ISSN No:-2456-2165

Application of Machine Learning Techniques in Prediction of Breast Cancer, Thyroid, and PCOS: An Overview

Akanksha Gautam, Sunita Jalal, Chetan Singh Negi
Department of Computer Engineering, College of Technology, GBPUAT Pantnagar

Abstract:- Because of the rapid population growth in diseases in recent years, early disease identification has become a critical problem. With massive population growth, the probability of death from breast cancer is increasing exponentially. Breast Cancer is the second leading most severe of the cancers that have already been identified, and following this, others are Thyroid and Polycystic Ovary Syndrome (PCOS) disease. An automatic disease diagnosis system assists medical personnel in disease detection, provides dependable, effective, and immediate responses, and reduces the death risk. In this paper, we study different machine learning techniques and their utility in predicting breast cancer, thyroid and PCOS to get the best result. The performance measures such as accuracy, specificity, sensitivity, precision, recall, F1 score, and receiver operating curve are discussed to assess the performance of machine learning algorithms. The paper explores the research work using machine learning algorithms done in detecting breast cancer, thyroid, and PCOS.

Keywords:- *Breast Cancer, Thyroid, PCOS, Dataset, Machine Learning.*

I. INTRODUCTION

Women's health varies from men's in numerous ways. Women's health is an illustration of population health, which the World Health Organization defines as "a state of complete physical, mental, and social well-being, rather than merely the absence of disease or infirmity." While the accelerations of the foremost causes of death, cancer, cardiovascular and lung diseases, are comparable in men and women, women have distinctive experiences. Lung cancer has surpassed all other cancers as the foremost reason for cancer death in females, followed by various cancers such as breast cancer, ovarian cancer, colorectal cancer, uterine cancer, and cervical cancer.

- *Breast Cancer:*

Today, breast cancer is a common and dangerous condition. According to World Health Organization research, 1,000,000 women are practically sure to acquire breast cancer, and half will die because the disease is often detected late. One in every eight women in the US will be diagnosed with breast cancer. Breast cancer is the most dominant breast malignant disorder. Malignant cells evolve destructively and uncontrolled way. Breast cancer symptoms include lumps in the breast, shape changes, skin dimpling, and flaky patches. Changes in eating habits, a lack of fitness, chemical substitution therapy during the menopausal phase, and consistent family history are all significant causes of these manifestations. According to estimates, approximately 316,120 females were diagnosed with various breast cancers in the USA in 2017.

- *Thyroid:*

Thyroid disease is one of the most common public health issues, particularly among pregnant women. It harms metabolism, mental functions, energy level, weakness, and bowel movements. Most thyroid diseases diagnosed during pregnancy are caused by a lack or excess of iodine in the diet. Thyroid disease poses a significant challenge to pregnant women's physiology and has substantial maternal and foetal implications. There is a growth in thyroxine (T4) and triiodothyronine (T3) generation in women with normal thyroid function. It hampers thyroid-stimulating hormone (TSH) in the first three months of pregnancy due to an increased human chorionic gonadotropin (hCG) level, which enables the TSH receptor.

- *PCOS:*

PCOS is a disease in which the ovary creates a varying amount of androgens hormone, generally present in small amounts in women. During the reproductive period, approximately 6-10% of women develop Polycystic Ovary Syndrome (PCOS). It manifests as various symptoms in women of reproductive age resulting from a hormonal disorder. PCOS manifests itself in three ways. The first is a menstrual irregularity, followed by an increase in male hormones in women, and finally, an ultrasound-detected polycystic ovarian structure. Polycystic Ovary Syndrome is suspected in women who exhibit these symptoms, but the diagnosis is determined by examining ultrasound images and blood test results.

Machine learning (ML) refers to an algorithm's ability to 'learn' by detecting patterns in large datasets. In other words, the 'outcomes' yielded by machine learning algorithms are statistical inferences drawn from huge datasets. Machine learning techniques utilise vast amounts of computing power to determine rules the human brain would be unable to process. The more data fed into a machine learning model, the more

complex the rules become – and the more exact the predictions become. It is an artificial intelligence application in which computers are programmed to mimic how humans think and learn. Machine learning helps in healthcare collect and manage patient data, identify trends, recommend treatments, and more. Hospitals and healthcare organisations have begun recognising machine learning's ability to improve decision-making and reduce risk in the medical field, resulting in a slew of new and exciting career opportunities. Machine learning aims to enhance patient outcomes and generate previously unavailable medical insights. Jalal et al. [22] discussed the use of machine learning techniques: Latent Dirichlet Allocation and k-Medoids in web services clustering. Web services can provide access to different modules of healthcare application software. Arushi et al. [23] explored the application of various machine learning models in Covid-19 prediction. The paper is organized as follows. Section 2 gives descriptions of breast cancer, thyroid, and PCOS datasets available in different repositories. The overview of various machine learning algorithms and performance measures is presented in section 3. Section 4 outlines the related work in detecting breast cancer, thyroid, and PCOS. The conclusion of the paper is drawn in section 5.

II. DATASETS DESCRIPTIONS

This section gives an overview of different dataset used by researchers in detection of breast cancer, thyroid, PCOS.

A. Breast Cancer Datasets:

Wisconsin Breast Cancer Dataset1 [24]: The dataset is available on the UCI repository, and it is of the year 1992. It contains 699 instances where the cases are either benign or malignant. There are 458 benign cases (65.50%) and 241 malignant cases (34.50%). The dataset has two classes such as 2 or 4, where 2 represents the benign case, and 4 illustrates the malignant case. The dataset contains the following attributes: Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, and Marginal Adhesion.

Wisconsin Breast Cancer Dataset2 [25]: The dataset is published in the year 1995. It is available on the kaggle repository. The dataset contains 569 samples. The dataset is popularly used for many applications because it includes noise-free instances. The dataset considers ten attributes: texture, radius, smoothness, area, perimeter, concave points, symmetry, compactness, concavity, and fractal dimension. Attributes are real-valued.

MIAS Dataset [26]: Some researchers used MIAS dataset. They used machine learning techniques on dataset to detect breast cancer in mammograms. The dataset contains 322 sample instances. In which nearly 189 images were of normal cases and 133 images were of abnormality. The attributes presented in the dataset are defined in Table 1:

Table 1: Details of Attributes

Name	Type
Character of background tissue	Nominal - F: Fatty, G:Fatty-glandular, D:Dense-glandular
Classes of Abnormality	Nominal- CALC, CIRC, SPIC, MISC, ARCH, ASYM, NORM
Severity of Abnormality	Nominal: Benign, Malignant
x image coordinate of centre of abnormality	Integer
y image coordinate of centre of abnormality	Integer
Approximate radius of a circle enclosing the abnormality	pixels

B. Thyroid Dataset:

The different thyroid disease data sets are available on UCI machine learning repository. These source datasets are classified using different machine learning algorithms by various researchers.

Throid dataset [27]: The dataset contains 7200 instances and 29 attributes. Some important attributes are age, sex, Thyroid stimulating hormone (TSH), triiodothyronine(T3), Thyroid binding globulin(TBG), Thyroxin utilization rate(T4U), Total Thyroxin, Free(TT4) Thyroxin Index (FTI).

C. PCOS Dataset:

Machine learning algorithms are used to diagnose polycystic ovary syndrome (PCOS) in women using a dataset freely available in the Kaggle repository, the dataset is called Polycystic Ovary Syndrome.

PCOS dataset [28]: The dataset is available on github. It contains 119 instances with 18 attributes. Some of the attributes are: Regularity of menstrual periods, weight gain, depression and anxiety, history of diabetes and hyper tension, fast food intake.

PCOS dataset [29]: The dataset is available on Kaggle repository. It contains 541 instances with 43 attributes. Clinical parameters are TSH, FSH, Hb, AMH, LH Vit, D3, PRG, and other metabolic, hormonal, and biochemical indicators. Whereas physical factors included measurements of height, weight, age, waist, and hip.

III. OVERVIEW OF MACHINE LEARNING TECHNIQUES AND EVALUATION MEASURES

This section explores different machine learning techniques such as Logistic Regression, Linear Regression, Naive Bayes classification technique, K-Nearest Neighbors (KNN), Decision Tree Random Forests, Artificial Neural Networks (ANNs), Support Vector Machine (SVM), and deep learning. Some evaluation measures such as accuracy,

sensitivity, specificity, precision, recall, and F1 score are discussed.

A. Machine Learning Techniques

➢ *Logistic Regression:*

Logistic Regression is an analytical modeling technique that associates a level's likelihood with a set of explicative variables. It analyses a dataset in which one or more independent variables determine the outcome. A binary variable is used to assess the outcome (in which there are only two possible results). Given a set of independent variables, it predicts a binary result (True/False, 1/0, Yes/No).

➢ *Linear Regression:*

The linear regression algorithm shows a linear relationship between a dependent (y) variable and one or more independent (x) variables. Because linear regression demonstrates a linear relationship, it determines how the value of the dependent variable changes with the value of the independent variable. Different regression models differ in the type of relationship they consider between dependent and independent variables, as well as the number of independent variables used.

➢ *Naive Bayes Technique:*

The Nave-Bayes method is a supervised learning algorithm that customs the Bayes theorem to resolve classification difficulties. It is a simple and effective classifier primarily used for text classification with an extensive training set. The Nave Bayes Classifier helps develop fast machine learning models capable of constructing quick predictions. It proposes classification based on an object's probability.

➢ *Decision Tree:*

The decision tree classification method classifies training data into a tree. We can find rules from the tree. The performance of the decision tree is determined using test data drawn randomly from training data using the hold-out method.

➢ *Random Forests:*

Random Forest classifier utilises many decision trees on various subsets of a dataset and computes average accuracy to enhance the model's predictive accuracy. Rather than depending on the decision of a single decision tree, the random forest considers the assumptions from each tree, and it computes the final outcome based on the majority of votes concept. The considerable number of trees increases the accuracy and lowers the risk of overfitting. The random forest classifier is an extremely effective supervised classification tool. The RF classification is an ensemble method that can be thought of as a type of nearest neighbour predictor. Ensemble learning is the process of strategically developing and incorporating statistical methods such as classifiers or experts to solve a specific computational intelligence problem. Instead of a single classification tree, RF generates a forest of classification trees from a given dataset.

➢ *K-Nearest Neighbors:*

The supervised classification algorithm k-Nearest Neighbors (K-NN) takes a large number of categorized points and uses them to show itself how to tag other points. The K-nearest neighbor algorithm is used for pattern recognition and grouping. It is widely used in forecasting. When new data arrives, the K-NN algorithm identifies existing data points that are closest to it. Any attribute that varies on a large scale may have a significant impact on the interval between data points. K-NNs are based on the assumption that data samples are represented in a metric space.

➢ *Support Vector Machine (SVM):*

The support vector machine is an extension of a natural classifier known as the maximal edge classifier. The meaning of hyperplane in an n-dimensional space is accompanied by the maximal edge classifier. The hyperplane has (n- 1) dimensions and a level subspace that does not have to go through the root. Because drawing a hyperplane in a higher dimension is difficult, (n- 1) dimensional level subspace is still used. In p-dimensions, the hyperplane used is as follows:

$$\beta\ 0 + \beta 1X1 + \beta 2X2 + \cdots + \beta p\ Xp = 0$$

where X1, X2,..., and Xp are data points in the p-dimensional sample space and 0, 1, 2,..., and p are hypothetical values.

➢ *Artificial Neural Networks (ANNS):*

The artificial neural network algorithm is inspired by biological neurons and works by mimicking the dendrite, soma, and axon workflow. An artificial neural network's basic architecture is a collection of interconnected neurons organized into three layers: input, hidden, and output. In general, this type of network learns to perform tasks by considering a sufficient number of examples.

➢ *Deep Learning:*

The deep learning technique teaches computer machines to do what human beings instinctively do: learn by doing. Deep learning is the leading technology behind self-driving cars, letting them understand a stop sign or differentiate between a road user and a physical static object. Deep learning allows a computer-based system to perform classification tasks directly from images, text, or sound.

B. Evaluation Measures

➢ *Confusion Matrix:*

A confusion matrix contains summarized data of prediction used to present the model's performance. It is a table that contains both predicted and actual values.

Actual\Predicted	A	B
A	TP	FN
B	FP	TN

Here true positive TP is the outcome that means machine learning model predicts Class A correctly. True negative value TN is the outcome that the model accurately predicts Class B. False positive FP denotes the result that the classification model mispredicts Class B. False negative value FN defines the outcome that the model mispredicts Class A.

➢ *Accuracy:*

Accuracy of a machine learning model gives the percentage of tuples correctly classified by the model.

$$accuracy = \frac{no.of\ true\ positive + no.of\ true\ negative}{(no.of\ true\ positive + no.of\ true\ negative + no.of\ false\ positive + no.of\ false\ negative)}$$

➢ *Sensitivity or Recall:*

Sensitivity is simply the true positive rate. It can also be defined as the proportion of correctly classified positive tuples.

$$Sensitivity\ or\ Recall = \frac{no.of\ true\ positive}{(no.of\ true\ positive + no.of\ false\ negative)}$$

➢ *Specificity:*

Specificity is defined as negative rate that is the fraction of correctly classified negative tuples.

$$Specificity = \frac{no.of\ true\ negative}{(no.of\ false\ positive + no.of\ true\ negative)}$$

➢ *Precision:*

It calculates the proportion of true positive versus all positive outcomes.

$$Precision = \frac{no.of\ true\ positive}{(no.of\ false\ positive + no.of\ true\ positive)}$$

➢ *F1 Score:*

F1 score is computed by taking harmonic mean of precision and recall.

$$F1\ score = \frac{2 * (Precision * Recall)}{(Precision + Recall)}$$

➢ *Receiver Operating Curve:*

It is a graphical aid to compare the performance of two or more machine learning models. It plots true positive rate against false positive rate.

IV. LITERATURE REVIEW

This section presents existing research works done in prediction of breast cancer, thyroid, and PCOS.

A. Machine learning techniques used in prediction of breast cancer:

Gopal et al. [1] explored the machine learning techniques: Naïve Bayes, linear regression, logistic regression, SVM, decision tree, random forest, and neural network in predicting breast cancer with IOT devices. The main objective of the research work done in [2] is to predict and diagnosis breast cancer, using machine-learning algorithms: Logistic Regression ,Support Vector Machine (SVM), Random Forest ,K-Nearest Neighbors (KNN) and Decision tree (C4.5), and find out the most effective results with respect to confusion matrix, accuracy and precision. In this study [3], the authors described a Computer-Aided Diagnosis (CAD) technique with deep learning to classify breast cancer patients into 3 classes (cancer, no cancer, and non-cancerous). Islam et al. [4] presented a model that predicts breast cancer disease using machine learning techniques such as SVM and K-Nearest Neighbors. The SVM-based model works with an accuracy value of 99.68% in the training phase. The proposed model is beneficial for the patient and medical team. Yue et al. [5] reviewed machine learning techniques such as artificial neural network, support vector machine, k-nearest neighbor, decision tree and their applications in breast cancer diagnosis and prediction. Gayathri et al. [6] summarized the works done on breast cancer diagnosis using different machine learning methods to enhance the accuracy of detecting breast cancer. The study helps us understand the research works proposed to diagnose breast cancer.

Islam et al. [7], compared the use of K-nearest neighbours, support vector machine (SVM), logistic regression, artificial neural networks (ANNs), and random forests supervised machine learning techniques in breast cancer prediction. The sensitivity, accuracy, precision, F1 score, negative predictive value, specificity, Matthews Correlation Coefficient, false-negative rate, and false-positive rate are used to assess the performance of machine learning techniques. Kumar et al. [8] explored and compared the performance of J48, Naïve Bayes, and SVM classification algorithms and combination of these algorithms using voting classifier technique for the prediction of breast cancer. Voting is one of the ensemble approaches where we can combine multiple models for the better classification. In [9], CART classifier has been applied for the classification of various breast cancer datasets. The bagging technique is used to assess the classifier's performance concerning accuracy and classification time. Harinishree et al. [10] examined different datasets and machine learning models and accuracy of various types of machine leaning models used to predict breast cancer. Yarabarla et al. [11] discussed the use of KNN and random forest machine learning models to predict whether the patient is having breast cancer or not.

B. Machine learning techniques used in prediction of thyroid:

Raisinghani et al. [12] discussed the most recent machine learning algorithms: Support vector machine, Decision trees, Logistic regression, Random forest, and Multilayer Feed Forward Neural Network used in the detection and prediction of thyroid detection. The proposed scheme is used to predict thyroid disease among patients based on various thyroid symptoms and reports. Duggal et al. [13],discussed several feature extraction and classification techniques for thyroid disease diagnosis which are related to machine learning techniques problems. The methods for selecting features introduced are Tree Based Feature Selection, Recursive Feature Elimination, and, Univariate Selection. Three classification techniques were used: Support Vector Machines, Random Forest, and Naive Bayes. Abbad et al. [14], examined various classifiers such as logistic regression, Naive Bayes, decision tree, K-nearest neighbour (KNN), and support vector machine that are implemented with or without feature selection techniques. Tyagi et al. [15], introduced various machine learning techniques such as K-NN, support vector machine, decision tree and analysis for thyroid preventative measure to predict a patient's risk of developing thyroid disease. Anari et al. [16] explained the goals of deep learning in thyroid cancer imaging and performed a review of literature on its limitations, potential, and current application in this field. They provided an overview of recent developments in thyroid cancer analysis using deep learning methods and mentioned different challenges and practical issues that may limit the growth and integration of deep learning into medical workflow.

C. Machine learning techniques used in prediction of PCOS

Denny et al. [17] developed a method for the initial finding and estimate of PCOS focused on an optimal and slight but promising set of medical and metabolic limitations that act as an early marker for the PCOS disease. It primarily focuses on creating algorithms that can access the datasets offered and utilize the data for the network's learning purposes. Bharti et al. [18] explained its model-based detection for polycystic ovary syndrome in women patients. Machine learning algorithms such as logistic regression, gradient boosting, random forest, and hybrid random forest and logistic regression (RFLR) are used to apply to a dataset that is openly accessible in the Kaggle source. Chauhan et al. [19] used Machine Learning techniques such as Decision Tree Classifier, K-Nearest Neighbor(KNN), Logistic Regression(LR), Naive Bayes, Support Vector Machine(SVM) to create an application for initial PCOS guess. The Decision Tree Classifier was discovered to be the most adequate model for PCOS forecast based on the accuracy and confusion matrix. An application was created to assist users in predicting PCOS at an early stage. Vikas et al. [20] compared the accuracies and other concert methods of the previously mentioned data mining techniques such as Neural Networks, Decision Tree and Naïve Bayes for predicting whether a patient is likely to have PCOS or not. In [21], seven classifiers were used to apply various machine algorithms. The results show that the Linear Discriminant classifier performs best in terms of accuracy, while the KNN classifier performs best in terms of sensitivity. Table 2 compares research works done in predicting breast cancer, thyroid, and PCOS using different machine learning techniques.

Table 2: Comparative Analysis of Existing Works

Reference	Disease Name	Machine Learning Techniques	Dataset	Evaluation Measures
[1]	Breast Cancer	Logistic regression, Naïve Bayes, SVM, linear regression, random forest, decision tree, neural network	Wisconsin Prognostic Breast Cancer Chemotherapy (WPBCC), Wisconsin Diagnostic Breast Cancer (WDBC)	Precision, Recall, F-Measure and accuracy
[2]	Breast Cancer	Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest, Logistic Regression, and Decision tree (C4.5)	Wisconsin Breast Cancer Diagnostic	Accuracy
[3]	Breast Cancer	Support Vector Machine (SVM), Deep Learning- Convolutional Neural Network (CNN), and Random Forest	Dataset from Kaggle repository	Accuracy
[4]	Breast Cancer	K-Nearest Neighbors and Support Vector Machine	Wisconsin breast cancer diagnosis data	Accuracy, False Discovery Rate Specificity, Sensitivity,
[5]	Breast Cancer	Artificial neural network, SVM, decision trees, and K-Nearest Neighbors.	Wisconsin breast cancer database (WBCD)	Accuracy
[6]	Breast Cancer	Reinforcement learning, Supervised learning, Unsupervised learning, Semi-supervised learning, Transduction, Learning to learn.	Wisconsin Breast Cancer Diagnostic	Accuracy

[7]	Breast Cancer	Random forests, Support vector machine (SVM), logistic regression, artificial neural networks (ANNs) and K-nearest neighbors	Wisconsin Breast Cancer	Accuracy, Precision,F1score,
[8]	Breast Cancer	J48, Naïve Bayes, and SVM	Wisconsin University database	Accuracy, Sensitivity, Specificity and misclassification rate
[9]	Breast Cancer	Decision tree, bagging, CART	Wisconsin University database	Accuracy
[10]	Breast Cancer	Supervised Learning, Neural Networks, Naïve Bayes, Decision Tree	Wisconsin Breast Cancer Dataset, MIAS Dataset	Accuracy
[11]	Breast Cancer	KNN, Random forest	WBCD data set from UCI machine learning repository	Recall,F1-score,support
[12]	Thyroid	Decision trees, Random forest, Support vector machine, Logistic regression and Multilayer Feed Forward Neural Network	UCI machine learning repository	Accuracy, Precision, F1 score, Recall
[13]	Thyroid	Support vector machines, Naive Bayes, and Random Forest	UCI Machine Learning Repository	Accuracy
[14]	Thyroid	L1 and L2 Norm-Based Model Feature Selection, KNN, SVM, Naive Bayes, Decision Tree and Logistic regression	DHQ Teaching Hospital, Dera Ghazi Khan, Pakistan	Precision, Accuracy, and Receiver Operating Curve
[15]	Thyroid	Support vector machine (SVM), K-NN, Decision Trees	UCI machine learning repository	Accuracy, Mean absolute error
[16]	Thyroid	Deep learning, ANN, CNN, GAN,	-	Accuracy, Sensitivity, Specificity
[17]	PCOS	Logistic Regression, Naïve Bayes Classifier, K-NN, Classification and Regression Trees (CART), Support Vector Machine (SVM), Random Forest Classifier,	Sampled datasets of 541 samples	Accuracy, Sensitivity, Specificity, Precision, F1 score
[18]	PCOS	Gradient boosting, random forest, logistic regression, and hybrid random forest and logistic regression (RFLR)	Dataset from Kaggle repository	Accuracy
[19]	PCOS	Logistic Regression, K-Nearest Neighbor, Naive Bayes, Decision Tree, Support Vector Machine,	Sampled datasets of 267 samples	Accuracy, Precision, Recall, Specificity, F1 score
[20]	PCOS	Naïve Bayes, Decision Tree and Neural Networks	Survey conducted based on the lifestyle of the women.	Sensitivity, Specificity, Precision, Accuracy
[21]	PCOS	Logistic Regression, Linear Regression, KNN, Naïve Bayes, SVM, Decision Tree,	Kaggle dataset name Polycystic Ovary Syndrome.	Accuracy, Precision, Sensitivity, Specificity

V. CONCLUSION

This survey presents an overview of applications of different machine learning techniques in predicting breast cancer, thyroid and PCOS. Existing research works incorporate various machine learning algorithms such as Support vector machine (SVM), K-nearest neighbors, decision trees, random forests, artificial neural networks (ANNs), logistic regression, Naive Bayes techniques, and deep learning. These machine learning models are evaluated using different performance measures such as accuracy, sensitivity, specificity, precision, recall, f1 score, and receiver operating curve. The study presented in the survey concludes that machine learning techniques provide automatic disease detection with high performance.

REFERENCES

[1]. Gopal, V. N., Al-Turjman, F., Kumar, R., Anand, L., & Rajesh, M.(2021).Feature selection and classification in breast cancer prediction using IoT and machine learning. *Measurement,178*, 109442.

[2]. Naji, M. A., El Filali, S., Aarika, K., Benlahmar, E. H., Abdelouhahid, R. A., & Debauche, O. (2021). Machine learning algorithms for breast cancer prediction and diagnosis. *Procedia Computer Science*, *191*, 487-492.
[3]. Allugunti, V. R. (2022). Breast cancer detection based on thermographic images using machine learning and deep learning algorithms. *International Journal of Engineering in Computer Science*, *4*(1), 49-56.
[4]. Islam, M. M., Iqbal, H., Haque, M. R., & Hasan, M. K. (2017, December). Prediction of breast cancer using support vector machine and K-Nearest neighbors. In *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)* (pp. 226-229). IEEE.
[5]. Yue, W., Wang, Z., Chen, H., Payne, A., & Liu, X. (2018). Machine learning with applications in breast cancer diagnosis and prognosis. *Designs*, *2*(2), 13.
[6]. Gayathri, B. M., Sumathi, C. P., & Santhanam, T. (2013). Breast cancer diagnosis using machine learning algorithms–a survey.
[7]. Islam, M., Haque, M., Iqbal, H., Hasan, M., Hasan, M., & Kabir, M. N. (2020). Breast cancer prediction: a comparative study using machine learning techniques. *SN Computer Science*, *1*(5), 1-14.
[8]. Kumar, U. K., Nikhil, M. S., & Sumangali, K. (2017, August). Prediction of breast cancer using voting classifier technique. In *2017 IEEE international conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM)* (pp. 108-114). IEEE.
[9]. Lavanya, D., & Rani, K. U. (2012). Ensemble decision tree classifier for breast cancer data. *International Journal of Information Technology Convergence and Services*, *2*(1), 17-24.
[10]. Harinishree, M. S., Aditya, C. R., & Sachin, D. N. (2021, April). Detection of Breast Cancer using Machine Learning Algorithms–A Survey. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1598-1601). IEEE.
[11]. Yarabarla, M. S., Ravi, L. K., & Sivasangari, A. (2019, April). Breast cancer prediction via machine learning. In *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 121-124). IEEE.
[12]. Raisinghani, S., Shamdasani, R., Motwani, M., Bahreja, A., & Raghavan Nair Lalitha, P. (2019, April). Thyroid prediction using machine learning techniques. In *International Conference on Advances in Computing and Data Sciences* (pp. 140-150). Springer, Singapore.
[13]. Duggal, P., & Shukla, S. (2020, January). Prediction of thyroid disorders using advanced machine learning techniques. In *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 670-675). IEEE.
[14]. Abbad Ur Rehman, H., Lin, C. Y., Mushtaq, Z., & Su, S. F. (2021). Performance analysis of machine learning algorithms for thyroid disease. *Arabian Journal for Science and Engineering*, *46*(10), 9437-9449.
[15]. Tyagi, A., Mehra, R., & Saxena, A. (2018, December). Interactive thyroid disease prediction system using machine learning technique. In *2018 Fifth international conference on parallel, distributed and grid computing (PDGC)* (pp. 689-693). IEEE.
[16]. Anari, S., Tataei Sarshar, N., Mahjoori, N., Dorosti, S., & Rezaie, A. (2022). Review of Deep Learning Approaches for Thyroid Cancer Diagnosis. *Mathematical Problems in Engineering*, *2022*.
[17]. Denny, A., Raj, A., Ashok, A., Ram, C. M., & George, R. (2019, October). i-hope: Detection and prediction system for polycystic ovary syndrome (pcos) using machine learning techniques. In *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)* (pp. 673-678). IEEE.
[18]. Bharati, S., Podder, P., & Mondal, M. R. H. (2020, June). Diagnosis of polycystic ovary syndrome using machine learning algorithms. In *2020 IEEE Region 10 Symposium (TENSYMP)* (pp. 1486-1489). IEEE.
[19]. Chauhan, P., Patil, P., Rane, N., Raundale, P., & Kanakia, H. (2021, June). Comparative analysis of machine learning algorithms for prediction of pcos. In *2021 International Conference on Communication information and Computing Technology (ICCICT)* (pp. 1-7). IEEE.
[20]. Vikas, B., Anuhya, B. S., Chilla, M., & Sarangi, S. (2018). A critical study of Polycystic Ovarian Syndrome (PCOS) classification techniques. *IJCEM Int J Comput Eng Manage*, *21*(4).
[21]. Hdaib, D., Almajali, N., Alquran, H., Mustafa, W. A., Al-Azzawi, W., & Alkhayyat, A. (2022, May). Detection of Polycystic Ovary Syndrome (PCOS) Using Machine Learning Algorithms. In *2022 5th International Conference on Engineering Technology and its Applications (IICETA)* (pp. 532-536). IEEE.
[22]. Jalal, S., Yadav, D. K., & Negi, C. S. (2019). Web service discovery with incorporation of web services clustering. *International Journal of Computers and Applications*, 1-12.(Online)
[23]. Arushi , Jalal, S. , Negi, C. S. (2022, October). Covid-19 Prediction based on Symptoms using Machine Learning, *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT)*, *11*(10).
[24]. Breast Cancer Wisconsin (Original) Dataset, [Online]. https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wiscon sin/breast-cancer-wisconsin.data
[25]. Breast Cancer Wisconsin dataset,[Online]. https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data
[26]. Breast Cancer Dataset, [Online]. http://peipa.essex.ac.uk/info/mias.html
[27]. Thyroid dataset. http://archive.ics.uci.edu/ml/machine-learning-databases/thyroid-disease/

[28]. PCOS dataset. https://github.com/PCOS-Survey/PCOSData/blob/master/PCOS-Data.csv
[29]. PCOS dataset. https://www.kaggle.com/datasets/prasoonkottarathil/polycystic-ovary-syndrome-pcos