USING A PREDICTIVE MACHINE LEARNING MODEL, TREAT THYROID ILLNESS

Dr. Anjaiah Adepu, Professor, CSE, St. Peter's Engineering College(A), Telangana, India
SHANKU VEERABADRAIAH Assistant Professor, CSE, St. Peter's Engineering College(A), Telangana, India,
Thakur Usha rani, Assistant Professor ,Malla Reddy pharmacy College , Telangana, India
[1]anjaiah@stpetershyd.com, [2]veerabadra@stpetershyd.com, [3]thakur.usharani316@gmail.com

ABSTRACT
The thyroid appears to be a component of the endocrine system that is located in the front of the neck and generates thyroxine, which is crucial for our general well-being. If it fails, either too little or too much thyroid hormone will be produced. Processing vast volumes of data requires the use of machine learning techniques and data mining, notably in the health care system where there has been an enormous quantity of information and data that has to be managed. We applied machine learning techniques to our investigation into thyroid illness. In our study, we collected data from individuals, some of whom had hypothyroidisms in addition to overactive thyroid glands, therefore general algorithms were applied. These studies seek to categories this condition into hypothyroidism, ordinary thyroid illness, and hyperthyroidism. KNN, naive-bayes, logistic regressions, decision trees, random forests, discriminant function analysis, and multilayer perceptron's are examples of support vector machines (MLP). to the categorization of thyroid disorders.

KEYWORDS: prediction models, machine learning, and thyroid illness.

INTRODUCTION
One of the diseases that is the most misunderstood and underdiagnosed is thyroid disease, a branch of endocrinology [1] [2]. The second most prevalent endocrine condition in the world, after diabetes, is an issue with the thyroid gland, according to the WHO. Hypothyroidism affects 1% and 2% of persons, respectively, and hyperthyroidism affects 2%. The predominance of men is around one eighth that of women. Thyroid gland dysfunction, which may follow pituitary gland dysfunction or come third to hypothalamus dysfunction, can result in hyperthyroidism and hypothyroidism. A dietary iodine shortage, which occurs in around 15% of cases, can lead to goitre or functional thyroid nodules in some regions. The thyroid gland is susceptible to damage by endogenous antibodies (autoantibodies), and it can potentially be the location of many cancers [3].
According to experts, early sickness detection, diagnosis, and treatment are essential for halting disease progression and mortality. Early diagnosis and potential therapies increase the likelihood of effective therapy for a number of illnesses. Despite several clinical studies, detection is typically regarded as a difficult task [4]. the thyroid gland, which is shaped like a butterfly and is located near the throat's base. Triiodothyroxine (T3), which controls blood pressure, heart rate, and body temperature, and levothyroxine (T4), two functional thyroid hormones, make up this hormone. Also, one of the most prevalent diseases in the nation is thyroid illness, which is typically brought on by an iodine deficiency but can also be brought on by other circumstances. While the thyroid gland seems to be an endocrine organ that produces hormones and transfers them throughout the body. The two forms of thyroid issues are hyperthyroidism and hypothyroidism. Data mining [5] appears to be a method of semi-automated relationship search in huge databases. ML techniques are among the most effective methods for solving a variety of challenging issues [6].

We investigated and classified here simply as ml algorithms play a very important role in characterising thyroid disease as they are effective and 3178 successful and aid in classification [7]. Classification seems to be a method of data gathering (machine learning) that is used for detecting & recognising various diseases, including thyroid cancer. Although machine learning and artificial intelligence have been employed in medicine since their debut [8], there has lately been a drive to address the need for robots to understand healthcare solutions. Because of this, scientists predict that ML will become a common healthcare system in the years to come [9]. When the thyroid gland produces an excessive number of thyroid-related hormones, thyroid hyperthyroidism results. Hyperthyroidism is brought on by an increase in thyroid hormone levels [10]. There are several symptoms, including increased sensitivity to heat, dry skin, weight loss, hair thinning, higher heart rate, excessive perspiration, high blood pressure (BP), neck enlargement, jitteriness, shorter menstrual cycles, hands shaking, and strange stomach movements. The thyroid gland's improper operation makes hypothyroidism a condition. Hypothyroidism results from a reduction in the production of thyroid hormones. Although Hypo is not medical jargon; it simply means "insufficient" or "less." Hypothyroidism has two primary causes: inflammation and thyroid gland issues. A few minor symptoms include low heart rate, obesity, neck swelling, increased temperature sensitivity, hand numbness, hair issues, dry skin, heavy menstrual periods, and stomach discomfort. These symptoms might progressively worsen if they are left untreated [12].

knowledge base and associated works

Several studies in the literature centre on the identification of thyroid disorders utilising hormonal tests and patient-specific information, such as age and sex. For instance, some research uses more complex CNN structures while others use machine learning classifications and predictive models. This study uses a variety of categorized models to identify thyroid disorders based on goitre, TSH, and T4U levels. Many grouping algorithms, like K-nearest Neighbor [13], support the claim. The Naive Bayes and support vector machine algorithms are employed. This research used the quick miner tool, and the findings indicate that K-nearest neighbour is more accurate than Naive Bayes in the diagnosis of thyroid issues. For the purpose of diagnosing thyroid illness, the researchers used data mining classifiers. Thyroid problems are a crucial factor to take into account while diagnosing an illness. KNN and Naive Bayes classifiers were used in this study. The Fast miner software is being compared with these two classifiers. In comparison to Naive Bayes classifiers, which had an accuracy of 22.58%, it was discovered that the K-nearest neighbour classifier had 93.44% accuracy. The recommended KNN technique improves classification accuracy, leading to improved results. Naive Bayes provides linear, circular, or parabolic decision boundaries as a result; hence, KNN's decision boundary consistency is highly favourable. KNN outperforms most approaches because the components are interconnected. Diagnosing sickness in the world of medicine is a challenging endeavour. A range of data \smining methods are applied in the decision-making procedure. We used J48 and other classification algorithms for decision making data mining to characterise hypothyroidism, and we decreased dimensionality to determine the selection of attributes from the prior findings. The uncertainty matrix is used to evaluate the \sprecision and error rate of a classifier's output. J48 Algorithm provides higher Compared to the decision stump tree, this method has a higher accuracy percentage (99.58%) and a lower rate of error [14]. Under supervised learning, classification is a key data mining technique that is employed to define predetermined data sets. In the healthcare sector, classification is widely used to guide decision-making for both medical diagnosis and oversight. The information used in this study was obtained from a famous lab in Kashmir. The entire research study will be conducted using the ANACONDA3-5.2.0 platform. In experimental research, many classification techniques, such as Decision Trees, K Closest

Neighbors, Naive Bayes, etc., may be used. The Judgment Tree is the most accurate of the four classes, with a 98.89 percent accuracy rate [15]. Thyroid illness is a chronic disorder that affects people all over the world. Data mining in healthcare has proven to produce excellent results in the prognosis (prediction) of many illnesses. High levels of accuracy and low prediction costs are characteristics of data mining techniques. The fact that prediction just requires a little period is another significant benefit. It is not as accurate as Random Forest and J48. We found that the Random Forest model had a prediction accuracy of 99.3%. Nonetheless, the model's development took longer than the next two versions. As a consequence, J48 is the best-yielding model for hypothyroid prediction, with 99% accuracy and an execution time that is significantly smaller than Random Forest (0.2 sec) [16]. The objective of this position is to provide a data mining-based strategy for increasing the accuracy of hypothyroidism diagnostics by fusing patient inquiries and test results throughout the diagnosis process. Lowering the risks connected to dialysis interventional trials is another goal. The decisive answer if fresh samples \sare hypothyroid, was deduced using statistics from machine learning.

Many sampling techniques were used during the data collection process to remove the unbalanced distribution, and hypothyroidism models were built utilising K Nearest Neighbor, Logistic Regression, and SVM classifiers. In this situation, the

The impact of sampling techniques on the analysis of hypothyroidism was established by hypothesis. The Logistic Regression classifiers produced the greatest outputs \samongst the entire model created. Accuracy - 97.8%, F-Score - 82.26%, RoC - 93.2%, and Matthews Correlation Coefficient - 81.8% [17] are the performance metrics for the study's model trained with the dataset using oversampling methodologies. The objective of this work is to create a machine learning system that accurately and early detects diabetes in patients using the Random Forest approach. For the application's classification and regression phases, Random Forest techniques are employed. When compared to other approaches, the performance ratio is higher. The results demonstrate that the system is accurate, efficient, and rapid in detecting diabetes, with the suggested model producing the best prediction results for diabetes diagnosis. [18]. After all other cancers, breast cancers are the second most common malignancy in women. The goal of the research paper is to give breast diagnostics utilising cutting-edge techniques.

Integrating new scientific breakthroughs into breast cancer survival prediction \smodels. We used three common data mining techniques and a sizable dataset to create prediction models. We used 10-folded cross-validation techniques to assess the objectivity of all the topic prediction models evaluation. Findings indicate that it is safe to visit the Bay. The next-best predictors are the RBF Networks, with 93.44% accuracy on the proposal sample, and the Nave Bayes, with 97.36% accuracy on the holdout sample. In this work, we evaluated several breast cancer survivability prediction models, whether benign or malignant, using two criteria [19]. In a more recent study, the classification of thyroid disorders into the two most prevalent thyroid dysfunctions in the general population was the emphasis. Researchers examined and compared four classification models: decision trees, multilayer perceptrons, naive bayes, and radial basis function networks. The findings demonstrate the high accuracy of all of the aforementioned classification models, with the Decision Tree model producing the highest level of score categorization. Romanian data website and the UCI machine learning repository's data were utilised to create and assess the classifier. The classification models were \sdeveloped and evaluated utilising data mining techniques as the backbone. Through a literature review, a variety of learning about thyroid classification is employed various data mining techniques to create powerful classifiers. The authors \sexamined how four classification models may be utilised on thyroid data to aid \scategorise thyroid dysfunctions including hyperthyroidism along with \shypothyroidism. In every situation that was looked at, the decision tree model served as an appropriate classification model.

proposed strategy

Data Gathering

Methods are used in the quick and successful treatment of thyroid issues as well as many other ailments since they now hold a significant relevance in the healthcare field and help us in the diagnosis and categorization of disorders.

As a result, we were able to compile a lot of data on thyroid disorders, which we then utilised in our investigations to identify diseases. The information was gathered from Iraqi residents, with a particular emphasis on thyroid disease. A total of 1250 individuals, both men and females, with ages ranging from one year to one year, had their data collected. These samples comprised healthy persons without thyroid illness as well as patients with hypothyroidism and hyperthyroidism. The data was acquired over a period of one to four months with the intention of classifying thyroid illnesses using machine learning techniques. Among the information acquired are age, gender, triiodothyronine (T3), thyroid stimulating hormone (TSH), thyroxine (T4), and a few additional features. Due of the 17 characteristics or features that were included in the data collection, including (id, gender, age, sick, thyroid surgery, pregnant, query hyperthyroid, etc.), all of which have been included in our research.
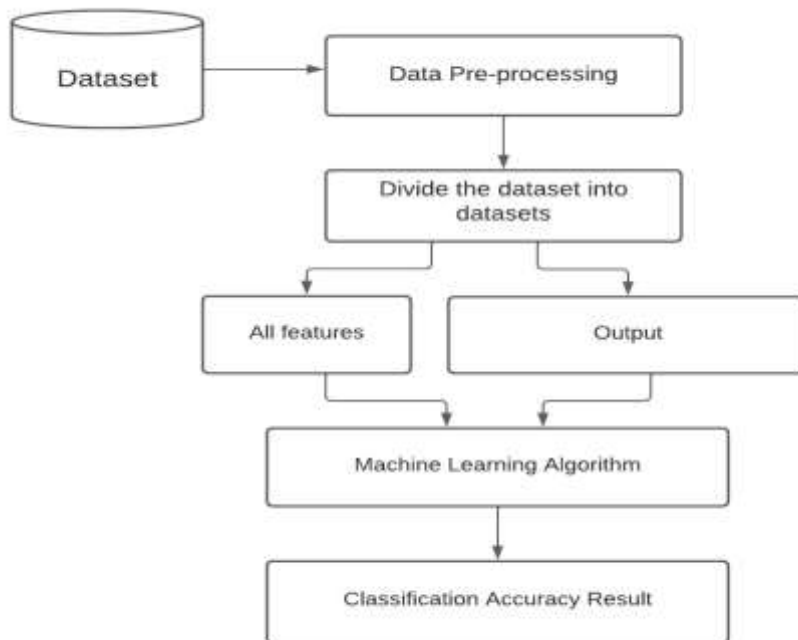

Fig. 1 Representing Architecture of data processing

Data Pre-processing

Pre-processing statistics is a very critical and important step in the data extraction process because it has a beneficial effect on the data and is used to reveal information by analyzing the available data and locating the missing data because it thoroughly examines the data. Cleansing data, preparing the data, and so on are all \spart of the pre-processing process. In this phase or step, we cleaned and organized the records that we were able to obtain, and within them, we identified a set of incomplete data in these statistics where the lacking characteristics were recognized. These missing characteristics were T4 by

quantity 155 and T3 by quantity 115, and within these records, we identified a set of incomplete information in this statistics where the lacking characteristics were recognized. After working in this method, we were able to obtain the information in a higher as well as positive way and be freed of missing data. We were able to interpret this erased information by replacing it with the valuation of an intermediate. We utilised normalisation strategies while using the MLP approach.

Approaches for Data Machine Learning
Three different thyroid diseases may be distinguished using machine learning methods. The very first category is hyperthyroidism, followed by hypothyroidism and stable individuals without thyroid problems.

Stability Vector Machines
The SVM method, a machine learning-based algorithm associated with supervised learning and data mining techniques, may be used to identify the most accurate energy use predictions. The researchers used well-known categorization techniques to answer our question, including optimal boosting trees, subset selection, and generalised additive model. To find a subset of predictor which \smore strongly predicted eating with a linear connection, we utilised \sbackward, forward and optimum subset selection. The SVM offered a method that employed a tree-based algorithm and iterative binary splitting to divide the predictor spaces into sample regions. The boosting tree method was selected since it was originally thought to be one of the most common tree-based algorithm solutions. SVM is overly effective at dealing with data with a lot of dimensions.
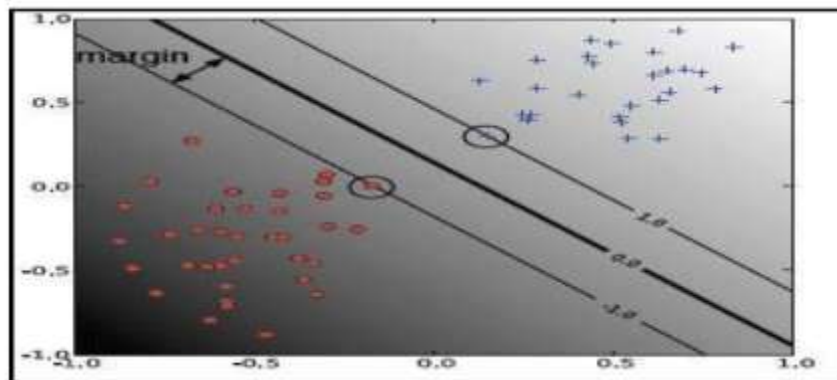
Support vector machine classifiers, Figure 2.

Random forest
The random forest estimates the average response of each energy consumption \spredictor. The overall amount of distance between each response would come from the database's averages from each sample, and the random forest would then add the relative distance between each response based on the mean of each predictor.

In each sample, those who were frequently removed from the answer variable will have a high maximum distance. Rates that repeatedly categorise the data were found using a computer that calculated the mode of each response. If the kind of response accounted for more than 90% of the total number of questions, the answer was flagged by the study as potentially excessive in power consumption. There have been many responses that have been recorded.
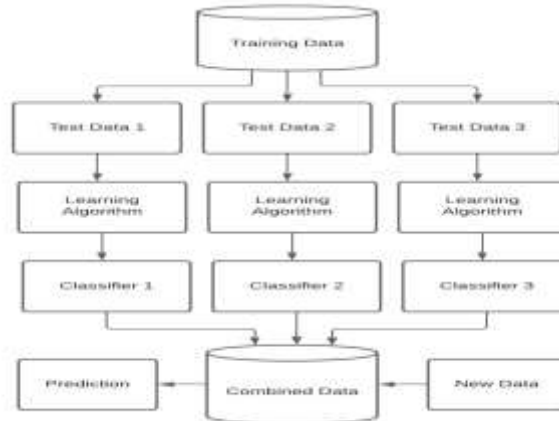
Fig. 3: RF categorization flowchart

Choice Tree

The goal of the decision tree based algorithm is to use a tree-based approach to determine the most significant predictors of consumption by relying on the decision-boosting system that was examined for predicting energy usage component. To do this, I used the decision tree approach. In order for the product to evolve over time, hundreds of trees that are created utilising information from the previous tree must be fitted. The shrinkage value, the number of trees, and the number of breaks in each tree are all tuning parameters.

Naïve Bayes

Many generalized additive models with independent variables for subset selection, variables with the greatest relative influence on classification, and a combination of distinct components from both may be compared using Nave Bayesian. It directly compared the forecast accuracy of the top models. It reduced correlations between individual predictors and responses by fitting naive Bayes using a variety of splines, two-degree polynomials, and sequential predictor variables. We employ a greater proportion of polynomials and splines to build predictors with nonlinear connections to our response variable.
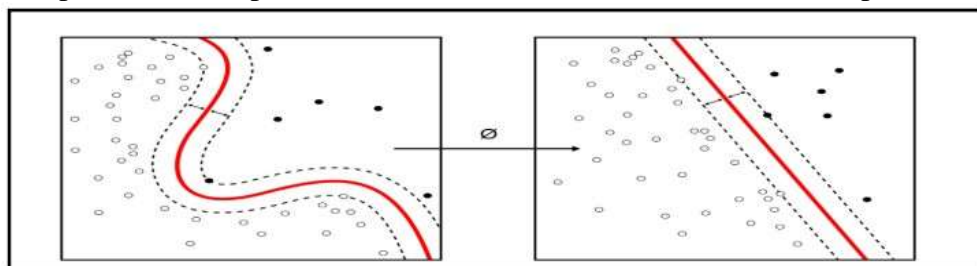
Naive Bayes representation in Figure 4

Rational Regression

The Supervised Learning techniques, which include logistic regression, are frequently used in Machine Learning (ML) algorithms. Using a set of independent variables, these techniques are used to estimate a dependent categorical variable.

A dependent categorical variable's overall contribution is predicted using LR. The value obtained must thus be categorical or singular. It may return probabilistic values in the middle rather than exact numbers like 0 and 1, such as false or true, 0 or 1, No or Yes, and so on. Logistic regression resembles linear

regression more in terms of application base. And in this way to solve \sregression issues, linear regression is applied, whereas logistic regression is \sbeing utilise to solve classification difficulties [25].

KNN

In contrast to older methods, the k-nearest neighbour algorithms employ the data immediately for classification rather than first building the models [26]. With this form of output, no more model construction is required, and the only part of the model that can be utilised to calculate the class label is the number of nearest neighbours, or k. The result of p(y/x) is essentially the ratio of elegant users between some of the kth nearest neighbours of x. Variations in k's value have an impact on the stability of this model. The ease of usage of KNN above all other algorithms is one of its main advantages. It's possible that neighbours can explain the categorization result. When black-box models are inadequate, using case-based reasoning may be helpful. The determination of the \scase neighbourhood, which needs establishing.

Multi-Layer Perceptron's

an advance feed multi-layer perceptrons are an artificial neural network (ANN) that builds a sequence of results from a set of inputs. Many layers of input nodes form directed graphs between the input and output layers of an MLP. Back propagation is used by MLP to trains the networks. A DL approach is MLP. The multilayer perceptron's are neural network which joins multiple layers in the directed graphs, with the signal going entirely in single direction among \snodes. Each node has a non-linear activation function aside from the input's node.

MLPs employ the supervised learning method known as back propagation. DL is MLP based strategy that utilises several layers of neurons. There are a few problems with supervised learning, including examination of concurrent distributed processing.

MLP is another technique in computational biology that is often used. Among the applications are speech recognition, image recognition, and automatic translation [28].

Linear Analysis

It's one among the main popular approaches for lowering dimensionality. Among \sother things, it's useful in the pattern categorization and machine learning \stechniques. LDA is a technique for converting objects from 3D to 2D space. It is done to prevent common dimensional issues and reduce operational and real estate costs. A classifier method called discriminant function analysis is used to develop machine learning techniques. Predictions and image identification are only two examples of the many applications that make use of framework for decreasing dimensionality [29].

Result analysis

The suggested research employed a vast range of machine learning (ML) techniques to \sanalyse our data (Random Forest, Decision Tree, Logistic Regression, SVM, Naive \sBaye, Linear Analysis, KNN, and Multi-Layer Perceptron). As this is the first time, we have utilised these data, we first split the dataset in two, using 70% of it for training and the remaining 30% for testing. At the initial step, we applied the characteristics in our data to the range of mechanics specified in the tables below, and the results of the application procedure were shown. This \spractical component was created in Python and is deemed a complete and \svarious interconnected systems. All sixteen items are treated as inputs and one \soutput have been got held into account.

Table 1: Representing all other classification models' algorithms

S.	Apply Algorithms	Accuracy

No		
1	MLP	97.7
2	DT	98.54
3	RF	98.98
4	SVM	92.32
5	KNN	91.25
6	LR	91.48
7	NB	82.12

After deleting the three features, the Naive Bayes method appears to have a \shigher accuracy of about 90.67, whereas the logistic regression technique, SVM \salgorithm, and K-Neighbours Classifier algorithm have risen greatly in accuracy \swhereas the accuracy value of the remaining algorithms has declined. Experience has shown that the accuracy values of the various algorithms applied on the dataset above alter depending on the data's quality vary. The accuracy of certain algorithms decreased while the accuracy value of others increased when three of the characteristics were removed from the algorithms.

Conclusion

The prevalence of this disease, which affects the whole population of the country, is growing. In light of medical data that show significant irregularities in thyroid illnesses, our study is focused on classifying thyroid disease into hypothyroidism and hyperthyroidism. Algorithms \swere employed to categorise this sickness. Combining many strategies and building two models using machine learning resulted in successful outcomes. In the first model, which had 16 inputs and 1 output, the random-forest method's accuracy value was 98.97%, which was the highest accuracy number among all examined methods. According to earlier study, the second incarnation lost the following characteristics: One of the qualities that was deleted was thyroxine. 2-search for hyperthyroid 3 searches for hypothyroid. In this part, we've included both the improved accuracy value of certain algorithms and the control of the accuracy value of others. The accuracies of the Naive Bayes \smachine learning algorithms were found to grow by 90.67 percent. The greatest precision of the MLP algorithm is 96.4.

REFERENCES

1. AKGÜL, Göksu, et al. "Hipotiroidi Hastalığı Teşhisinde Sınıflandırma Algoritmalarının Kullanımı." Bilişim Teknolojileri Dergisi 13.3 (2020): 255268.
2. Aswad, Salma Abdullah, and Emrullah Sonuç. "Classification of VPN Network Traffic Flow UsingTime Related Features on Apache Spark." 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). IEEE, 2020.
3. Aytürk Keleş and Keleş, Ali. "ESTDD: Expert system for thyroid diseases diagnosis." International Research Journal of Engineering and Technology (IRJET) Volume: 03 Issue: 11 | Nov -2017 34.1 (2017):242- 246
4. Azar, a.T, Hassanien, A.E. and Kim, T. Expert system based on neural fuzzy rules for thyroid disease diagnosis, Computer Science, Artificial Intelligence, arXiv:1403.0522, Pp. 1-12,2012.
5. Banu, G. Rasitha. "A Role of decision Tree classification data Mining Technique in Diagnosing Thyroiddisease." International Journal of Computer Sciences and Engineering 4.11 (2016): 64-70.
6. Banu, G. Rasitha. "A Role of decision Tree classification data Mining Technique in Diagnosing Thyroid disease." International Journal of Computer Sciences and Engineering 4.11 (2016): 64-70.

7. Begum, Amina, and A. Parkavi. "Prediction of thyroid disease using data mining techniques." 20195th International Conference on Advanced Computing & Communication Systems (ICACCS). IEEE, 2019.

8. Bolotnikova, Anastasia, Hasan Demirel, and Gholamreza Anbarjafari. "Realtime ensemble based face recognition system for NAO humanoids using local binary pattern." *Analog Integrated Circuits and Signal Processing* 92, no. 3 (2017): 467-475.

9. C. Fan, F. Xiao, Z. Li, J. Wang. Unsupervised data analytics in mining big building operational data for energy efficiency enhancement: A review. Energy Build. 2018, 159, 296–308.

10. Chandel, Khushboo, et al. "A comparative study on thyroid disease detection using K-nearest neighbour and Naive Bayes classification techniques." CSI transactions on ICT 4.2-4 (2016): 313-319.

11. Chandio, Jamil Ahmed, et al. "TDV: Intelligent system for thyroid disease visualization." 2016International Conference on Computing, Electronic and Electrical Engineering (ICE Cube). IEEE, 2016.

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

13. D. Mora, G. Fajilla, M. Austin, D. Simone. Occupancy patterns obtained by heuristic approaches: Cluster analysis and logical flowcharts. A case study in a university office. Energy Build. 2019, 186, 147–168

14. Dasarathy B. Nearest neighbor pattern classification techniques. Silver Spring, MD: IEEE Computer Society Press; 1991.

15. Dr. Srinivasan B, Pavya K "Diagnosis of Thyroid Disease: A Study" International Research Journal of Engineering and Technology Volume: 03 Issue: 11 | Nov – 2016

16. Dreiseitl, Stephan, and Lucila Ohno-Machado. "Logistic regression and artificial neural network classification models: a methodology review." Journal of biomedical informatics 35.5-6 (2002): 352-359.

17. Gupta, Puneet, Vijay Kumar Sharma, Naman Mittal, Raghav Bansal, and Himanshu Gupta. "Ai Enabled Virtual Environment Simulator." (2020).

18. Heuck, "World Health Organization," 2000. [Online]. Available: https://www.who.int/.

19. Iqbal, Md, Vimal Kumar, and Vijay Kumar Sharma. "Krishi Portal: Web Based Farmer Help Assistance." International Journal of Advanced Science and Technology 29 (2007): 4783-4786.

20. Keles, A. ESTDD: Expert system for thyroid diseases diagnosis, Expert Syst Appl., Vol. 34, No.1,Pp.242–246, 2008.

21. Khushboo Taneja, Parveen Sehgal, Prerana "Predictive Data Mining for Diagnosis of Thyroid Disease using Neural Network" International Journal of
Research in Management, Science & Technology (E-ISSN:2321- 3264) Vol. 3, No. 2, April 2016

22. Kouroua, K., Exarchosa, T.P. Exarchosa, K.P., Karamouzisc, M.V. andFotiadisa, D.I. (2015) Machinelearning applications in cancer prognosis and prediction, Computational and Structural BiotechnologyJournal, Vol. 13, Pp.8–17.

23. Ripley B. Pattern recognition and neural networks. Cambridge: Cambridge University Press; 1996.

24. Sharma, Vijay Kumar, Vimal Kumar, Md Iqbal, Sachin Tawara, and Vishal Jayaswal. "Virtual Mouse Control Using Hand Class Gesture."

25. Shukla, A. & Kaur, P. (2009). Diagnosis of thyroid disorders using artificial neural networks, IEEE International Advance computing Conference (IACC 2009)– Patiala, India, pp 1016-1020.
26. Sindhya, Mrs K. "EFFECTIVE PREDICTION OF HYPOTHYROID USING VARIOUS DATAMINING TECHNIQUES."
27. Tiwari, Durgesh, and Vijay Kumar Sharma. "A Review on Conventional and Lightweight Security Techniques in Mobile and IoT Devices."
28. Travis B Murdoch and Allan S Detsky. The inevitable application of big data to health care. Jama,309(13):1351–1352, 2013.
29. Tyagi, Vidhi, Shivam Arora, Sattyam Gupta, Vijay Kr Sharma, and Vimal Kumar. "Architecture of an IoT-based Women Safety System." *Architecture* 29, no. 5 (2020): 3670-3676.
30. Umar Sidiq, Dr, Syed Mutahar Aaqib, and Rafi Ahmad Khan. "Diagnosis of various thyroid ailments using data mining classification techniques." Int J Sci Res Coput Sci Inf Technol 5 (2019): 131-6.
31. V. Cerqueira, L. Torgo, M. Mozetic. Evaluating time series forecasting models: An empirical study on performance estimation methods. Mach. Learn. 2020, 109, 1997–2028.
32. VijiyaKumar, K., et al. "Random Forest Algorithm for the Prediction of Diabetes." 2019 IEEEInternational Conference on System, Computation, Automation and Networking (ICSCAN). IEEE, 2019.
33. W. Kleiminger, C. Beckel, T. Staake, S. Santini. Occupancy Detection from Electricity Consumption Data. In Proceedings of the 5th ACM Workshop on Embedded Systems for Energy-Efficient Buildings, Rome, Italy, 14–15 November 2013; pp. 1–8.