Rev. Roum. Sci. Techn.– Électrotechn. et Énerg.
Vol. **66**, *3*, pp. 201–206, Bucarest, 2021

CATEGORY BOOSTING MACHINE LEARNING ALGORITHM FOR BREAST CANCER PREDICTION

HARSHIT GUPTA[1], PRITAM KUMAR[1], SHUBHAM SAURABH[1], AND SUNIL KUMAR MISHRA[1], BHARGAV APPASANI[1], AVADH PATI[2], CRISTIAN RAVARIU[3], AND AVIRENI SRINIVASULU[4]

Key words: Breast cancer, Category boosting, Machine learning, Prediction model.

Cancer is considered the worst of all diseases. It is a category of diseases that enable irregular growth that may enter or spread to certain body areas. These contrast with healthy, not multiplying tumors. There are 100 different cancer forms that impact humans. With the emergence of machine learning (ML), its uses have been identified in many fields, particularly medical research. It also used for cancer detection when a correct dataset is available. This paper suggests a category boosting (CatBoost) ML algorithm for predicting the different stages of breast cancer, facilitating early diagnosis. The proposed CatBoost algorithm is an efficient method to train and test the available data. To show the CatBoost method's efficacy a detailed comparative analysis has been carried out with other prominent ML approaches. It has been established that the CatBoost is accurate compared to the other ML methods.

1. INTRODUCTION

Regardless of the poor cancer survival rate, the early detection of the disease increases the probability of survival through treatments such as chemotherapy. Breast cancer is a common disease in women, and its early detection can be of immense benefit [1,2]. Machine learning (ML) is the analysis of machine models by experience [3]. It is a branch of artificial intelligence (AI) and deep learning superset. This constructs a predictive model based on experimental data called training data and partly executes the analysis on the testing data. Recently, several ML algorithms have been introduced for the diagnosis of breast cancer [4]. A brief survey of these ML methods for breast cancer diagnosis is presented.

In [5], a comparative analysis on breast cancer detection was carried out using a k-fold cross-validation method. A support vector machine (SVM) model was introduced in [6], with semi-supervised learning for breast cancer classification. In [7], the research showed that gene mutation profiles might be successfully used to classify clinically distinguishable subgroups in breast cancer utilizing unsupervised ML techniques. The k-nearest neighbours (KNN), logistic regression (LR), decision tree (DT), SVM, random forest (RF), adaptive boosting (AdaBoost), and gradient boosting machines were applied by [8]. Two different feature selection techniques were implemented wherein the SVM ML algorithm yielded the best results. In [9], the sorting of benign (initial stage) and malignant (advanced stage) breast cancer cells were implemented with an Extreme Learning Machine (ELM) algorithm. In [10], an optical method for identifying and treating breast cancer, focused on computed tomography laser mammography (CTLM), was suggested. In the CTLM system, two supervising ML methods, SVM and multi-layer neural network were compared to diagnose angiogenesis.

The suggested scheme in [11] was built on the extreme gradient boosting (XGBoost) classifier with the following four equivalent strategies: transition, re-sampling, clustering, and collaborative learning to enhance balanced training performance. The findings indicated that the highest predictive performance for an empiric event was the XGBoost associated with re-sampling and clustering techniques. Three tuning strategies were examined in [12]. The grid search and particle swarm optimization developed more precise classifications for correct diagnosis compared to the Weka tool. A review of ML methods, which included artificial neural networks (ANN), Naïve Bayes (NB), and SVM ML methods, was suggested again in [13] for the detection of breast cancer. The linear SVM and the medium KNN, using eight characteristics, recorded the highest accuracy in [14].

A brief survey was presented in [15], again centered on the comparative study of ML strategies such as KNN, classification and regression trees, NB and SVM. Next, an infrared, high-precision and hand-held ML algorithm focusing on breast cancer identification was developed in [16], where the classifier was built on a trained linear SVM and a convolutional network (CNN) to evaluate the decision boundary.

The AdaBoost-ELM was introduced in [17], and it was seen that the suggested collective learning could successfully enhance the memory and consistency of the classification. In [18], supervised ML techniques such as DT, KNN, RF, and Gaussian NB were used to determine the risks associated with breast cancer by evaluating the biomarkers concerned. Reference [19] also introduced ML strategies such as RF, NB, SVM, and KNN to diagnose breast cancer. Detailed research on traditional ML and a deep learning solution for the multi-classification of breast cancer histopathology photos was presented [20], which concentrated on a comparative analysis of various testing methods for breast cancer. Cross-task ELM for sorting the breast cancer images with deep convolutional features was then discussed in [21], and an analog ML classifier for the sorting of breast cancer was analyzed in [22].

The SVM, DT, RF, LR, AdaBoost, ELM, CLTM, XGBoost, NB, KNN, ANN, and CNN have been used in the above-discussed research for breast cancer diagnosis. Nonetheless, none of the reports have applied the Category

[1] School of Electronics Engineering, Kalinga Institute of Industrial Technology Deemed to be University, Bhubaneswar, India; Email ID: 1730161@kiit.ac.in
[2]Department of Electrical Engineering, National Institute of Technology, Silchar, India
[3]Faculty of Electrical, Electronic & Computer Engineering, Universitatea Politehnica Bucuresti, București-060042, Romania
[4]Dept. of Electronics and Communication Engineering, K R Mangalam University, Gurgaon, India

Boosting (CatBoost) method to classify breast cancer to the best of the authors' understanding. Therefore, in this article, the classification of breast cancer that classifies cancer into malignant (cancerous) or benign (non-cancerous) tumors was applied using the CatBoost algorithm.

CatBoost consists of two words **Cat**egory and **Boost**ing. It works well with categorical data. CatBoost [23] is an open-sourced ML algorithm. This can quickly be combined with deep learning applications like TensorFlow. This is a high-performance library designed to improve gradient boosting on DT. It produces the results usually needed by other ML techniques, with intensive data training. The main function of this library is outstanding efficiency without parameter tuning. This paper also presents a detailed analysis comparing the CatBoost with previously developed techniques. The LR, KNN, SVM, NB, DT, RF, and XGBoost have been implemented for the comparative analysis. The Python programming language has been used to implement all ML techniques [24].

The discussion from the next section onwards will be as follows: Section 2 discusses the overview of the breast cancer prediction model and its input-output samples for classification. Description of CatBoost and other ML algorithms/classifiers in Section 3 whereas in Section 4, the implementation of ML models and results analysis is presented. The concluding remarks and future scopes is provided in Section 5.

2. OVERVIEW OF BREAST CANCER PREDICTION MODEL AND MACHINE LEARNING

2.1. BREAST CANCER PREDICTION MODEL

This breast cancer data samples are collected from the UCI ML repository [25]. The total samples are 699. The attributes that the data contains are:

A1: Sample Identity Number,
A2: Clump Thickness
A3: Uniformity of Cell Size,
A4: Uniformity of Cell Shape,
A5: Marginal Adhesion,
A6: Single Epithelial Cell Size
A7: Bare Nuclei,
A8: Bland Chromatin,
A9: Normal Nucleoli,
A10: Mitoses,
A11: Class – Benign (non-cancerous) or malignant (cancerous).

The simple block diagram for predicting breast cancer is shown in Fig. 1. Eight attributes (**A2-A10**) mentioned above are taken as input to the model, whereas the final **A11** is the output to the model. Based on the sample data in [26], the goal is to develop an effective and productive model with the best possible precision for potential patients. Several ML methods have been executed to attain this objective. Next, we provide an overview of ML and its categories.

2.2. MACHINE LEARNING

The ML is a sub-category of AI. It deals with learning from data sets. The ML has already been found to be an important area of biomedical science with multiple uses for a specific collection of biological samples, utilizing various techniques and algorithms. Supervised and unsupervised learning are the two popular forms of ML approaches.

In supervised learning, the collection of training information is used to approximate the input data to the target value. On the other side, no designated instances are available under the unsupervised learning approaches, and so there is no sense of output during the learning experience. Consequently, it is up to the learning scheme to identify correlations or discover the classes of input data. This method may be conceived of as a grouping problem of supervised learning. The classification function corresponds to a learning process that classifies the data into a collection of finite classes. The other two rising ML functions are regression and clustering. In the case of regression problems, the learning method provides the value of the data as an output. Clustering is a simple unsupervised process in which one seeks to identify groups or clusters to classify data objects. Based on this method, each new sample may be allocated to one of the defined clusters with related characteristics.

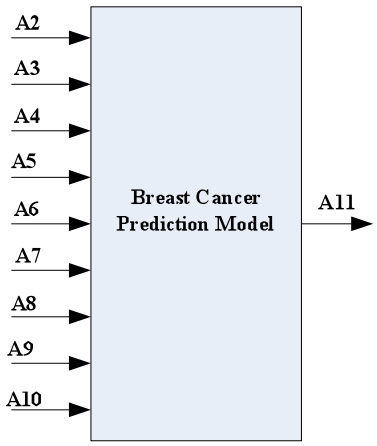


Fig. 1 – Block representation of breast cancer prediction model.

3. DESCRIPTION OF CATBOOST AND OTHER MACHINE LEARNING ALGORITHMS

There is a wide variety of classification applications from medicine to marketing. Classification models include: (i) CatBoost, (ii) XGboost, (iii) LR, (iv) KNN, (v) Kernel SVM, (vi) NB, (vii) DT (viii) RF and so on. In this segment, a brief account of these methods are presented.

3.1. CATEGORY BOOSTING

CatBoost is a gradient boosting execution that uses conditional DTs as basic predictors. A DT model is constructed by a recursive division of the feature space in many tree nodes centered on the values of many of the splitting attributes. The binary variables are used as parameters. It means that certain features surpass a certain threshold. Each final area (tree leaf) is given a value that is the region response estimation for the regression task. More details of the algorithm can be found in [23, 27–29]. The main features of CatBoost algorithm are:

- Great quality without parameter tuning
- Categorical feature support
- Fast and scalable CPU version
- Improved accuracy
- Fast prediction

3.2. EXTREME GRADIENT BOOSTING

The XGBoost [11] is designed using the rule of gradient boosting trees. Trees could be designed easily, and concurrent processing is possible. XGBoost is an efficient ensemble learning process that can be applied for numerous medical purposes. Other explanations for using XGBoost include multiple organized or categorical variables in the data collection, data distribution assumption is not needed, and tree-based approaches also work fine on unbalanced datasets.

3.3. LOGISTIC REGRESSION

The LR method [8] is a binary sorting approach, where designers forecast with two classes. The LR tests the association between the reliant variable and one or even more independent variables by calculating the probability. First, the equation of linear regression is taken as:

$$y = b_0 + b_1 x\,. \tag{1}$$

Using the sigmoid function whose equation is:

$$P = \frac{1}{1+e^{-y}}\,. \tag{2}$$

Taking the value of y from Eqn. (2) and substituting in Eqn. (1), we get the equation of LR as:

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 x\,. \tag{3}$$

The above Eqn. (3) represents the expression of LR.

3.4. K-NEAREST NEIGHBOURS

The KNN may be used for both classification and regression statistical issues [8]. Nevertheless, it is most commonly used for labeling issues in the sector. The KNN algorithm is a quick, easy-to-implement supervised ML algorithm. For example, suppose the graph between dependent and independent variables is given and the data has been classified into two categories. The new data point is allocated to the group with the most neighbors. The nearest neighbors of the new data point are given according to their Euclidean distance, whose equation is specified as:

$$d = \sqrt{(h_2 - h_1)^2 + (v_2 - v_1)^2}\,, \tag{4}$$

where, h_2–h_1 is the distance along the horizontal axis, and v_2–v_1 is the vertical axis.

3.5. KERNEL SUPPORT VECTOR MACHINE

The SVM is about finding the best-fit decision boundary, which will help us separate our space into classes [8]. It is valid for linearly separable data. For non-linear data, kernel SVM is used. For these types of data, we have to use kernel SVM. If the data is non-linear, it is first mapped to a higher dimension, and then a separator exists in a higher dimension. When the separator is found, it is projected back into the original dimension.

3.6. NAÏVE BAYES

For ML, the NB classifier is a family of classifiers that focus on interpreting the Bayes theorem [22]. There are two of the easiest Bayesian structures. Nevertheless, it should be paired with kernel density estimation and high precision rates. This functions solely based on the Bayes theorem. The principle of Bayes defines the likelihood of an occurrence based on previous information of the circumstances that could be connected to the event.

3.7. DECISION TREE

The concept of splitting the criterion resides behind the knowledge of every DT classifier [8]. The DTs are viewed as analogous to a flow map, with a tree layout in which cases are listed according to their characteristic values. The node in the DT signifies the instance, the test results are characterized by the branch, and the leaf node epitomizes the class name. The DT acquires data in the form of a tree but may also be interpreted as a series of discrete laws to make things simpler to grasp. The most significant benefit of the DT classifier is the capacity to utilize various subsets of features and decision-making at specific points of classification.

3.8. RANDOM FOREST

The RF method yields a set of DTs, randomly selected subsets of the training sample [8]. It further integrates the votes of the several DTs to conclude the last class of the test item. The RFs build DTs on randomly chosen data sets, make projections from each tree and choose the best approach through a vote. This also offers a relatively clear predictor of the value of the function. The elementary parameters for the RF approach may be the total number of trees to be created and the tree-related decision parameters, such as minimum split, split criteria, etc.

4. IMPLEMENTATION OF PROPOSED MODELS AND RESULTS ANALYSIS

In order to implement the proposed CatBoost based breast cancer diagnosis model and its comparative analysis with other classifiers, the Python coding language on Google Collaborator platform has been selected [30]. The advantage of this platform is that many ML classifiers are inbuilt in this. For assessing the performance of any classifier techniques, three parameters are generally defined: accuracy, confusion matrix, and receiver operating characteristic (ROC) graph.

Accuracy (A_{CC}) is a metric of the right estimation of the classifier, which gives general knowledge about how frequently objects are misclassified. It is given as:

$$A_{cc} = \frac{TN + TP}{TN + TP + FN + FP}\,, \tag{5}$$

where, *TP*: Number of true positives; *FP*: Number of false positives; *TN*: Number of true negatives; *FN*: Number of false negatives.

Table 1.

Confusion Matrix

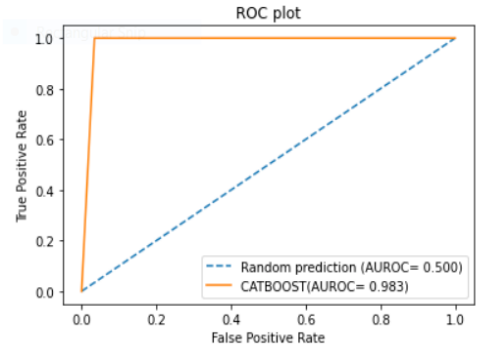
	NO (predicted)	YES(predicted)
NO(actual)	TN	FP
YES(actual)	FN	TP

Next, the Confusion Matrix (CM), as provided in Table 1 comprises the classifier's real and projected sorting results. This matrix assesses the behavior of the ML models. The ROC plot is a schematic diagram. It displays the predictive potential of a two-level classifier. In the ROC plot, the true positive rate (TPR) is presented. The TPR is the advantage

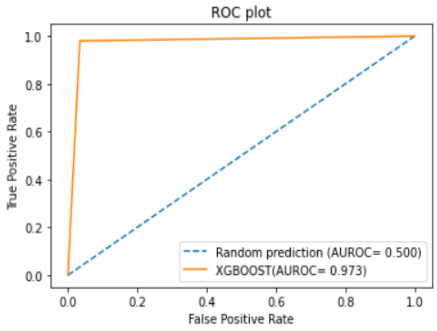
on the vertical axis in contrast to the false positive rate (FPR). It is the cost on the horizontal axis. Both horizontal and vertical axes are in the range [0,1]. The TPR and the FPR for each potential threshold value are acquired, and then the graph is plotted. The ROC represents the efficiency of the prediction models by demonstrating the trade-off of the expense and the benefit. The region below the ROC shows the efficiency of the ML algorithm. For training and testing the breast cancer diagnosis model, a total of 137 samples has been selected. 87 samples were chosen for training, whereas 37 samples were selected for testing the models. The Python codes were written for building the breast cancer diagnosis models. Based on Python programming, the CM, ACC, and ROC graphs are obtained.

Figure 2 presents the ROC curve for all models. As shown in the figure, the area under ROC (AUROC) is highest in CatBoost as compared to other ML algorithms.

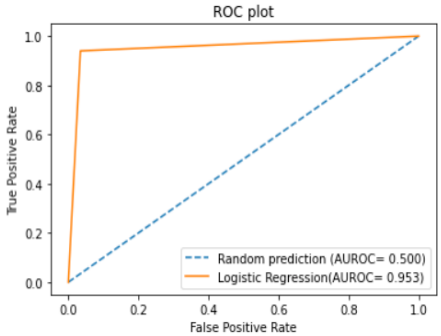
Next, the ACC for all diagnosis models is presented in Table 2. The **CatBoost** model gives an accuracy of 97.81 % on the training set. Out of 87 benign tumors, only three were predicted incorrectly, while the remaining 84 were predicted correctly. Similarly, the 50 malignant tumors were predicted correctly.



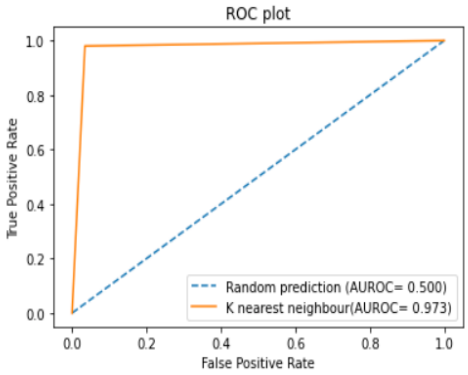
(a) CatBoost Model



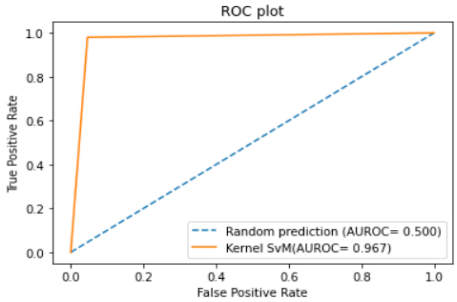
(b) XGBoost Model



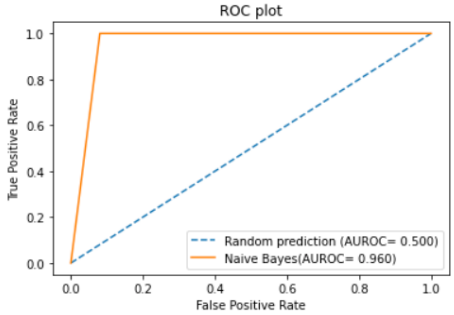
(c) LR Model



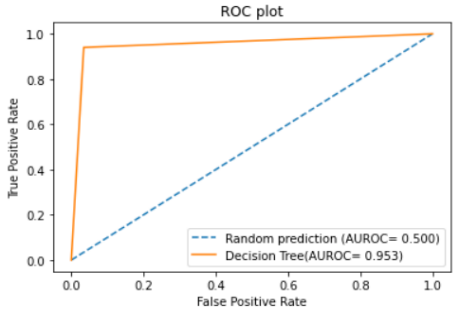
(d) KNN Model



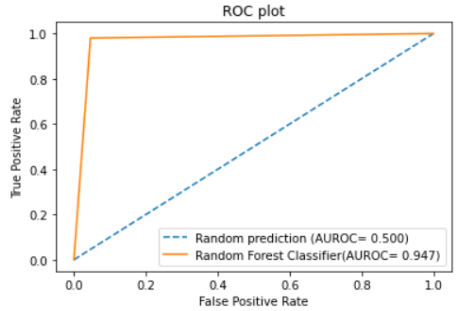
(e) Kernel SVM Model



(f) NB Model



(g) DT Model



(h) RF Model

Fig. 2 – ROC curves for CatBoost and different breast cancer diagnosis models.

The **XGBoost** model gives an accuracy of 97.08 % on the training set. Out of 87 benign tumors, only three were predicted incorrectly, while the remaining 84 were predicted correctly. Similarly, out of 50 malignant tumors, only one was predicted incorrectly, while the remaining 49 were predicted correctly. The **LR** model gives an accuracy of 95.6 % on the training set. Out of 87 benign tumors, only three were predicted incorrectly, while the remaining 84 were predicted correctly.

Similarly, out of 50 malignant tumors, only three were predicted incorrectly, while the remaining 47 were predicted correctly. This **KNN model** gives an accuracy of 97.08 % on the training set. Out of 87 benign tumors, only three were predicted incorrectly, while the remaining 84 were predicted correctly. Similarly, only one was predicted incorrectly out of the 50 malignant tumors, while the remaining 49 were predicted correctly. Finally, confusion matrices for all models are given in Table 3.

Table 2.

Accuracy values for different breast cancer models

ML Algorithms	**Accuracy (%)**
CatBoost	97.80
XGBoost	97.08
LR	95.60
KNN	97.00
Kernel SVM	96.30
NB	94.80
DT	95.60
RF	94.80

The **Kernel SVM** model gives an accuracy of 96.3 % on the training set. Out of 87 benign tumors, only four were predicted incorrectly, while the remaining 83 were predicted correctly. Similarly, out of 50 malignant tumors, only one was predicted incorrectly, while the remaining 49 were predicted correctly. The NB model gives an accuracy of 94.8 % on the training set. Out of 87 benign tumors, only seven were predicted incorrectly while the remaining 80 were predicted correctly.

Similarly, all 50 malignant tumors were predicted correctly. The **DT** model gives an accuracy of 95.6 % on the training set. Out of 87 benign tumors, only three were predicted incorrectly, while the remaining 84 were predicted correctly. Similarly, out of 50 malignant tumors, only three were predicted incorrectly, while the remaining 47 were correctly predicted. Based on the above analysis, the CatBoost is found to give a better prediction than all other classifiers.

Table 3.

Confusion Matrix for different breast cancer models

ML Algorithms	**Confusion Matrix**
CatBoost	$\begin{bmatrix} 84 & 3 \\ 0 & 50 \end{bmatrix}$
XGBoost	$\begin{bmatrix} 84 & 3 \\ 1 & 49 \end{bmatrix}$
LR	$\begin{bmatrix} 84 & 3 \\ 3 & 47 \end{bmatrix}$
KNN	$\begin{bmatrix} 84 & 3 \\ 1 & 49 \end{bmatrix}$
Kernel SVM	$\begin{bmatrix} 83 & 4 \\ 1 & 49 \end{bmatrix}$
NB	$\begin{bmatrix} 80 & 7 \\ 0 & 50 \end{bmatrix}$
DT	$\begin{bmatrix} 84 & 3 \\ 3 & 47 \end{bmatrix}$
RF	$\begin{bmatrix} 83 & 4 \\ 3 & 47 \end{bmatrix}$

5. **CONCLUSION**

It is crucial to diagnose deadly diseases like breast cancer at an early stage. In this regard, the CatBoost based ML algorithm seems to be a better classifier for breast cancer prediction than other classifiers considered in this study. It gives an accuracy of 97.8 % that is greater than the rest of the ML procedures. The confusion matrix also provides superior prediction than other approaches. Therefore, the proposed CatBoost based breast cancer classification model could be used to diagnose the disease at an early stage. This would be helpful in reducing the fatality rate due to the disease.

Received on December 1, 2020

REFERENCES

1. G. Kacso, et al., *Using a thermal sensing bra in early locating the breast tumors*, Rev. Roum. Sci. Techn.– Électrotechn. et Énerg, **61**, *1*, pp. 89-93, 2016.
2. G.M. Vasilescu et al., *Two devices equipped with temperature sensors used to detect and locate incipient breast tumors*, Rev. Roum. Sci. Techn.– Électrotechn. et Énerg, **61**, *4*, pp. 441-445, 2018.
3. N. Sabri, A. Tlemçani, A. Chouder, *Real-time diagnosis of battery cells for stand-alone photovoltaic system using machine learning techniques*, Rev. Roum. Sci. Techn.– Électrotechn. et Énerg, **66**, *2*, pp. 105-110, 2021.
4. A. Al Nahid,Y. Kong, *Involvement of machine learning for breast cancer image classification: A survey*, Comput. Math. Methods Med., 2017.
5. Z. Nematzadeh, R. Ibrahim, A. Selamat, *Comparative studies on breast cancer classifications with k-fold cross validations using machine learning techniques*, 2015 10th Asian Control Conf. Emerg. Control Tech. a Sustain. World, ASCC 2015, 2015.
6. Z. Nawel, A. Nabiha, D. Nilanjan, S. Mokhtar, *Adaptive semi supervised support vector machine semi supervised learning with features cooperation for breast cancer classification*, J. Med. Imaging Heal. Informatics, **6**, *1*, pp. 53–62, 2016.
7. S. Vural, X. Wang, C. Guda, *Classification of breast cancer patients using somatic mutation profiles and machine learning approaches*, BMC Syst. Biol., **10**, *3*, 2016.
8. S. Turgut, M. Dagtekin, T. Ensari, *Microarray breast cancer data classification using machine learning methods*, 2018 Electr. Electron. Comput. Sci. Biomed. Eng. Meet. EBBT 2018, 2018.
9. A. Toprak, *Extreme learning machine (ELM)-based classification of benign and malignant cells in breast cancer*, Med. Sci. Monit., **24**, pp. 6537–6543, 2018.
10. A. Jalalian, B. Karasfi, *Machine learning techniques for challenging tumor detection and classification in breast cancer*, 2018 9th Conf. Artif. Intell. Robot. 2nd Asia-Pacific Int. Symp. AIAR 2018, pp. 19–24, 2018.
11. C.C. Chang, S.H. Chen, *Developing a novel machine learning-based classification scheme for predicting spcs in breast cancer survivors*, Front. Genet., **10**, pp. 1–6, 2019.
12. A. Idri, M. Hosni, I. Abnane, J.M. Carrillo de Gea, J.L.F. Alemán, *Impact of parameter tuning on machine learning based breast cancer classification*, Adv. Intell. Syst. Comput., **932**, pp. 115–125, 2019.

13. D.A. Omondiagbe, S. Veeramani, A.S. Sidhu, *Machine learning classification techniques for breast cancer diagnosis*, IOP Conf. Ser. Mater. Sci. Eng., **495**, *1*, 2019.
14. T. Xieet al., *Machine learning-based analysis of MR multiparametricradiomics for the subtype classification of breast cancer*, Front. Oncol., **9**, pp. 1–10, 2019.
15. M.M. Kaklamanis, M. Filippakis, *A comparative survey of machine learning classification algorithms for breast cancer detection*, ACM Int. Conf. Proceeding Ser., 2019.
16. H.T. Iqbal, B. Majeed, U. Khan, M.A. Bin Altaf, *an infrared high classification accuracy hand-held machine learning based breast-cancer detection system*, BioCAS 2019 - Biomed. Circuits Syst. Conf. Proc., pp. 1–4, 2019.
17. M. Sharifmoghadam, H. Jazayeriy, *Breast cancer classification using adaboost- extreme learning machine*, 5th Iran. Conf. Signal Process. Intell. Syst. ICSPIS 2019, 2019.
18. R. Ray, A.A. Abdullah, D.K. Mallick, S.R. Dash, *Classification of benign and malignant breast cancer using supervised machine learning algorithms based on image and numeric datasets*, J. Phys. Conf. Ser., **1372**, *1*, 2019.
19. Y. Khourdifi, M. Bahaj, *Applying best machine learning algorithms for breast cancer prediction and classification*, 2018 Int. Conf. Electron. Control. Optim. Comput. Sci. ICECOCS 2018, 2019.
20. S. Sharma, R. Mehra, *Conventional machine learning and deep learning approach for multi-classification of breast cancer histopathology images — a comparative insight*, J. Digit. Imaging, **33**, *3*, pp. 632–654 , 2020.
21. P. Wang et al., *Cross-task extreme learning machine for breast cancer image classification with deep convolutional features*, Biomed. Signal Process. Control, **57**, p. 101789, 2020.
22. S.T. Chandrasekaran, R. Hua, I. Banerjee, A. Sanyal, *A fully-integrated analog machine learning classifier for breast cancer classification*," Electron., **9**, *3*, pp. 1–10, 2020.
23. A.V. Dorogush, V. Ershov, A. Gulin, *CatBoost: gradient boosting with categorical features support*, pp. 1–7, 2018. Available at: https://arxiv.org/pdf/1810.11363.pdf.
24. M. Popovic et al., *Formal verification of local and distributed python software transactional memories*, Rev. Roum. Sci. Techn.– Électrotechn. et Énerg, **64**, *4*, pp. 423-428, 2019.
25. O.L. Mangasarian, W.N. Street, W.H. Wolberg, *Breast cancer diagnosis and prognosis via linear programming*, Operations Research, **43**, *4*, pp. 570-577, 1995.
26. Available at https://archive.ics.uci.edu/ml/datasets.php .
27. Y. Zhang, Z. Zhao, J. Zheng, *CatBoost: A new approach for estimating daily reference crop evapotranspiration in arid and semi-arid regions of Northern China*, J. Hydrol., **588**, p. 125087, 2020.
28. G. Huang et al., *Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions*, J. Hydrol., **574**, pp. 1029–1041, 2019.
29. S. Sakila, S. Garg, T. Yeole, H. Yadav, *Earthquake time prediction using catboost and SVR*, Int. J. Innov. Technol. Explor. Eng., **9**, *1*, pp. 225–229, 2019.
30. B. Kordic et al., *An evolutionary computational system architecture based on a software transactional memory*, Rev. Roum. Sci. Techn.– Électrotechn. et Énerg, **61**, *1*, pp. 47-52, 2021.