, while the evaluation indexes for the deep learning-based method trained to the testing samples are 74.69%, 63.10% and 80.20%, respectively. Deep learning turned out as best approaches.

Hitesh Garg. [8] Feed Forward Neural Network is used for feature extraction and segmentation from Ultrasound images to predict the tumors. The accuracy and other factors were measured and all the average values were above 86%.

IV. PROPOSED METHODOLOGY

Supervised learning is a record mining project that involves inferring a feature of particular educational data. Education facts consist of an order of explanatory setting. In controlled adaptation, each state is a pair consisting of an input object (usually a vector) and a preferred output value (also known as a control flag). A supervised knowledge account checks education records and produces an indirect property. which can be used to draw new artwork. Better optimization will take account into account to efficiently select class names for unseen instances. This requires calculating the summarization from the training records to the hidden conditions in a "realistic" way.

SYSTEM ARCHITECTURE

ISSN: 2366-1313

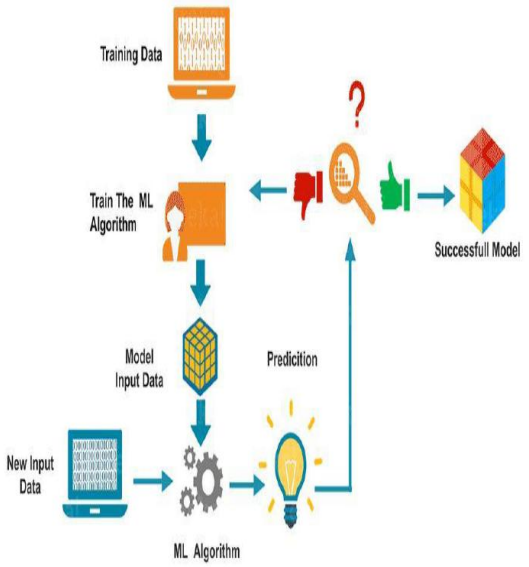


Fig.1 System architecture

A) Decision Tree (DT)Algorithm

A decision tree is a supervised machine learning that defines the algorithm for both type and regression problems. But it's miles generally used in class problems. It is a tree-based classifier containing internal nodes, branches, and leaves where internal nodes represent a dataset, branches represent selection policies, and leaf nodes are rendered. It is a graphical representation showing all possible effects based on given situations. It is known as a decision tree because it is very similar to the shape of a tree, starting from the root node and the branches and ending with the leaves. A set of rules is used to build a CART tree (Classification and Regression Tree). It is entirely based on circumstance as it is divided into sub-trees. We choose the quality attribute in the data set and use the attribute selection measure. There are strategies for trait selection, measurement, information acquisition, and a genetic index.

B) Support Vector Machine

Support vector machine is considered as anassorted research algorithm that helps inperforming the analysis in a precise way.Support vector machine is an approach that is commenced with a concept of an aceof separating hyper plane to aid in the distribution for sampling of data. A hyperplane or multiple planes are created by the support vector machine classifier in highdimensional space. The training data samples are being separated as a positive andnegative data sample by the hyper plane.

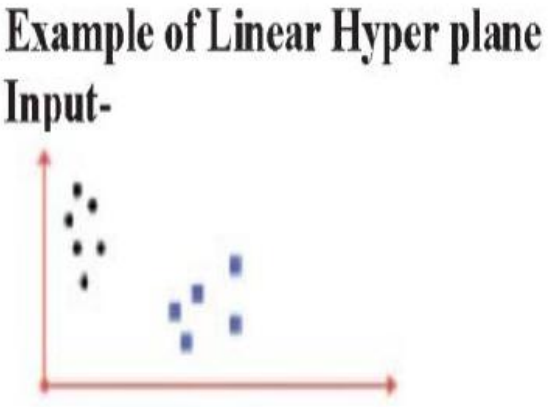


Fig.2 The block dots and blue squares represents two different classes

Output:

ISSN: 2366-1313

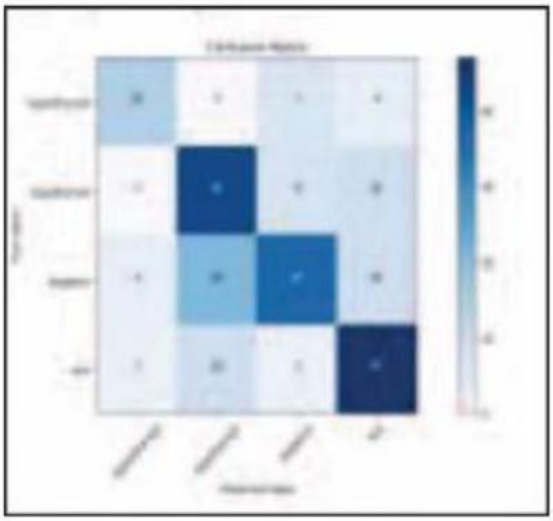


Fig.3 Confusion matrix obtained by the SVM

C) K Nearest Neighbor (KNN):

When given a training tuple K-Nearest Neighbour simply stores it and waits until it isgiven a test tuple. Hence it is a “lazy learner” as it stores the training tuples or the“instances”, they are also known as “Instance Based Learners” [21 22]. Thus, k is apositive integer and decides how ma nyneighbors influence the classification.“Closeness” delineates as a distance metric such as “Euclidean Distance” or“Manhattan Distance”.

D) Artificial Neural Network:

The neural network presents a familiar and practical technique in teaching discrete and vector-valued functions, which is a parallel system that mainly relies on neural machines to master real-valued, discrete and vector-valued functions and is a fully human-based parallel system, which has many corresponding modulators recognized mainly as neurons, acting in a combinatorial manner to solve specific problems. Backpropagation is often used to gain knowledge of ANN technologies. It is a three-layered architecture that is placed inside algorithms within neural networks. It is composed of a three-layer structure: an input layer, a hidden layer, and an output layer. The main layer, the input layer, feeds the input into this backlog, and the second layer, i.e., the hidden layer, gives the output of the input layer. Finally, the output layer transmits the prediction of the population. This mini-grid provides classes with new stats.

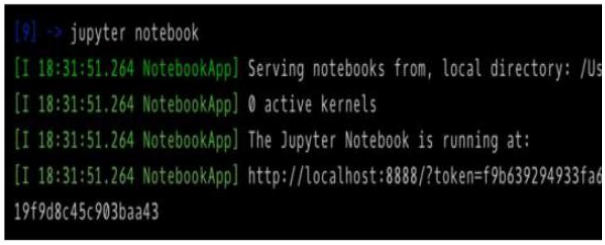
V. RESULTS AND DISCUSSIONS

In these experiments, we are using python programming language The Python programming language is an Open Source, crossplatform, high level,dynamic, interpreted language. The Python 'philosophy' emphasizes readability,clarity and simplicity, whilst maximizing the power and expressiveness available tothe programmer. To implement the project here we are using Jupyter notebook tool.

Jupyter notebook

ISSN: 2366-1313

The Jupyter Notebook is an open-source web application that allows you tocreate and share documents that contain live code, equations, visualizations andnarrative text. Uses include: data cleaning and transformation, numerical simulation,statistical modelling, data visualization, machine learning, and much more.



The below steps shows the project development using jupyternotbook

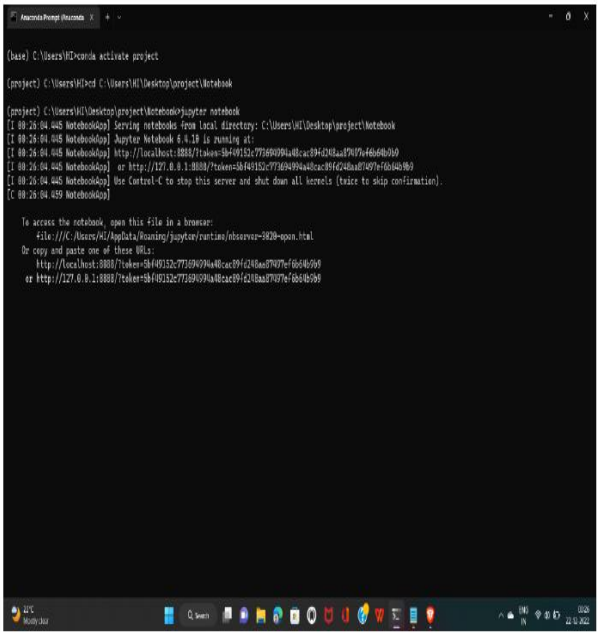


Fig.4 Loading project file

The figure.4 shows the loading of project file

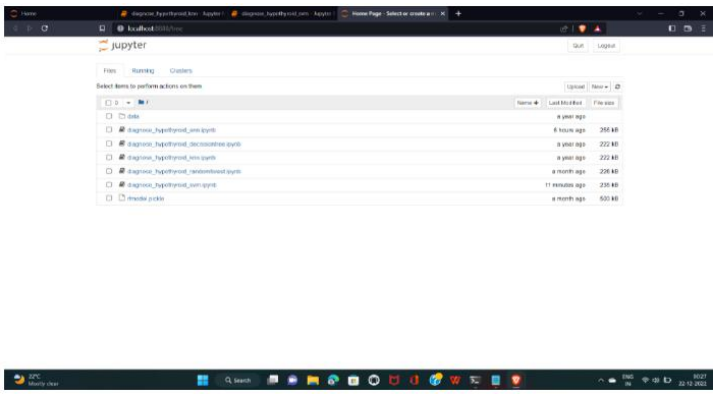


Fig.5 Load jupyter Notebook

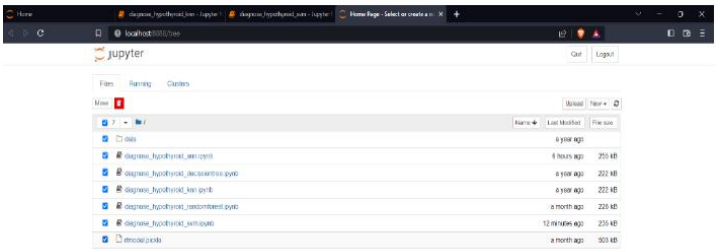


Fig.6 Run the Notebook files

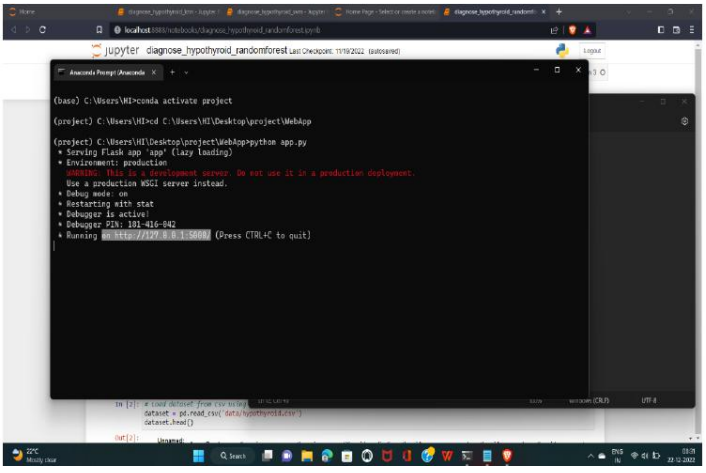


Fig.7 Run the Python URL

ISSN: 2366-1313

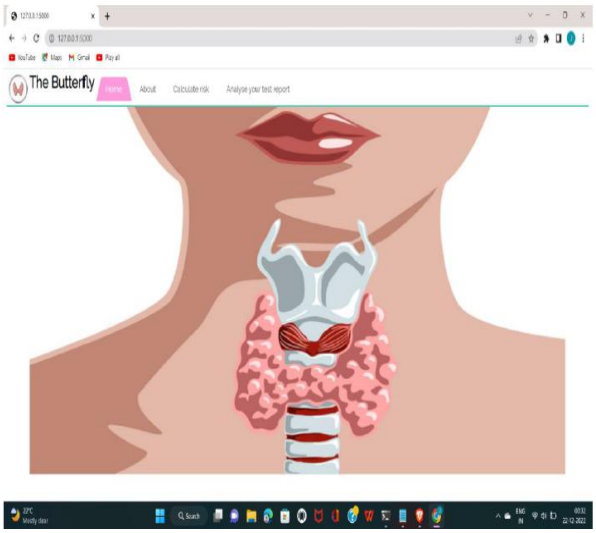


Fig.8 Application interface

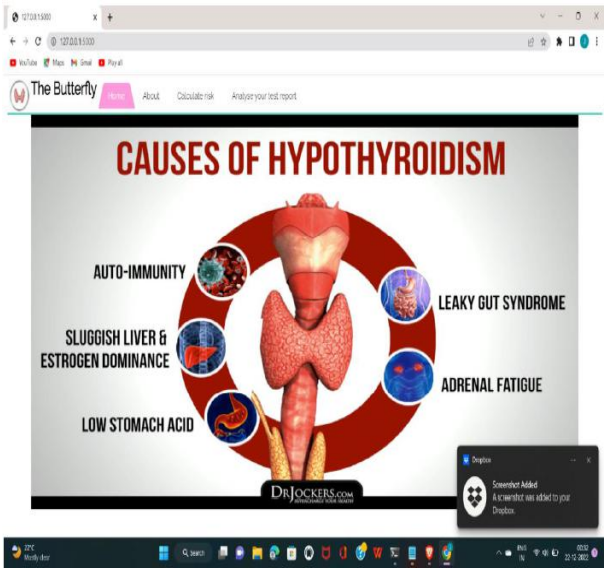


Fig.9 Application Interface

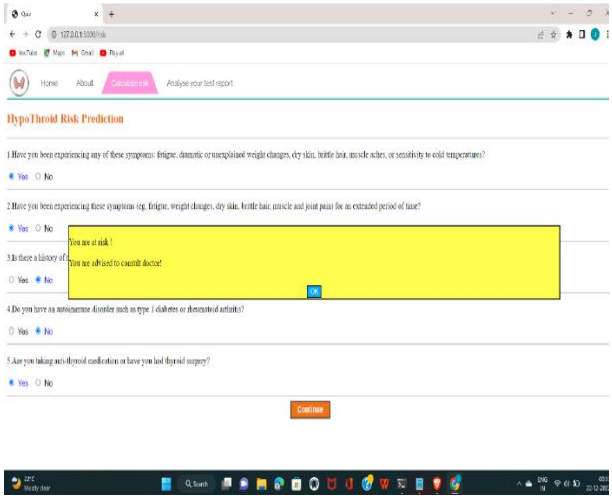


Fig.10 Prediction of HypoEndocrine.

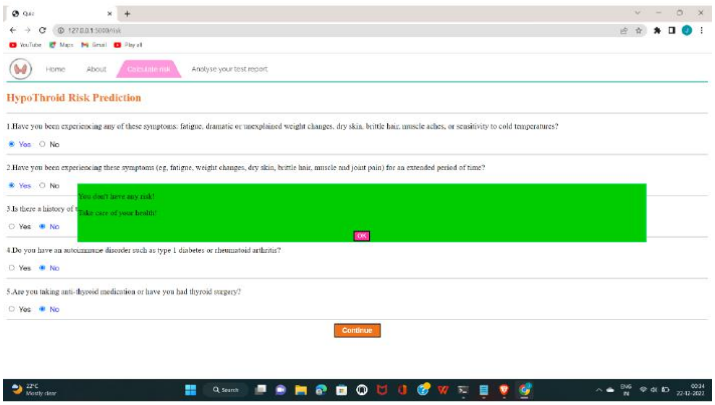


Fig.11 Calculating Hypoendocrine

VI. PERFORMANCE RESULTS

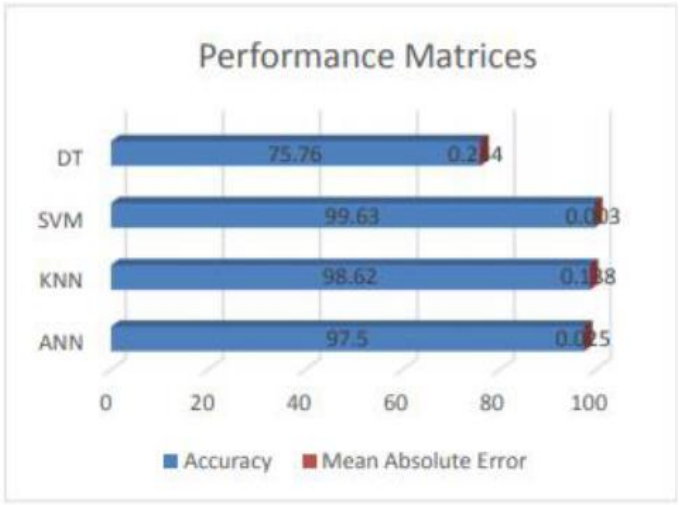


Fig. Final Performance matrices

VII. CONCLUSION

The intent of our work to be done further is to cater the research of idiosyncratic techniques of machine learning that can be mobilized in the diagnosis of thyroid diseases. There are numerous approachable analyses that are delineated and are being used in the latter years of

ISSN: 2366-1313

adequate and competent thyroid disease diagnosis. The analysis shows that different technologies are used in all the papers showing different accuracies. In most research papers it is shown that neural network outperforms over other techniques. On the other hand, this is also given that support vector machine and decision tree has alsoperformed well. There is no doubt that researchersworldwide have attained a lot of success to diagnose thyroid diseases,but it is suggested to decrease the number of parameters used by the patients for diagnosis of thyroid diseases. More attributes mean a patient has to undergo a greater number of clinical tests which is both cost effective as well time consuming. Thus, there is a need to develop such type of algorithms and thyroid disease predictivemodels which require minimum number of parameters of a person to diagnose thyroid disease and saves both money and time of the patient.Indeed, the expansion of our unified representative is a constructive mechanism to predictthyroid disease based on the limited dataset that isavailable with us. The model can further beenhanced to any desired level by increasing thenumber of inputs and outputs and dynamic data canbe generated as more data can be fed to it.

FUTURE ENHANCEMENT

In the future, to better generalize our findings it is necessary to further expand the set of data and attributes considered. With more data the training process is likely to produce more effective classifiers also allowing a more reliable estimate of the exhibited performance. Finally, another aspect that could be investigated concerns the presence of any secondary thyroid disease linked to the patient, to understand if there is a particular additional thyroid disease that can affect hypothyroidism. In fact, it often happens that patients are suffering **from more than one thyroid disease at the same time.**

REFERENCES

[1] L. Ozyılmaz and T. Yıldırım, “Diagnosis ofthyroid disease using artificial neuralnetwork methods,” in: Proceedings of ICONIP’02 9 th international conference onneural information processing (Singapore: Orchid Country Club, 2002) pp. 20332036.

[2] K. Polat, S. Sahan and S. Gunes, “A novel hybrid method based on artificialimmune recognition system (AIRS) with fuzzy weighted pre-processing for thyroiddisease diagnosis,” Expert Systems with Applications,vol. 32, 2007, pp. 1141 1147.

ISSN: 2366-1313

[3] F. Saiti, A. A. Naini, M. A. Shoorehdeli, and M. Teshnehlab,"Thyroid DiseaseDiagnosis Based on Genetic Algorithms Using PNN and SVM," in 3rd InternationalConference on Bioinformatics and Biomedical Engineering, 2009. ICBBE 2009.

[4] G. Zhang, L.V. Berardi, "An investigation of neural networks in thyroid functiondiagnosis," 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, NewYork, 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. StatisticalModeling: the two cultures. Stat Sci. 2001;16:199 231..

[8] Ehrenstein V, Nielsen H, Pedersen AB,Johnsen SP, Pedersen L. Clinicalepidemiology in the era of big data: new opportunities, familiar challenges. ClinEpidemiol. 2017; 9:245 250

[9] Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature.2015;521: 452 459.