JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 2, Maret 2025, hlm. 297 – 304 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i2.3661
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

PREDICTING OF BREAST CANCER RISK USING MACHINE LEARNING WITH FEATURE SELECTION THROUGH XGBOOST

Cahya Mutiara Al Azhar[1*], Pujiono[1]
[1]Information System Universitas Dian Nuswantoro
email: *112202106720@mhs.dinus.ac.id

Abstract: Breast cancer is the leading cause of death for women globally, exacerbated by late detection. This study proposes a breast cancer risk prediction framework using XGBoost with SelectKBest feature selection. It aims to improve the accuracy and efficiency of early detection through exploratory data analysis, coding, SMOTE to address class imbalance, and feature selection (k=29). As a result, the XGBoost model achieved 98.1% accuracy, 98.1% recall, 98.1% f1-score, and 98.2% precision on test data, highlighting the importance of feature selection. These results are promising in patient prioritization (triage) for further examination, helping medical personnel identify high-risk patients, thus improving resource allocation efficiency. These findings validate SelectKBest and pave the way for the development of a machine learning-based clinical decision support system for breast cancer early detection workflows. This research contributes significantly to the application of machine learning to support early breast cancer detection.

Keywords: breast cancer; feature selection; machine learning; risk prediction; XGBOOST.

Abstrak: Kanker payudara menjadi penyebab utama kematian wanita global, diperparah deteksi yang terlambat. Penelitian ini mengusulkan kerangka prediksi risiko kanker payudara menggunakan XGBoost dengan seleksi fitur SelectKBest. Tujuannya meningkatkan akurasi dan efisiensi deteksi dini melalui analisis data eksploratif, pengkodean, SMOTE untuk mengatasi ketidakseimbangan kelas, dan seleksi fitur (k=29). Hasilnya, model XGBoost mencapai akurasi 98.1%, recall 98.1%, f1-score 98.1%, dan presisi 98.2% pada data uji, menyoroti pentingnya seleksi fitur. Hasil ini menjanjikan dalam penentuan prioritas pasien (triage) untuk pemeriksaan lebih lanjut, membantu tenaga medis mengidentifikasi pasien berisiko tinggi, sehingga meningkatkan efisiensi alokasi sumber daya. Temuan ini memvalidasi SelectKBest dan membuka jalan bagi pengembangan sistem pendukung keputusan klinis berbasis machine learning untuk alur kerja deteksi dini kanker payudara. Penelitian ini berkontribusi signifikan dalam penerapan machine learning untuk mendukung deteksi dini kanker payudara.

Kata kunci: kanker payudara; pembelajaran mesin; prediksi risiko ; seleksi fitur; XGBOOST.

INTRODUCTION

In 2020, breast cancer claimed the lives of approximately 685,000 women, representing 16% or one in six cancer-related deaths among women[1]. WHO introduced the Global Breast Cancer Initiative to address the lack of public health response to the breast cancer issue[2]. Asia, home to 59.5% of the global population, accounts for 50% of all cancer cases and 58.3% of cancer-related deaths. Europe (9.7% of global population) contributes 22.8% of cases,

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 2, Maret 2025, hlm. 297 – 304 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i2.3661
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

while the United States reaches 20.9% cases with a mortality rate of 14.2%.[3]. Early detection of breast tumors increases the chances of survival. Besides being easier to treat at an early stage, it also provides insight into cancer progression[4].

Breast cancer occurs when breast cells develop abnormally and divide rapidly, forming a tumor. Symptoms of advanced disease include bone pain, enlarged lymph nodes, difficulty breathing, and jaundiced skin[5]. Neoadjuvant chemotherapy has been increasingly used for breast cancer patients in recent decades[6]. To get a significant patterns and insights from complex has changed due to the field of machine teaching (ML). Large-scale datasets proven effective in fields such as data mining, pattern recognition, and biotechnology[7]. Building robust machine learning-based prediction models is a complex process influenced by various factors. Addressing these concerns is crucial to maximizing machine learning's potential in advancing breast cancer diagnosis and therapy[8].

Breast cancer prediction using machine learning methods is also underway[9]. XGBoost's power and efficiency enable the identification of non-linear relationships, reduction of overfitting, and management of missing data. Careful hyperparameter optimization can enhance system performance[10]. XGBoost was chosen due to its ability to handle null values, fine-tune hyperparameters, correct errors, and tolerate unbalanced data scales. This is important for achieving maximum accuracy[11]. Boosting techniques train the model repeatedly. These models, which have basic prediction rules slightly better than random guesses ("weak learning"), focus on "hard" examples that are difficult to predict, which is the basis of boosting[12].

Because of its superiority, the XGBoost algorithm has been proven in previous studies. As research conducted by [13] in their research resulted in the I-XGBoost classifier excelling in precision (99%), recall (1,000%), and f1-score (0.999%), according to the data. Another study [14] through a case study, XGBoost was implemented to develop a prediction model. One of them, the results of XGBoost + LP are with accuracy results (78.66%). Furthermore, research conducted by [15] with the results reveals that XGBoost surpasses multiple linear regression (MLR), support vector regression (SVR), and random forest (RF) in predicting wave run-up on sloped beaches, achieving a correlation coefficient (R2) with a mean absolute percentage error (MAPE) of 6.635% and a root mean square error (RMSE) of 0.03902. Then research conducted by [16] shows the XGBoosting classifier algorithm achieves the best performance with mung bean yields. The experimental results indicate a testing accuracy of 98.65% and a training accuracy of 99.8%. Another research by [17] the modified XGBoost model exhibited a 17% improvement in performance on the test set, with a 28% reduction in root mean square error, demonstrating significant enhancements after parameter optimization.

This study seeks to enhance the efficiency and precision of XGBoost in breast cancer detection using SeleckKBest feature selection method with previous findings that show how important it is to select features to achieve optimal performance [18].

METHOD

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 2, Maret 2025, hlm. 297 – 304 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i2.3661
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

This research consists of several stages, as shown in Figure 1.

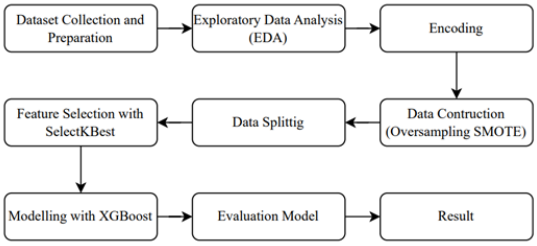

Figure 1. Research Stages

Dataset Collection and Preparation

This study used the 'Breast Cancer Wisconsin (Diagnostic) Dataset' from Kaggle. This dataset consists of 569 records and 33 columns with a multivariate data type on Figure 2. [19]. The feature 'unnamed:32' was removed as it did not provide useful information for analysis and modeling[20], [21].

	index	nan	dtype
0	id	0	int64
1	diagnosis	0	object
2	radius_mean	0	float64
3	texture_mean	0	float64
4	perimeter_mean	0	float64
5	area_mean	0	float64
6	smoothness_mean	0	float64
7	compactness_mean	0	float64
8	concavity_mean	0	float64
9	concave points_mean	0	float64
10	symmetry_mean	0	float64
11	fractal_dimension_mean	0	float64
12	radius_se	0	float64
13	texture_se	0	float64
14	perimeter_se	0	float64
15	area_se	0	float64
16	smoothness_se	0	float64
17	compactness_se	0	float64
18	concavity_se	0	float64
19	concave points_se	0	float64
20	symmetry_se	0	float64
21	fractal_dimension_se	0	float64
22	radius_worst	0	float64
23	texture_worst	0	float64
24	perimeter_worst	0	float64
25	area_worst	0	float64
26	smoothness_worst	0	float64
27	compactness_worst	0	float64
28	concavity_worst	0	float64
29	concave points_worst	0	float64
30	symmetry_worst	0	float64
31	fractal_dimension_worst	0	float64
32	Unnamed: 32	569	float64

Figure 2. Data Information

Exploratory Data Analysis (EDA)

Exploratory Data Analysis stage, shows that the diagnosis features have class imbalance as shown in Figure 3.

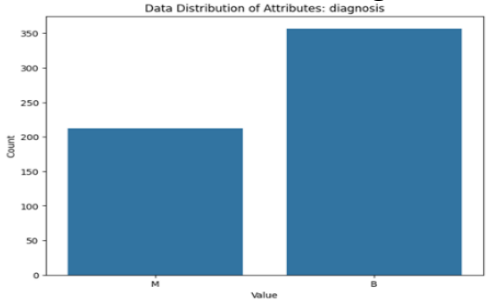

Figure 3. Diagnosis Distribution

Encoding

In this study, encoding was applied to transform categorical diagnosis features into numerical. The 'diagnosis' feature is encoded using one-hot encoding to become numeric or the number Benign becomes 0 and Malignant becomes 1.

Data Construction (Oversampling)

Oversampling used with the aim that the machine learning model is unbiased and able to learn patterns from all classes properly [22]. The oversampling method used is SMOTE, which creates synthetic samples based on minority class data.

Data Splitting

The dataset is divisible into training data and test data to train and test the performance of the model [23]. Data division with a ratio of 70:30.

Feature Selection

Feature selection involves identifying a subset of the most pertinent and informative. This process used to filter out important information from the data and ignore irrelevant things [24].

XGBoost

XGBoost (Extraordinary Angle Boosting) is one of the beat calculations in Machine Learning (ML). The parameters that will be utilized in this consider are, n_estimators parameter indicates the number of trees within the XGBoost gathering demonstrate, whereas max_depth sets the most extreme profundity of each tree, learning_rate manages each tree's contribution to the final outcome, subsample determines the proportion of data samples used for training each tree, and colsample_bytree specifies the fraction of features utilized

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 2, Maret 2025, hlm. 297 – 304 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i2.3661
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

in training each tree [24].

This algorithm operates by minimizing an objective function, comprising a loss function term to evaluate predicting errors on training data and a regularization term to penalize overly complex models, thus preventing overfitting, as defined in equation 1.

$$Obj(\Theta) = \sum_{i=1}^{n} L(yi, \hat{y}i) + \sum_{k=1}^{K} \Omega(fk) \quad (1)$$

Where *Obj(Θ)* is Objective function to be minimized to obtain the best model, *L(yi,y î)* is Loss function measuring the model's prediction error on the i-th data point, *Ω(fk)* Regularization function penalizing model complexity, preventing overfitting.

$$\hat{y}i = \sum_{k=1}^{K} f_k(x_i) \quad (2)$$

Where $y\hat{i}$ is Prediction for data point I and $\sum_{k=1}^{K} f_k(x_i)$ Prediction is the sum of the outputs of K functions, each evaluated on data point i (x_i).

Evaluation Model

After training, the model undergoes an initial evaluation to assess its performance using test data. The model evaluation is conducted thoroughly using the confusion matrix as the main visualization tool. A confusion matrix classifies predictions into four categories, True Negatives (TN) representing accurately predicted negative cases, True positives (TP) indicating correctly identified positive cases, False Negatives (FN) where positives cases were mistakenly predicted as negative, and False Positives (FP) denoting negative cases misclassified as positive. The confusion matrix then produces an accuracy metric that is used to measure on the whole validity of the model's predictions. The formulas used to compute evaluation metrics, including accuracy as the proportion of correct predictions, are outlined in equation 3. Precision for accurate positive predictive propositions in equation 4. Recall for the proportion of positive cases successfully detected in equation 5. F1-score for the harmonic mean between precision and recall in equation 6.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

RESULT AND DISCUSSION

This research evaluates the performance of the prediction algorithm by using various values of k. The value of k is an independent variable and can be freely modified. Figure 4 shows a correlation heatmap between numeric features, aiding in identifying linear relationships and informing encoding and feature selection. Several features exhibit strong positive correlations, such as radius_mean and perimeter_mean. Figure 5 presents a clustered correlation heatmap grouping features based on correlation similarity, assisting in identifying redundant features for feature selection, using a correlation threshold above 0.75. For example, radius_worst and perimeter_worst are grouped together due to their very high correlation. This visualization reveals features strongly correlated with the diagnosis variable Figure 5.

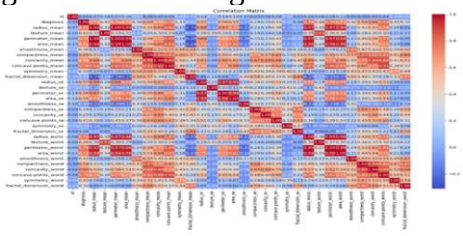

Figure 4. Correlation Heatmap

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 2, Maret 2025, hlm. 297 – 304 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i2.3661
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

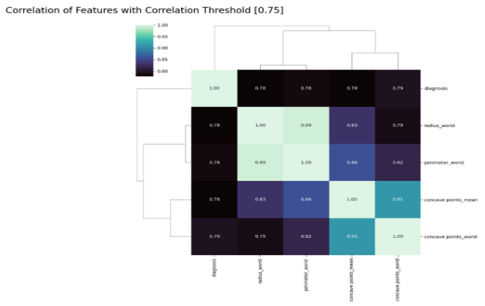

Figure 5. Clustered Heatmap Correlation

The next stage is data construction. Figure 6 shows the data distribution for the 'diagnosis' attribute. The data is divided into two classes: '1' (malignant) and '0' (benign), with the number of samples in class '0' (357) exceeding that of class '1' (212). This distribution indicates a class imbalance, prompting oversampling using SMOTE.

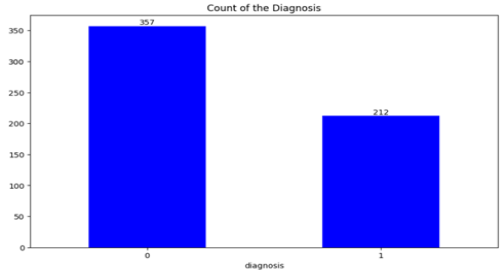

Figure 6. Distribution Class

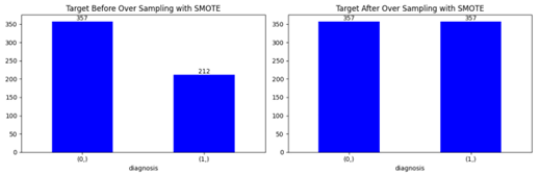

Figure 7. Compare Distribution Class Before and After Oversampling

Figure 7 illustrates the distribution of the target variable 'diagnosis' before and after oversampling with SMOTE. Before SMOTE, a significant class imbalance existed, with the malignant class 212 being much smaller than the benign class 357. After SMOTE, both classes have the same number of samples 357. This oversampling aims to address the class imbalance and improve the model's performance in predicting the minority class. 70% of the dataset was allocated for model training, and 30% for performance testing. The parameters used were: n_estimators=400, max_depth=10, learning_rate=0.1, subsample=0.8, and colsample_bytree=0.9.

Table 1. Result Feature Selection k=5

	Precision	Recall	F1-score	Support
0	0.93	0.98	0.95	108
1	0.98	0.93	0.95	107
Accuracy			0.95	215
Macro avg	0.96	0.95	0.95	215
Weighted avg	0.95	0.95	0.95	215

Table 2. Result Feature Selection k=10

	Precision	Recall	F1-score	Support
0	0.96	0.99	0.97	108
1	0.99	0.95	0.97	107
Accuracy			0.97	215
Macro avg	0.97	0.97	0.97	215
Weighted avg	0.97	0.97	0.97	215

Table 3. Result Feature Selection k=15

	Precision	Recall	F1-score	Support
0	0.96	0.99	0.97	108
1	0.99	0.95	0.97	107
Accuracy			0.97	215
Macro avg	0.97	0.97	0.97	215
Weighted avg	0.97	0.97	0.97	215

Table 4. Result Feature Selection k=20

	Precision	Recall	F1-score	Support
0	0.94	1.00	0.97	108
1	1.00	0.93	0.97	107
Accuracy			0.97	215
Macro avg	0.97	0.97	0.97	215
Weighted avg	0.97	0.97	0.97	215

Table 5. Result Feture Selection k=25

	Precision	Recall	F1-score	Support
0	0.94	1.00	0.97	108
1	1.00	0.93	0.97	107
Accuracy			0.97	215
Macro avg	0.97	0.97	0.97	215
Weighted avg	0.97	0.97	0.97	215

Table 6. Result Feature Selction k=29

	Precision	Recall	F1-score	Support
0	0.96	1.00	0.98	108
1	1.00	0.96	0.98	107
Accuracy			0.98	215
Macro avg	0.98	0.98	0.98	215
Weighted avg	0.98	0.98	0.98	215

Table 6. demonstrates that the XGBoost model with feature selection (k=29) achieved an accuracy of 0.98 on the test data, represent significantly high

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XI No 2, Maret 2025, hlm. 297 – 304 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v11i2.3661
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

performance compared to previous results.

Table 7. Result Feature Selection k=30

	Precision	Recall	F1-score	Support
0	0.96	1.00	0.98	108
1	1.00	0.95	0.98	107
Accuracy			0.98	215
Macro avg	0.98	0.98	0.98	215
Weighted avg	0.98	0.98	0.98	215

Table 8. Compare With Different k

k	Accuracy	Recall	F1-Score	Precision Score
5	0.953	0.953	0.953	0.955
10	0.972	0.972	0.972	0.973
15	0.963	0.963	0.963	0.964
20	0.967	0.967	0.967	0.969
25	0.967	0.967	0.967	0.969
29	0.981	0.981	0.981	0.982
30	0.977	0.977	0.977	0.978

Table 8 shows XGBoost model evaluation results on test data with various k values with number of selected features, SelectKBest method. The model with k=29 achieved the highest accuracy, recall, f1-score (0.981), and precision (0.982), indicating the best configuration for the XGBoost model.

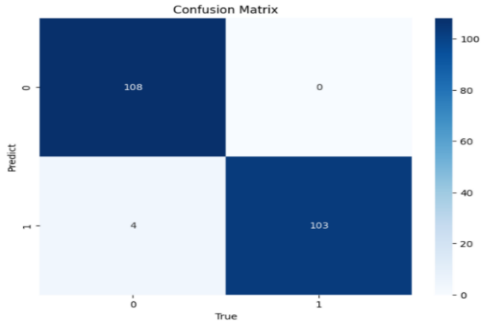

Figure 8. Confusion Matrix

Figure 8 presents the confusion matrix showing the classification model's evaluation results. True Negatives (TN) 108 negative data points correctly classified. False Positives (FP) 0, no negative data incorrectly classified as positive. False Negatives (FN) 4 positive data points incorrectly classified as negative. True Positives (TP) 103 positive data points correctly classified.

CONCLUSION

This study optimizes XGBoost for breast cancer risk prediction through SelectKBest feature selection. The model with k=29 produces optimal performance (accuracy, recall and f1-score 0.981, precision 0.982). SelectKBest (k=29) provides an XGBoost configuration that can be a benchmark. The results of this study have the potential to be integrated into a clinical decision support system for breast cancer screening and diagnosis. Further validation with a wider range of clinical data is required to test its generalizability and robustness. Future research is recommended to explore more advanced feature selection techniques, such as Recursive Feature Elimination (RFE) with cross-validation or genetic algorithm-based feature selection.

BIBLIOGRAPHY

[1] Y. S. Prabandari *et al.*, "'Alas … my sickness becomes my family's burden': A nested qualitative study on the experience of advanced breast cancer patients across the disease trajectory in Indonesia," *The Breast*, vol. 63, pp. 168–176, Jun. 2022, doi: 10.1016/j.breast.2022.04.001.

[2] M. Arnold *et al.*, "Current and future burden of breast cancer: Global statistics for 2020 and 2040," *The Breast*, vol. 66, pp. 15–23, Dec. 2022, doi: 10.1016/j.breast.2022.08.010.

[3] B. E. Patiño-Palma, L. López-Montoya, R. Escamilla-Ugarte, and A. Gómez-Rodas, "Trends in physical activity research for breast cancer - A bibliometric analysis of the past ten years," *Heliyon*, vol. 9,

no. 12, p. e22499, Dec. 2023, doi: 10.1016/j.heliyon.2023.e22499.

[4] S. M. Malakouti, M. B. Menhaj, and A. A. Suratgar, "ML: Early Breast Cancer Diagnosis," *Curr. Probl. Cancer Case Rep.*, vol. 13, p. 100278, Mar. 2024, doi: 10.1016/j.cpccr.2024.100278.

[5] Md. M. Hassan *et al.*, "A comparative assessment of machine learning algorithms with the Least Absolute Shrinkage and Selection Operator for breast cancer detection and prediction," *Decis. Anal. J.*, vol. 7, p. 100245, Jun. 2023, doi: 10.1016/j.dajour.2023.100245.

[6] A. De Luca *et al.*, "Neoadjuvant chemotherapy for breast cancer in Italy: A Senonetwork analysis of 37,215 patients treated from 2017 to 2022," *The Breast*, vol. 78, p. 103790, Dec. 2024, doi: 10.1016/j.breast.2024.103790.

[7] H. Xie, Y. Deng, J. Li, K. Xie, T. Tao, and J. Zhang, "Predicting the risk of primary Sjögren's syndrome with key N7-methylguanosine-related genes: A novel XGBoost model," *Heliyon*, vol. 10, no. 10, p. e31307, May 2024, doi: 10.1016/j.heliyon.2024.e31307.

[8] M. Darwich and M. Bayoumi, "An evaluation of the effectiveness of machine learning prediction models in assessing breast cancer risk," *Inform. Med. Unlocked*, vol. 49, p. 101550, 2024, doi: 10.1016/j.imu.2024.101550.

[9] V. Nemade and V. Fegade, "Machine Learning Techniques for Breast Cancer Prediction," *Procedia Comput. Sci.*, vol. 218, pp. 1314–1320, 2023, doi: 10.1016/j.procs.2023.01.110.

[10] S. Jafari, J.-H. Yang, and Y.-C. Byun, "Optimized XGBoost modeling for accurate battery capacity degradation prediction," *Results Eng.*, vol. 24, p. 102786, Dec. 2024, doi: 10.1016/j.rineng.2024.102786.

[11] C.-J. Tseng and C. Tang, "An optimized XGBoost technique for accurate brain tumor detection using feature selection and image segmentation," *Healthc. Anal.*, vol. 4, p. 100217, Dec. 2023, doi: 10.1016/j.health.2023.100217.

[12] N. Q. K. Le, D. T. Do, T.-T.-D. Nguyen, and Q. A. Le, "A sequence-based prediction of Kruppel-like factors proteins using XGBoost and optimized features," *Gene*, vol. 787, p. 145643, Jun. 2021, doi: 10.1016/j.gene.2021.145643.

[13] V. Jaiswal, P. Saurabh, U. K. Lilhore, M. Pathak, S. Simaiya, and S. Dalal, "A breast cancer risk predication and classification model with ensemble learning and big data fusion," *Decis. Anal. J.*, vol. 8, p. 100298, Sep. 2023, doi: 10.1016/j.dajour.2023.100298.

[14] M. Shanbehzadeh, H. Kazemi-Arpanahi, M. Bolbolian Ghalibaf, and A. Orooji, "Performance evaluation of machine learning for breast cancer diagnosis: A case study," *Inform. Med. Unlocked*, vol. 31, p. 101009, 2022, doi: 10.1016/j.imu.2022.101009.

[15] D. Tarwidi, S. R. Pudjaprasetya, D. Adytia, and M. Apri, "An optimized XGBoost-based machine learning method for predicting wave run-up on a sloping beach," *MethodsX*, vol. 10, p. 102119, 2023, doi: 10.1016/j.mex.2023.102119.

[16] A. M. Mequanenit, A. M. Ayalew, A. O. Salau, E. A. Nibret, and M. Meshesha, "Prediction of mung bean production using machine

learning algorithms," *Heliyon*, vol. 10, no. 24, p. e40971, Dec. 2024, doi: 10.1016/j.heliyon.2024.e40971.

[17] Z. Wang, X. Wu, and Y. Wu, "A spatiotemporal XGBoost model for PM2.5 concentration prediction and its application in Shanghai," *Heliyon*, vol. 9, no. 12, p. e22569, Dec. 2023, doi: 10.1016/j.heliyon.2023.e22569.

[18] T. Chen, X. Zhou, and G. Wang, "Using an innovative method for breast cancer diagnosis based on Extreme Gradient Boost optimized by Simplified Memory Bounded A*," *Biomed. Signal Process. Control*, vol. 87, p. 105450, Jan. 2024, doi: 10.1016/j.bspc.2023.105450.

[19] S. Batool and S. Zainab, "A comparative performance assessment of artificial intelligence based classifiers and optimized feature reduction technique for breast cancer diagnosis," *Comput. Biol. Med.*, vol. 183, p. 109215, Dec. 2024, doi: 10.1016/j.compbiomed.2024.109215.

[20] P. T. Teo *et al.*, "Determining risk and predictors of head and neck cancer treatment-related lymphedema: A clinicopathologic and dosimetric data mining approach using interpretable machine learning and ensemble feature selection," *Clin. Transl. Radiat. Oncol.*, vol. 46, p. 100747, May 2024, doi: 10.1016/j.ctro.2024.100747.

[21] V. Safavi, A. Mohammadi Vaniar, N. Bazmohammadi, J. C. Vasquez, O. Keysan, and J. M. Guerrero, "Early prediction of battery remaining useful life using CNN-XGBoost model and Coati optimization algorithm," *J. Energy Storage*, vol. 98, p. 113176, Sep. 2024, doi: 10.1016/j.est.2024.113176.

[22] X. Y. Liew, N. Hameed, and J. Clos, "An investigation of XGBoost-based algorithm for breast cancer classification," *Mach. Learn. Appl.*, vol. 6, p. 100154, Dec. 2021, doi: 10.1016/j.mlwa.2021.100154.

[23] P. Paulus, Y. Ruppert, A. Andreicovici, M. Vielhaber, and J. Griebsch, "Comparison of machine learning based methods on prediction quality of thin-walled geometries using laser-based Direct Energy Deposition," *Procedia CIRP*, vol. 124, pp. 781–784, 2024, doi: 10.1016/j.procir.2024.08.224.

[24] A. Maleki, M. Raahemi, and H. Nasiri, "Breast cancer diagnosis from histopathology images using deep neural network and XGBoost," *Biomed. Signal Process. Control*, vol. 86, p. 105152, Sep. 2023, doi: 10.1016/j.bspc.2023.105152.