A Review on Different Machine Learning Algorithms to Breast Cancer Risk Prediction

RAM BABU BURI
(Assistant Professor)
Computer Science & Engineering
Arya College of Engineering & I.T.
Jaipur
burirambabuapex2009@gmail.com

Dr. VISHAL SHRIVASTAV
(Professor)
Computer Science & Engineering
Arya College of Engineering & I.T.
Jaipur
vishalshrivastava.cs@aryacollege.in

Abstract— Breast cancer is the second most fatal kind of cancer in women, accounting for almost a quarter of all cancer deaths. About 10 percent of women worldwide are diagnosed with breast cancer at some time in their lives, making it one of the most frequent malignancies among women today. However, although the treatment for this cancer is now accessible in practically all first world and some third world countries, the primary problem occurs when the cancer is not appropriately detected at the very beginning of the disease's progression. Several researchers have made significant contributions to early diagnosis, improved prognosis, and better treatment of BC during the previous two decades, resulting in a decrease in the death rate. Machine Learning has shown to be quite useful in this discipline, particularly in the prediction of illnesses such as cancer. So far, classification and data mining approaches have shown to be dependable and successful means of categorizing data. These strategies have been used to forecast and make judgments in a variety of fields, particularly the medical industry. Throughout this study, we examined the current state of the art in BC prediction, which included breast cancer diagnosis, BC risk prediction using several machine learning algorithms, and breast cancer recurrence prediction.

Keywords— Breast Cancer, Risk Predication, Breast Cancer Types, Cancer Stages, Machine Learning, Classification Algorithms.

I. INTRODUCTION

As stated by the WHO (2018), cancer is the second foremost reason for death worldwide. In 2018, it is projected to have been responsible for 9.6 million fatalities. On a global scale, cancer is responsible for around 1 in every 6 fatalities. Cancer is the foremost reason for mortality in low- & middle-income nations, accounting for around 70% of all cancer fatalities. Cancer control funding allocated to developing countries account for just 5 percent of total worldwide funds allocated to cancer control, and other such countries have extremely limited human and material resources accessible to them. Cancer, along with the American Cancer Society, is a category of illnesses defined by uncontrolled development & spread of abnormal cells throughout the body. If the spread is not controlled, it might end in death. It is cancer that affects breast tissue, and it most usually arises from the inner lining of milk ducts or lobules that feed milk to ducts. It is the most prevalent kind of cancer in women. Increasing rates of breast cancer incidence, [1] and death are being reported throughout the globe, and this poses a substantial and increasing concern to the developing world. Breast cancer is on the rise in developing countries, mostly as a result of increased life expectancy as well as changes in lifestyle, including such women fewer children, as well as hormonal treatment, including such post-menopausal hormone treatment.

Breast cancer (BC) is second on the list of leading causes of mortality among women (after lung cancer). In the US it is projected that 246,660 novel instances of invasive BC would be identified in women throughout 2016, with an estimated death toll of 40,450 women. Breast cancer, which develops in the breast tissue, affects women more often than males. Proliferating cells becomes unrestrained as cancer develops. Tumours formed by breast cancer cells are often visible on an x-ray or felt as a bump on the breast surface. BC may spread if cancer cells enter the bloodstream or lymphatic system & are transported to other regions of the body by the immune system. Changes and mutations in DNA are among the factors that contribute to the development of breast cancer. Invasive DCIS are two of frequent kinds of breast cancer, but there are many more. Many other types of cancer are less prevalent, such as phyllodes tumour and angiosarcoma. There are a plethora of algorithms available for predicting breast cancer outcomes.

Classification of benign and malignant cancers using ML approaches[2] When diagnosed early, Breast Cancer patients are more educated about when and how to begin clinical therapy, increasing their chances of survival. The classification of benign tumours may assist patients in avoiding unnecessary therapy. To correctly diagnose breast cancer and classify individuals into malignant and benign categories, research is needed. Machine learning is widely recognised as the tool for the prediction of breast cancer because of its advances in the discovery of essential characteristics from complicated datasets. It is possible to use data mining methods in the medical profession to better anticipate results, reduce costs of drugs, improve people's health, and even save lives.

Classification methods of machine learning may be used to better categorise benign and malignant tumours. Numerous datasets on breast cancer have been the subject of a great deal of ML & DM research. There is a lot of evidence that classification systems may accurately anticipate the kind of tumour that will be found. It has become more common in recent years for useful information and expertise to be concealed in data from a wide range of industries. It is always a monumental undertaking to process and extracts useful information from such massive amounts of data. This may be accomplished with the use of data mining. Current ages have seen significant development in the use of data mining tools in breast cancer research. Breast cancer data categorization may help forecast the prognosis of certain illnesses or find the genetic behaviour of tumours. Breast cancer patterns may be predicted and classified using a variety of methods.[3].

II. BREAST CANCER: AN OVERVIEW

Breast cancer [4] is the most fatal and diverse disease that affects women all over the globe today, and it is responsible for the deaths of an immense number of people. Breast cancer is a cancerous tumour that develops when cells in the breast tissue multiply and divide unchecked, leading to uncontrolled cell proliferation as well as expansion. When it comes to all tissues and organs in the human body, the cell is a fundamental unit. Normal cells in the breast expand as well as a divide to generate new cells when they are required. As normal cells become old or get injured, these dice, as well as new cells take their place but now and then, the system continues haywire and indeed the old or damaged cells don't die as they should. As a cause of death, it is the second - most common sickness among women. The prediction concerning breast cancer is being made using a variety of machine learning and data mining methods. Outcome The most acceptable and suitable algorithms for the prediction of breast cancer is a significant problem. This is extremely difficult work. Breast cancer develops as a result of malignant tumours, which form when a cell's development becomes uncontrollably large. A large number of fatty & fibrous tissues in the breast begin to grow abnormally, which eventually leads to the development of BC. The cancer cells spread all through tumours, causing various stages of cancer to manifest themselves.

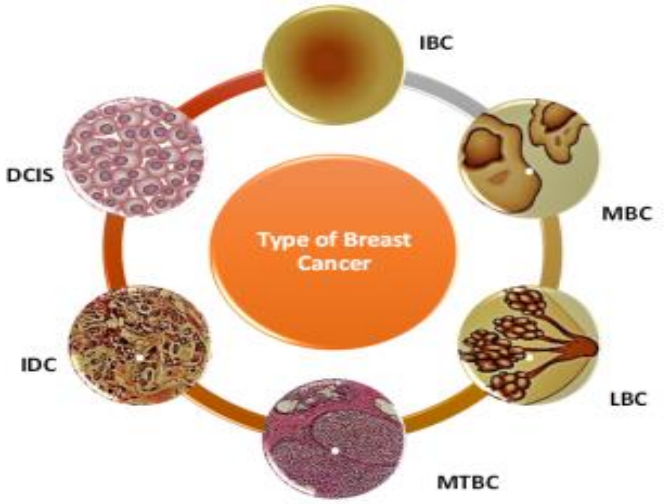

Figure 1: Demonstration of major types of BC

A. Breast Cancer Stages

BC is separated into the above-given categories that reveal the intensity of breast cancer [5][6].

Stage 0 means there could be a risk of cancer. In Stage 1, breast cancer cells begin to form, as well as the tumour size is around 2 cm or less. In Stage 2, the normal tumour size is somewhere between 2 to 5 cm. The tumours are developing inside the breast tumour or migrating to adjacent lymph nodes. Stage 2 is classified into Stage 2A as well as Stage 2B based on the size of the tumour but whether breast cancer has progressed to lymph nodes. Those three levels are classified as early stages [7]. Breast cancers Stages 1, as well as 2, are commonly treated with breast-conserving surgery as well as radiation treatment. Dependent on the kind of cancer cells as well as other health conditions, hormone treatment can be indicated [8].

In Stage 3, breast cancer is characterised as locally progressed when tumours have expanded to nearby lymph nodes including muscles but have not travelled to distant organs. Stage 3 is further split into sub-stages 3A, 3B, and 3C based on the size of the tumour and if BC has progressed to lymph nodes or adjacent tissues. The therapies for these phases may include mastectomy plus radiation for local treatment, including hormone therapy or chemo for treatment response. Most of the individuals diagnosed with Terminal 3 will react effectively with a combination of the two or more therapies. In [8], they stated that “stage 3 breast cancer often needs induction chemotherapy to down-size the tumour to permit breast-conserving surgery. Inflammatory breast cancer, although deemed stage 3, is aggressive as well as needs induction treatment accompanied by the mastectomy, but instead of breast-conserving operations, but also auxiliary lymph node removal and chest wall radiotherapy”

Stage 4 breast cancer signifies that cancer is spreading to other regions of the body, like the brain, bones, lungs, as well

as liver. “Prognosis is poor in women with recurring or metastatic (stage IV) breast cancer, therefore treatment choices must balance advantages in an increase in life expectancy and decreased discomfort against hazards from therapy”[8][6].

B. Common Types of Breast Cancer [6]

- **Ductal Carcinoma in Situ (DCIS):** Stage 0 is either not recognised to be cancer or DCIS, which is a non-invasive form of the disease. The cancer is still confined to the milk duct and has never really spread further. Hormone treatment may, on the other hand, be recommended to patients to prevent cancer cells from spreading into the surrounding tissue.
- **Invasive Ductal Carcinoma (IDC):** IDC is a cancer that started developing in the duct and infiltrating the damaged tissue. IDC accounts for 80 per cent of all affected individuals. The treatment choices for IDC vary, based on cancer stage, medical information, coupled with a medical assessment, and other tests.
- **Inflammatory Breast Cancer (IBC):** IBC is uncommon yet aggressive as well as fast-growing. In this situation, cancer is spreading to the skin of the breast, stops the lymph arteries in the skin causing the breast to appear red and swollen also feels warm. IBC might be in phases 3B as well as 3C. With rigorous therapy, the rate of survival for IBC patients has increased dramatically in recent years.
- **Recurrent Breast Cancer (RBC):** RBC seems to be cancer that has returned after being treated.
- **Metastatic Breast Cancer (MBC):** MBC is also defined as Stage 4 breast cancer. Cancer is spreading to other regions of the body like the lungs, liver, bones or brain.

III. BREAST CANCER RISK PREDICTION (BCRP)

BC is a tumour that grows in breast tissue. This kind of cancer is prevalent in women across the world and is one of the most common reasons for death among women. Breast cancer, a specific kind of cancer that predominantly affects women, is produced by two fundamental elements: changeable or non-modifiable. Environmental and psychological factors are examples of controllable variables. Gender and medical history in the family are examples of non-modifiable factors. Since early intervention technologies for the presence of breast cancer are still insufficient in the accurate prediction of sickness, 1 in 28 Indian women is at risk of developing the disease, according to a survey. Furthermore, survival risks are increased due to a lack of information, safeguards, and treatment options. The early detection of these disorders may aid in the development of effective treatments for breast cancer [9].

Suggested [10] use a combination of methods to detect breast cancer in its earliest stages. UCI BC's database helped them to evaluate their model. They used the J48 approach to develop rules, which they subsequently submitted to customizable autonomy coordinators who were in charge of BC's forecasting system. Implemented [11] Breast cancer risk assessment using supervised machine learning. They used DT, Naive Bayes, Bayesian Networks, and SVM as classification approaches. Additionally, they used ten-fold cross-validation to evaluate the performance of the methods. Aside from that, they used ROC plots to show the algorithm's effectiveness visually. Binary classifiers are evaluated based on their selectivity and sensitivity, two statistical measures. The sensitivity metric shows the percentage of correctly classified positive events, while the precision metric shows the percentage of correctly recognised negative events. In order to improve classification performance, a feature selection approach is used throughout the classification process. More precise results are obtained by using the LibSVM implementation of the SVM. A [12] BCRP algorithms may be found in the PubMed dataset. A multi-method approach was needed to get the best possible risk prediction abilities, they concluded.

Table 1: Breast cancer risk prediction summary [13]

Author and Year	Technique Used	Advantages	Limitations	Remarks
Al-Imran Ahmed, 2014 [10]	A hybrid method with J48 algorithms as well as flexible autonomy coordinate.	BCRP.	The function of the user as well as autonomously agent is to be researched further.	UCI breast cancer dataset.
Aniket Bochare, Aryya Gangopadhyay, 2014 [11]	DT, Naive Bayes, Bayesian Networks, and SVM	BCRP.	Needs to evaluate ensemble methods	NHS dataset was used.
Anothaisinatawee et al., 2012 [12]	Performances of the proposed comparison review.	Diverse models for BCRP.	Additional study is necessary to enhance models.	A good review of risk prediction models.

IV. ML ALGORITHMS FOR BREAST CANCER PREDICTION

ML is the field of science that allows systems to learn without having to be specifically designed. Using ML, we shouldn't have to supply detailed instructions to Computer for responding to specific unique scenarios. ML is nothing more than the systematic systematization of the acquisition of information and skill via the use of mathematical and non-mathematical approaches. We ought to offer training to the computers to identify real-time answers for unique situations. The chess game is a good illustration whereby machine learning is being utilised to play chess. The programming enables the computer to learn as well as optimise itself over patterned behaviour. Continual efforts, expansion, gradual refinement and advancement in ML has allowed the effective construction of hundreds of AI expert systems which are today being utilised regularly in business and industries. The data can be more efficiently handled with help of ML. It is applied in cases where the exact information or pattern from the data is impossible to be interpreted. The requirement of ML is increasing day by day along with the rise of abundant and ever-growing datasets. Machine Learning is highly utilized to extract the requisite information by major industries such as from medicine to the military. Learning through existing data is the foremost tenacity of ML [14][15].

A. Type of Machine Learning

There are four types of ML methods that are described below [16][15]:

a) Supervised Learning

Supervised learning involves training information setwithlabelled information, or information with a defined output value (e.g. rural/not agricultural or home price). Classification, as well as regression issues, are tackled by supervised learning. Supervised learning methods are of two types: classification and regression. Classification algorithm predicts discrete value. It is used to categorize input data as a part of a particular group. Algorithms such as RF, BN and DT, SVM, ANN and KNN comes under Supervised Learning [14].

b) Unsupervised Learning

Unsupervised learning methods don't use a training set as well as to detect patterns or organization in the data by themself. Clustering difficulties may be tackled using an unsupervised technique. This model is opposite to Supervised Learning. There is no need to supervise a model. It works on unlabeled data sets and it performs complex processing to learn from data and form the groups or cluster-based on data having similar properties. Several algorithms such as Kmeans, hierarchical clustering come under Unsupervised Learning [14].

c) Semi-Supervised Learning

Semi-supervised learning may be regarded as a combination of the above-mentioned unsupervised with supervised approaches since it acts on both labelled and unlabelled. In the actual world, labelled data might be scarce in various scenarios, whereas unlabeled data are plentiful, whereas semi-supervised learning is useful. The ultimate aim of a semi-supervised learning model is to deliver an improved output for forecasts than that provided utilising labelled data alone without a model.

d) Reinforcement Learning

In this learning, there is no predefined data. Here an algorithm interacts with a dynamic environment wherein it accomplishes a certain goal. As it proceeds in solving the problem, feedback is provided to raise the performance. This model is not dependent on supervised or unsupervised learning, but a combination of both. The quantity of utilizations of these algorithms in the area of AI is developing quickly [14].

B. Machine Learning Algorithms

ML is an automated learning approach, the techniques are built to learn from the previous dataset, we input a huge volume of information, ML model evaluates that data & even on the foundation of that train model we can predict the results about future. Regarding breast cancer forecasts, prominent ML techniques are as follow [4]:

a) Artificial Neural Network (ANN)

ANN is a typical method for data extraction and processing. Input, hidden layer, as well as output layers, make up a neural network. The pattern that is too complicated may be extracted using this method. A procedural programming, distributed memory, collaborative resolution, as well as network architecture-based algorithm is used. [17].

b) Logistics Regression (LR)

One of the primary algorithms of ML is Regression which acts as a base for different ML algorithms. It is a linear methods that is applied to resolve classification problems & to calculate the probability of an event. Outcome of the problem is predicted as value 0 or 1 on the basis of threshold added. The primary goal of this method is to accurately predict the category of outcome.

c) K-Nearest Neighbor (KNN)

As the term suggests, K-NN is a kind of supervised machine learning algorithm. It's a great option for activities that need a lot of categorising. KNN measures the distance between the class labels of the closest training data with even a current test data according to the requirement of K value in order to

forecast the target label of a fresh test data. In order to identify the present test data class label, the K value is calculated by determining the number of extremely close data points in the dataset. KNN commonly employs a K variable value between 0 and 1o to define the distance between training data points.

d) Decision Tree (DT)

The classification and regression models serve as the foundation for the decision tree. Each subset of the data set is partitioned into a smaller number of subsets. With a limited collection of data, it is possible to create predictions with the maximum degree of accuracy. CART, C4.5, C5.0, & the conditional tree are all types of DT methods. [18].

e) Naive Bayes Algorithm (NB)

This model assumes that a large training dataset is available. The Bayesian strategy is used to calculate the probability. It is the most accurate method for calculating the probability of noisy data input. For comparing training datasets, this classifier uses an analogue method. [19].

f) Support Vector Machine (SVM)

Support Vector Machine is another well-known supervised learning model for regression and classification problems. This technique takes input data points in space and outputs a hyperplane or a line that categorize data into different classes. There can be several hyperplanes that can classify the input data. The idea is to select the best hyperplane that represents the biggest margin between the two classes.

g) Random Forest (RF)

Using supervised learning, the Random Forest technique may be utilised to tackle classification and regression issues. It's a machine learning building component that's used to predict future data based on past datasets. Random forests, also called random decision forests, consist of a huge number of trees that learn to classify and predict using ensemble techniques. Bagging as well as feature randomization are the characteristics it employs to generate such trees. The random forest has a benefit over the decision tree because is that it does not significantly overfit the data.

h) Extreme gradient boosting (XGBoost)

For both Gradient Boosting and or the Random Forests algorithms discussed above, the final model is generated by building an ensemble of individual models, often decision trees. As with neural networks, which employ gradient descent to optimise weights, the gradient is employed here to minimise the loss function. Gradient boosting using more precise estimates is known as XGBoost (or "Extreme Gradient Boosting" in short). Calculating the gradient of the loss function at the second-order level, eliminates over-fitting and enhances both the generalizability as well as the performances of the model. XGBoost is lightning-fast in its interpretation and can easily handle big datasets.

V. LITERATURE REVIEW

This section gives a literature review of the breast cancer risk prediction utilising machine learning approaches. They have been many additional studies on breast cancer prediction using various algorithms. Some of the relevant efforts previously taken on breast cancer detection by researchers utilising various methodologies are mentioned. According to the findings of the offered literature assessment, it is comparable to creating a precedent among the currently available deep learning methodologies. This section contains information regarding the research work that has previously been completed that is linked to the current study discussed below:

In this research, [20] are reported by applying seven multiple ML approaches for the categorization of BC. To acquire good outcomes, a processing procedure was performed. As a consequence, when comparing with various studies in literature, it was discovered that the level of effectiveness of several approaches improved. In trial findings, BayesNet was shown to be a classification technique with a prediction performance of 97.13 per cent.

At the moment, [21] WBCD has been utilised which has been taken from UCI. The objective is to study the dataset and assess the effectiveness of several ML techniques for detecting BC. Furthermore, SVM, LR, KNN, DT, NB and RFC have been built for categorising tumours into normal and abnormal. To select the best accurate method, the precision of each is calculated in accordance. Classifiers such as Random Forest as well as SVM surpass the others, with an accuracy rate of 96.5 percent, according to the research.

The studies [22] have introduced the use of effective ML techniques to construct models for a forecast which would identify BC happening rate. Breast cancer data consisting of 567 rows of 30 distinct cancer diagnostic qualities, including benign as well as malignant data, was utilised by the authors in their study. A forecast has been done to compute the proportion of persons afflicted with squamous cell carcinoma and visual findings have been shown as outcomes. A non-biased approximation is provided by the authors using supervised learning methods including KNN, NB as well as DT models as well as support vector machines. The findings revealed KNN is the main determinant with 91.6 percent accuracy.

This study [23] generated many machine learning algorithms for breast cancer categorisation utilising RF, SVM, KNN, GNB, Perceptron as well as LR. Three frequently utilised test data sets were employed; WBC Original, WDBC as well as WPBC. The findings reveal that datasets impact the

effectiveness of the training classifiers. Furthermore, the machine learning classifiers have varying characteristics with a particular breast cancer dataset

In this study, [24] six ML classification techniques, particularly regarding LR, kNN, DT, SVM, NB, and RF are executed to evaluate the effectiveness as well as the prognostication power of the prototype. The major goal of this study is to compare various algorithms achievements using the WBCD. There are several important roles in affecting taken into account, including precision, recall, f-1 score, as well as specificity. A support vector machine with just an error rate of 7.7 percent and a precision of 97.07 percent, compared to an NB model with a precision of 96%, exhibits the best results in our research. All these studies were done using SciKit.

In this study work, [25] breast cancer dataset was examined to predict breast cancer employing prominent two ensemble ML approaches. RF, as well as XGBoost, are being utilized for prognostic breast cancer. A total of 275 instances with 12 features were exploited for this inquiry. Employing the RF approach 74.73 per cent accuracy or 73.63 % applying XGBoost have been obtained in this research.

In this study, [26] 357 tumours were found to be malignant, whereas 212 were found to be benign, according to the University of Wisconsin's feature methodologies and assessments in breast mass pictures. The data was split into training & testing sets, by training runs for six different supervised ML techniques accepted out on each set (KNNs algorithm, RF algorithm, DT, NB algorithm, SVM and LR). The dataset was also subjected to principal components analysis and linear discriminant analysis, which performed the same functions. In this research, the performance values for all experimental models were raised after performing linear discriminant analysis and then a success rate of 96.49 percent was attained with LR. The purpose of this study was to pick an algorithm as well as collect data that would be useful for future research by building on the findings from this one.

VI. CONCLUSION

Breast cancer is currently the most often diagnosed cancer in women, & indeed the proportion of women diagnosed with it is steadily growing. If breast cancer is detected as well as treated at an early stage, there is a strong probability of a successful treatment outcome. It is one of the most difficult challenges to discover a suitable and acceptable technique for the prediction of breast cancer. This is extremely difficult work. Breast cancer develops as a result of malignant tumours, which form when a cell's development becomes uncontrollably large. Different ML and DM techniques are being utilised to make predictions about the possibility of breast cancer occurring in the future. The primary goal of this review is to summarise all of the previous studies of ML techniques that have been applied for BC prediction. Additionally, this paper also provides all of the essential knowledge for beginners who are interested in analysing machine learning algorithms. For this reason, some researchers have developed machine-learning approaches for forecasting the development of cancer cells using medical imaging modalities, which they have tested in the laboratory. There are just a few review studies on breast cancer diagnosis that have been published so far, and these studies synthesise some of the current research. The emphasis of this review is on the advancement of machine learning in the identification of breast cancer. Nonetheless, the breast cancer issue continues to be alarming, and more study is needed in the areas of improved diagnosis and prediction, as well as treatment options.

REFERENCES

[1] R. Rawal, “BREAST CANCER PREDICTION USING MACHINE LEARNING,” *J. Emerg. Technol. Innov. Res.*, vol. 7, no. 5, 2020.

[2] C. H. Shravya, K. Pravalika, and S. Subhani, “Prediction of breast cancer using supervised machine learning techniques,” *Int. J. Innov. Technol. Explore. Eng.*, 2019.

[3] I. O. B. and A. A. A., “a Survey on the Data Mining Techniques Used for Breast Cancer Predictions,” *Int. J. Res. Eng. Sci.*, vol. 3, no. 5, pp. 30–41, 2019, doi: 10.26808/rs.re.v3i5.04.

[4] N. Fatima, L. Liu, S. Hong, and H. Ahmed, “Prediction of Breast Cancer, Comparative Review of Machine Learning Techniques, and Their Analysis,” *IEEE Access*. 2020, doi: 10.1109/ACCESS.2020.3016715.

[5] N. Subhani, Y. Li, A. Ngom, and L. Rueda, “Alignment versus variation methods for clustering microarray time-series data,” 2010, doi: 10.1109/CEC.2010.5586111.

[6] Q. H. Pham, “Machine Learning Approaches for Breast Cancer Survivability Prediction,” 2020.

[7] S. Y. Kang *et al.*, “Prosaposin inhibits tumor metastasis via paracrine and endocrine stimulation of stromal p53 and Tsp-1,” *Proc. Natl. Acad. Sci. U. S. A.*, 2009, doi: 10.1073/pnas.0903120106.

[8] A. Alkhateeb, I. Rezaeian, S. Singireddy, and L. Rueda, “Obtaining biomarkers in cancer progression from outliers of time-series clusters,” 2015, doi: 10.1109/BIBM.2015.7359802.

[9] S. G. Durai, S. Hari Ganesh, and A. Joy Christy, “Prediction of breast cancer through classification algorithms: A survey,” *Int. J. Control Theory Appl.*, 2016.

[10] A. I. Ahmed and M. M. Hasan, “A hybrid approach for decision making to detect breast cancer using data mining and autonomous agent based on human agent teamwork,” 2003, doi: 10.1109/ICCITechn.2014.7073116.

[11] A. Bochare *et al.*, “Integrating domain knowledge in supervised machine learning to assess the risk of breast cancer,” *Int. J. Med. Eng. Inform.*, 2014, doi: 10.1504/IJMEI.2014.060245.

[12] T. Anothaisintawee, “Risk prediction models of breast cancer: a systematic review of model performances,” 2012, doi: https://doi.org/10.1007/s10549-011-1853-z.

[13] R. Aavula, “A Survey on Latest Academic Thinking of Breast Cancer Prognosis 1,” *Int. J. Appl. Eng. Res.*, vol. 13, no. 7, pp. 5207–5215, 2018.

[14] P. R. Maidamwar, M. M. Bartere, and P. P. Lokulwar, “A Survey on Machine Learning Approaches for Developing Intrusion Detection System,” *SSRN Electron. J.*, 2021, doi: 10.2139/ssrn.3833858.

[15] M. Meenakshi, “Machine Learning Algorithms and their Real-life Applications: A Survey,” *SSRN Electron. J.*, no. Icicc, pp. 1–4, 2020, doi: 10.2139/ssrn.3595299.

[16] V. Jain, A. Kulkarni, Y. B. T. Integrated, C. Engineering, and N. Mpstme, “Survey on Various Algorithms of Machine Learning and its Applications,” 2020.

[17] Y. Uzun and G. Tezel, “RULE LEARNING WITH MACHINE LEARNING ALGORITHMS AND ARTIFICIAL NEURAL NETWORKS,” *J. Selcuk Univ. Nat. Appl. Sci.*, 2012.

[18] N. Satyanarayana, C. Ramalingaswamy, and Y. Ramadevi, “Survey of Classification Techniques in Data Mining,” *IJISET -International J. Innov. Sci. Eng. Technol.*, 2014.

[19] A. A. Ibrahim, A. I. Hashad, and N. E. M. Shawky, “A Comparison of Open Source Data Mining Tools for Breast Cancer Classification,” 2017.

[20] B. Erkal and T. E. Ayyıldız, “Using Machine Learning Methods in Early Diagnosis of Breast Cancer,” in *2021 Medical Technologies Congress (TIPTEKNO)*, 2021, pp. 1–3, doi: 10.1109/TIPTEKNO53239.2021.9632975.

[21] S. Ara, A. Das, and A. Dey, “Malignant and Benign Breast Cancer Classification using Machine Learning Algorithms,” 2021, doi: 10.1109/ICAI52203.2021.9445249.

[22] N. Sinha, P. Sharma, and D. Arora, “Prediction model for breast cancer detection using machine learning algorithms,” 2021, doi: 10.1007/978-981-15-6876-3_33.

[23] K. J. Kurugh, M. A. Ahmad, and A. A. Babajo, “THE EFFECT OF DATASETS ON BREAST CANCER DETECTION MODELS,” *FUDMA J. Sci.*, 2021, doi: 10.33003/fjs-2020-0404-487.

[24] M. R. Ahmed, M. A. Ali, J. Roy, S. Ahmed, and N. Ahmed, “Breast Cancer Risk Prediction based on Six Machine Learning Algorithms,” 2020, doi: 10.1109/CSDE50874.2020.9411572.

[25] S. Kabiraj *et al.*, “Breast Cancer Risk Prediction using XGBoost and Random Forest Algorithm,” 2020, doi: 10.1109/ICCCNT49239.2020.9225451.

[26] S. Kaya and M. Yaganoglu, “An Example of Performance Comparison of Supervised Machine Learning Algorithms before and after PCA and LDA Application: Breast Cancer Detection,” 2020, doi: 10.1109/ASYU50717.2020.9259883.