Available online at http://scik.org

Commun. Math. Biol. Neurosci. 2025, 2025:96

https://doi.org/10.28919/cmbn/9392

ISSN: 2052-2541

ENHANCING BREAST CANCER DETECTION USING MACHINE LEARNING ON DATA FROM CUBAN WOMEN

KARLI EKA SETIAWAN[1,*], HAYYUN LISDIANA[2]

[1]Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia

[2]Department of Chemistry Education, Faculty of Mathematics and Natural Science, Universitas Negeri Jakarta, Jakarta 13220, Indonesia

Abstract: Breast cancer is becoming as the predominant cause of cancer-related mortality among women globally. Accurate and early diagnosis in detecting the presence of breast cancer can bring a positive impact in reducing mortality rates. This research explored the capabilities of a machine learning approach in detecting the presence of breast cancer in patients undergoing screening based on patient background parameters. This study utilized a publicly available dataset entitled Breast Cancer Risk Factors in Cuban Women obtained from Mendeley Data. This research contribution is the exploration and experiment of various machine learning models, such as support vector machine (SVM) using various kernels as our proposed model, logistic regression as our baseline model, and random forest as a comparison model with the best model in previous research that provided this dataset, with the result that our methodology, especially in handling preprocessing data and feature engineering, can improve most tested machine learning models to achieve perfect scores (100% accuracy, precision, recall, and F1-score), except for the SVM with radial basis function (RBF) kernel.

Keywords: breast cancer detection; machine learning; medical diagnosis; support vector machine; feature engineering.

2020 AMS Subject Classification: 68T01, 68T10, 97P80.

*Corresponding author

E-mail address: karli.setiawan@binus.ac.id

Received May 30, 2025

1. INTRODUCTION

There is a fact that breast cancer is becoming the most prevalent cancer among women around the world, and it is also becoming the primary cancer-related mortality [1]. Globally, approximately 2.3 million women were diagnosed with breast cancer in 2022, resulting in 670,000 fatalities, as per data from the World Health Organization [2]. The majority of research on breast cancer diagnosis has used mammography, ultrasound, and magnetic resonance imaging (MRI) [3]. Accurate and early diagnosis can play an important role in saving people from more lethal long-term consequences in the future. Moreover, there is a lack of access in rural and remote areas to detecting the presence of breast cancer. In an effort to mitigate this constraint, this investigation explored a machine learning approach that uses patient background parameters to predict the presence of breast cancer, utilizing a public dataset titled Breast cancer risk factor in Cuban women by Valencia et al. [4]. The objective and aim of this research are to improve the predictive capabilities of machine learning models through data preprocessing and feature engineering techniques, which may contribute to developing a more accessible, cost-effective, and accurate breast cancer screening tool. The motivation behind this research was to enhance and benchmark the results of a previous research, which was the pioneer in providing this dataset, and achieve 99.6% accuracy using a random forest model.

2. RELATED WORKS

There was research exploring the capability of a machine learning approach using a dataset titled "Breast Cancer Risk Factors in Cuban Women." The research was done by those who provided this dataset [4] [5]. They examined the efficacy of various machine learning models, including decision trees (DT), gradient boosting trees, random forests (RF), support vector machines (SVM), fast large margin classifiers, logistic regression, generalized linear models, and naïve Bayes, with random forests yielding the highest accuracy rate of 0.996. Another research investigating the capabilities of machine learning models in the research of breast cancer using public gene expression data was done by Wu et al. [6]. The machine learning models they used were SVM, K-Nearest Neighbors (KNN), Naïve bayes, and DT to perform binary classification on triple negative breast cancer (TNBC) and non-TNBC from the Cancer Genome Atlas, with SVM achieving the best result. Another research using a machine learning approach in breast cancer research was done by Rabiei et al., where they explored a dataset recorded between 2011 and 2021 from the Motamed Cancer Institute using random forest, multilayer perceptron, gradient

boosting trees, and genetic algorithm for binary classification in predicting malignant or benign with random forest as the best model by achieving 80%, 95%, 80%, and 0.53 for accuracy, sensitivity, specificity, and AUC, respectively [7].

Expanding on similar methodologies, Naji et al. proposed five machine learning algorithms, such as SVM, RF, logistic regression, DT, and KNN for binary classification on the Breast Cancer Wisconsin Diagnostic dataset [8]. Their best result was obtained from using the SVM model with a precision of 97.5% and 96.6% AUC. There was research performing a literature review study in combining biotechnology and a machine learning method for automation in early detection of breast cancer [9]. Biosensors were employed to detect specific biological analytes by converting cellular components like protein, DNA, or RNA into electrical signals applicable for breast cancer research. They found out that the Fuzzy Extreme Learning Machine with radial basis function (ELM-RBF) was the best model for breast cancer detection.

3. RESEARCH METHODOLOGY

3.1. Dataset, Preprocessing Dataset, and Feature Engineering

The dataset explored in this work was open data and was obtained from Mendeley Data titled Breast cancer risk factor in Cuban women by Valencia et al. [4]. Published on 31 August 2024, the dataset describes the risk factors for breast cancer in a patient group of Cuban women [5]. This dataset can be accessed from Mendeley data with the link https://data.mendeley.com/datasets/7jhddnpz2p/1 as open public data. The dataset contained 1697 breast cancer diagnosis cases in the 20–90 age group of Cuban women. This dataset comprised 23 patient background parameters, including identity number, age, menarche details, age at first childbirth, number of live births, duration of breastfeeding in months, quantity of first-degree family member diagnosed with breast cancer, number of breast biopsies, presence of atypical hyperplasia, patient race, year of data collection, body mass index (BMI), body weight, weekly exercise activity, alcohol/ethanol consumption status, tobacco-smoking, allergy information, emotional state, depression status, tumor histological classification, BI-RADS (Breast Imaging-Reporting and Data System) categories, and a final parameter used to classify the presence of cancer.

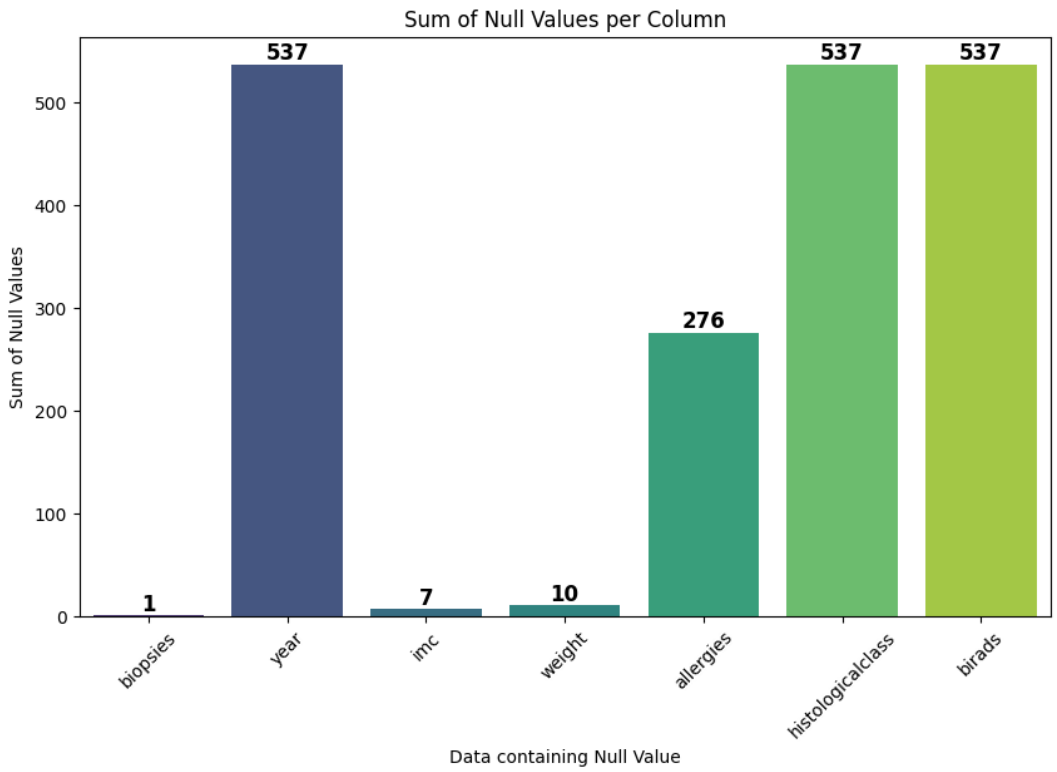

FIGURE 1. The quantity of missing values

The dataset contains many missing values, as explained in Figure 1, especially in year, allergy, histological class, and BIRADS parameters. To handle these missing values, first this research removed the data containing missing values on biopsies, IMC, and weight parameters due to small numbers in missing values. This research also removed the year parameter due to its irrelevance for predicting cancer based on unknown information containing year values between 2001 and 2018. Meanwhile, for allergies and BI-RADS data, this research imputed the missing value with a new categorical category named unknown as a new category, and for histological class data, this research imputed the missing value with the number 0 as in Figure 2. After handling missing values, now this research has 1686 data points.

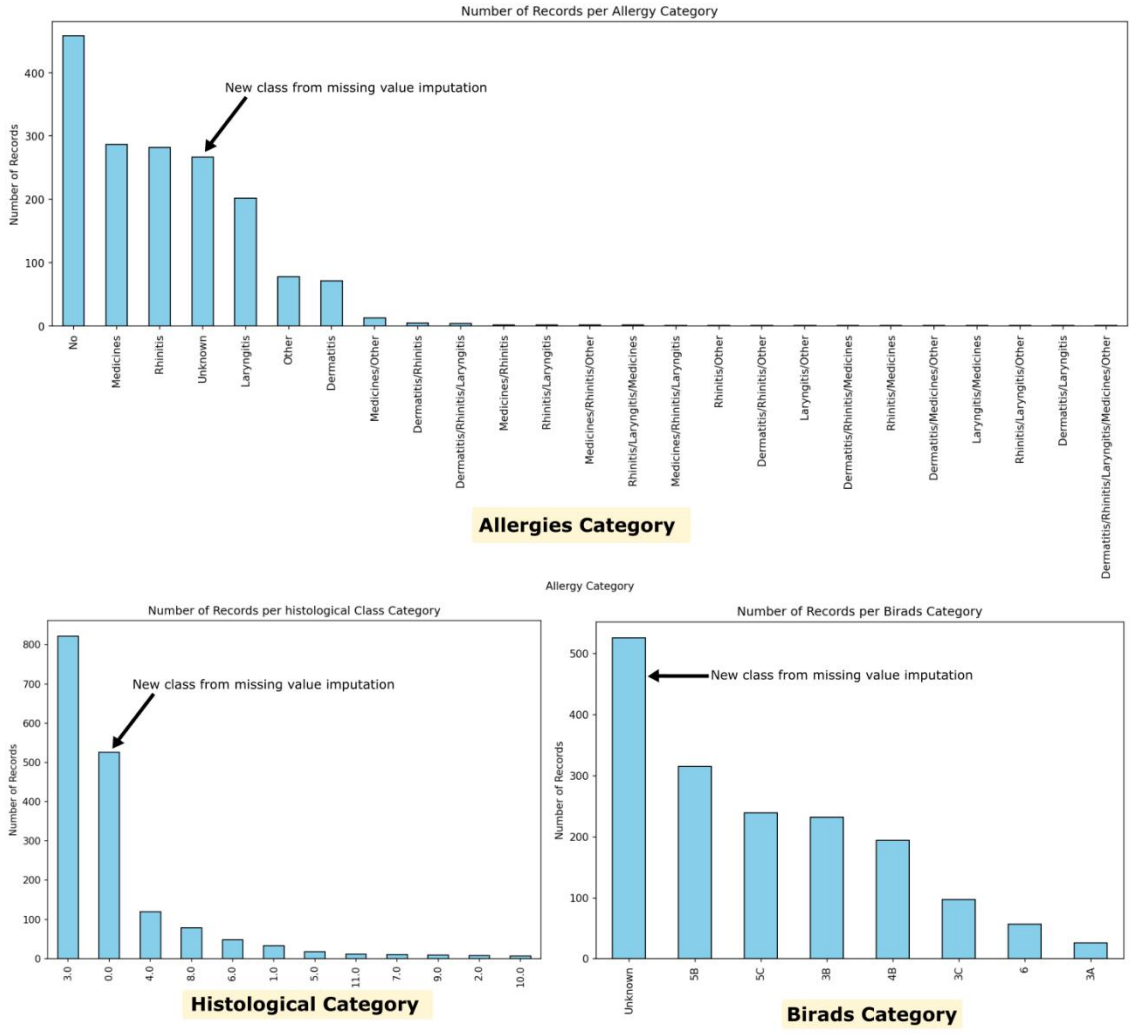

FIGURE 2. The process involves imputation of missing values into allergies, histological class, and BI-RADS data.

Some data from the dataset, such as age, menarche, biopsies, IMC, and weight, were initially in numeric or continuous data, so the distribution of those data can be illustrated with a histogram chart as in Figure 3. Meanwhile, some data from the dataset, such as hyperplasia, alcohol consumption, emotional status, tobacco consumption, depression status, and race, were initially in discrete or categorical data, so that the number of each unique value can be illustrated with a bar chart as in Figure 4. Meanwhile, certain data, including menopausal data, age at first birth, number of children, breastfeeding duration, exercise frequency, and the count of first-degree relatives with breast cancer, need specialized handling, as depicted in Figure 5.

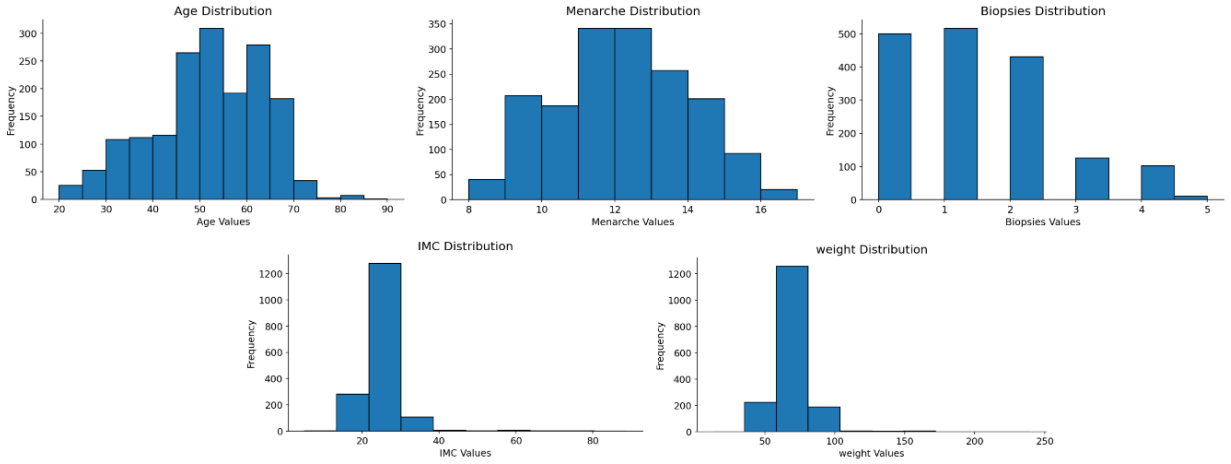

FIGURE 3. Other parameters distribution in dataset such as age, menarche, biopsies, IMC, and weight.

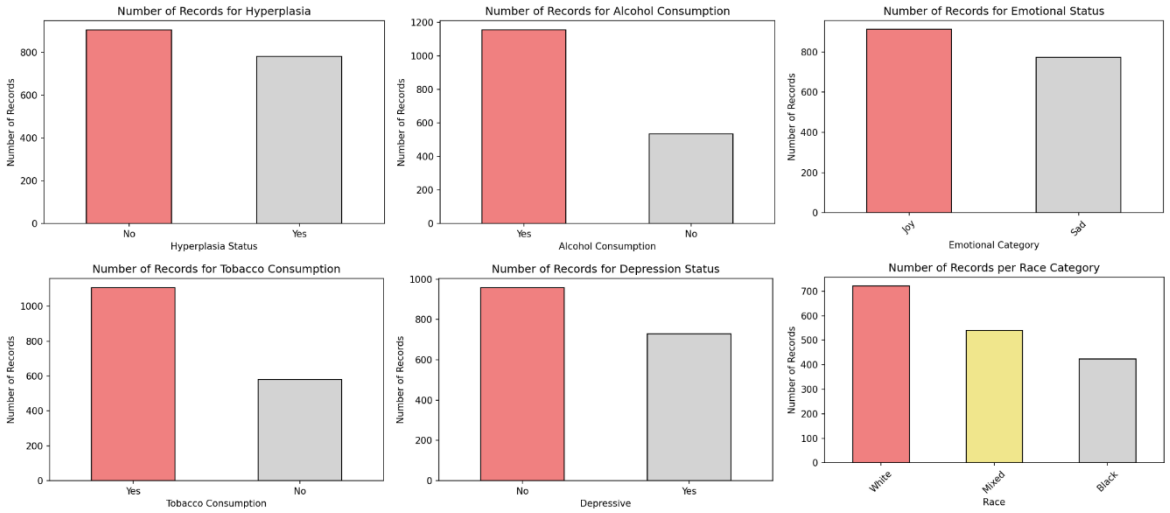

FIGURE 4. The dataset includes other parameters such as hyperplasia, alcohol consumption, emotional status, tobacco consumption, depression status, and race category distribution.

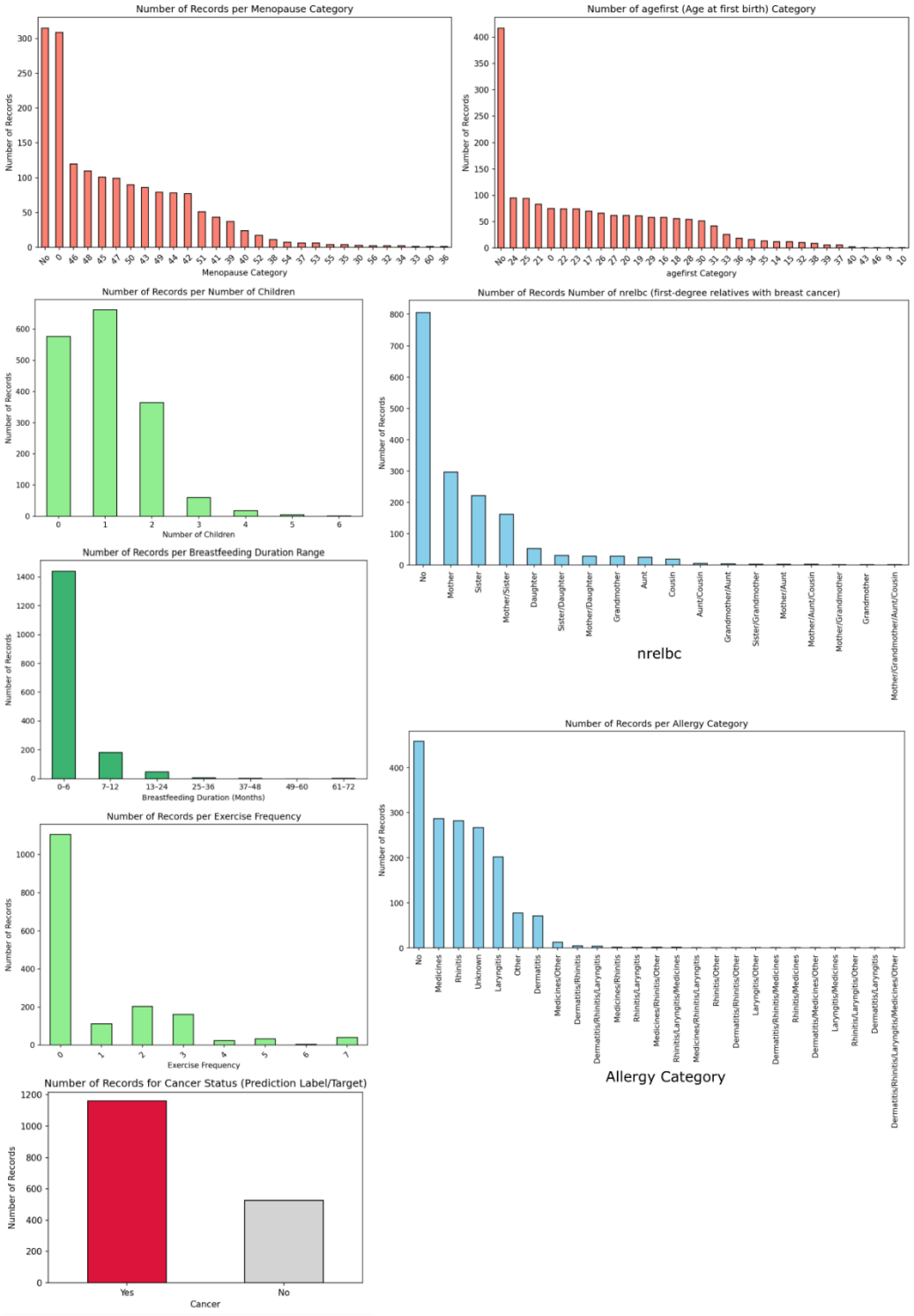

FIGURE 5. Other parameters were distributed in the dataset, such as menopause, age at first birth, number of children, breastfeeding duration, exercise frequency, first-degree relatives with breast cancer, and the last data used as a prediction label or target describing cancer.

The dataset recorded menopause data as integer values between 30 and 60. Based on our analysis of the dataset, as in Figure 5 on the number of records per menopause category, the data contain values of 0 (zero) and “No” data in large numbers. These two data categories mean that the patients have not experienced menopause, so this research transformed both the 0 (zero) and “No” classes into a single 0 (zero) class as feature engineering treatment. Like the menopause data, the agefirst column shows the age at which a patient first gave birth, recorded as whole numbers between 9 and 46. Our analysis, shown in Figure 5, reveals that there are two classes with values 0 (zero) and “No,” indicating that the patient has not given birth, so this study combined both the 0 (zero) and “No” classes into one 0 (zero) class as well as feature engineering treatment.

The children column, which describes the number of children had by patients, has 7 classes with values of 0, 1, 2, 3, 4, 5, and 5+. This research transformed 5+ class into 6 for representing the value of having more than five children so that the values will be 0, 1, 2, 3, 4, 5, and 6 as feature engineering treatment. The breastfeeding column, which describes the duration of breastfeeding in months containing values in integers between 0 and 72 (months), contained many inconsistent values so that this research needs to combine the duplicated classes with the same meaning value, for example, the “0” class, the “No” class, and the “No ” class into a single class with value 0, and another example, the “1” class and the “1 month” class into a single class with value 1. The exercise data having the same problem as the breastfeeding column, this research performed a similar treatment with the result shown in figure 5.

It can be seen in Figure 5 on the allergy category describing allergies suffered by patients and NRELBC data describing the number of first-degree family members with breast cancer data. This research implemented one-hot encoding using the MultiLabelBinarizer library from sklearn.preprocessing. So that, there is an increasing dimension for both nrelbc and allergy data from a single column becoming seven new columns, such as Nrelbc = Aunt, Nrelbc = Cousin, Nrelbc = Daughter, Nrelbc = Grandmother, Nrelbc = Mother, Nrelbc = No, and Nrelbc = Sister for nrelbc data and allergies = Dermatitis, allergies = Laryngitis, allergies = Medicines, allergies = No, allergies = Other, allergies = Rhinitis, and allergies = Unknown for allergy data. The newly generated columns from the one-hot encoding process will be binary class data, where the value 1 represents yes and the value 0 represents no. Finally, after performing feature engineering, this research used 32 feature data from number one to thirty-two and 1 target label at data number thirty-three as in Table 1.

TABLE 1. Data used for training machine learning

Data	Column	Data	Column
1	age	18	birads
2	menarche	19	Nrelbc = Aunt
3	menopause	20	Nrelbc = Cousin
4	agefirst	21	Nrelbc = Daughter
5	children	22	Nrelbc = Grandmother
6	breastfeeding	23	Nrelbc = Mother
7	biopsies	24	Nrelbc = No
8	hyperplasia	25	Nrelbc = Sister
9	race	26	allergies = Dermatitis
10	imc	27	allergies = Laryngitis
11	weight	28	allergies = Medicines
12	exercise	29	allergies = No
13	alcohol	30	allergies = Other
14	tobacco	31	allergies = Rhinitis
15	emotional	32	allergies = Unknown
16	depressive	33	Cancer (Target/Label)
17	histologicalclass		

3.2. Machine Learning Models

This study investigated and experimented with the capabilities of some machine learning models, such as SVM, logistic regression, and RF. Although this study focused on SVM, previous research by Valencia et al. found that RF was the best model/algorithm, so this study also included random forest [5]. This study also used logistic regression as a base model comparison for better analysis.

3.2.1 Support Vector Machine (SVM)

Cortes and Vapnik first presented the SVM as a way to find the most effective hyperplane in a multidimensional space [10]. Basically it perform binary classification by increasing the separation area using a hyperplane separation between the two classes [11]. The SVM method is often recognized for its outstanding effectiveness relative to other classifiers [12]. This research proposes three distinct SVM models for binary classification: linear SVM, polynomial SVM, and

a radial basis function (RBF)-SVM, all accessible in the Scikit-learn toolkit [13]. A kernel function is a mapping procedure used to improve the training set's resemblance to a linearly separated data set. Mapping enhances the dimensionality of data collection/datasets and is accomplished effectively by a kernel function.

3.2.2 Random Forest (RF)

One of the popular supervised machine learning models for regression and classification tasks is Random Forest (RF) [14]. RF models operate in a manner analogous to ensemble learning by bagging, comprising several weak learners that utilize decision trees [15]. Both ensemble learning with bagging and random forest will learn the data from different bootstraps of the data for every weak learner inside. The difference between both of them is that the RF model implements the arbitrary selection of features from the dataset so that every weak learner inside will learn different bootstrap data with different features that will create many various decision trees, which makes this model named Random Forest [16] [17].

3.2.3 Logistic Regression

The extension of linear regression for the classification case is logistic regression, where the output of the model is a discrete value [18] . This model can map independent variables (X) into outcome probabilities (p) between 0 and 1, which is suitable for this research case in binary classification, which makes this model become the base model that needs to be compared with the proposed model. The equation of logistic regression is notated in equation 1, where b_0 represents the intercept and b_1, b_2, b_3 until b_n represents the slope [19].

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 X \tag{1}$$

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \cdots + b_n X_n \tag{2}$$

3.3. Evaluation Metrics

This study implemented various evaluation and assessment metrics, including confusion matrix, precision, recall, accuracy, and F1-score, to evaluate and contrast the performance of our machine learning models [20]. A confusion matrix is a list that shows the evaluation of the model by comparing the predicted and actual class labels. The confusion matrix has four terms, such as correctly predicting a positive class as True Positive (TP), correctly predicting a negative class as True Negative (TN), mistakenly predicting a positive class as False Positive (FP), and mistakenly predicting a negative class as False Negative (FN) [21] [22]. The accuracy shows the correct prediction of all class in both positive and negative classes, denoted as in equation 3. The recall

exhibits sensitivity by measuring the number of correctly detected positive cases, denoted as in equation 5. Precision assesses the amount of trust in a model by assessing its correctness, denoted as in equation 4. We calculate the F1 score by balancing the importance of recall versus accuracy, as shown in equation 6.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1 - Score = \frac{(2 \times Precision \times Recall)}{(Precision + Recall)} \tag{6}$$

4. RESULTS AND DISCUSSION

This research employed various machine learning models, including three SVM models that utilized three different kernels (linear, polynomial two-degree, and RBF), logistic regression, and random forest. These models were fed with processed data, as presented in Table 1, to perform binary classification in identifying the presence of breast cancer. This research suggests that further study and exploration should focus on the SVM models. Meanwhile, the logistic regression was chosen as the base model, and random forest was chosen as the model comparison because it was the best model in previous research done by Valencia et al. [5], achieving an accuracy of 0.996. This research implemented all mentioned machine learning models using the Scikit-learn library. This research also split the dataset into a proportion of 75:25 for the train and test sets, respectively.

TABLE 2. Testing result evaluation using various machine learning models in accuracy, precision, recall, and F1-score.

Model	Overall		The Absence of Breast Cancer				The Presence of Breast Cancer			
	Accuracy	Support	Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
SVM with Linear Kernel	1.00	422	1.00	1.00	1.00	139	1.00	1.00	1.00	283
SVM with two-degree Polynomial Kernel	1.00	422	1.00	1.00	1.00	139	1.00	1.00	1.00	283
SVM with RBF Kernel	0.99	422	1.00	0.97	0.99	139	0.99	1.00	0.99	283
Logistic Regression	1.00	422	1.00	1.00	1.00	139	1.00	1.00	1.00	283
Random Forest	1.00	422	1.00	1.00	1.00	139	1.00	1.00	1.00	283

The result of testing using 25% of the dataset, called the test set, is illustrated in Table 2. The results indicate that our method, which involves preparing the data before using it in machine learning models, can greatly enhance the models' ability to distinguish between two classes: whether cancer is present or not. Almost all machine learning models can perfectly separate two different classes as measured in overall accuracy, precision, recall, and F1-score, except the SVM model with the RBF kernel.

In the research of health analytics, the presence of disease, for example, the presence of cancer as in this research, is necessary to be completely detected, measured by recall as a measurement metric. It is important and crucial for patients suffering from cancer to get treatment for their disease. Based on Table 2, all tested machine learning models achieved 100% completeness in revealing the patient as having cancer, and no patient with breast cancer was undetected for its

disease.

Based on Table 2, the support column in both classes means the amount of data of both the absence and the presence in the testing set. It appears that the support amount was an imbalanced number in both the absence and presence of cancer class. So for this research, it was suggested to use the F1 score because it measures the harmonic average of precision and recall [23]. To get more detailed data, this research showed the confusion matrix between all machine learning models in Table 3. Table 3 shows that all experiments in all models look excellent because all models never make mistakes in labelling the real cancer patient with a healthy label. Only the SVM model using the RBF kernel has four mistakes in labelling healthy people as having breast cancer. This is beneficial as it allows the four false positives to receive further treatment for their erroneous breast cancer diagnosis. However, we cannot allow the genuine breast cancer patients to remain undiagnosed or to receive a healthy label without further treatment.

Table 3. Testing result evaluation using various machine learning models in confusion matrix.

Model	**Confusion Matrix of Breast Cancer**			
	True Presence Cancer (True Positive)	**False Cancer (False Positive)**	**True Absence of Cancer (True Negative)**	**False Absence of Cancer (False Negative)**
SVM with Linear Kernel	283	0	139	0
SVM with two-degree Polynomial Kernel	283	0	139	0
SVM with RBF Kernel	283	4	135	0
Logistic Regression	283	0	139	0
Random Forest	283	0	139	0

For future work, this research may implement some data science techniques for handling imbalanced data like SMOTE (Synthetic Minority Over-sampling Technique). The development of deep learning for data prediction is also promising to be explored further so that this research may implement and explore the capability of a deep learning model for predicting breast cancer. This research hopes that it can help doctors identify and diagnose the presence of breast cancer using relevant health parameters.

5. CONCLUSION

This study was aimed to provide an insight into developing machine learning models for helping health professionals in diagnosing the presence of breast cancer using a dataset titled "Breast cancer risk factors in Cuban women." Our result showed a significant improvement from the original research that provided this dataset. Our methodology in handling and pre-processing the dataset until the machine learning model can learn the pattern in predicting the patient having breast cancer can achieve a perfect 100% in all used classification metrics used in this research for some models such as linear-SVM, polynomial-SVM, logistic regression, and RF. All our tested models were also pretty impressive in minimizing the risk of false negatives and never making a mistake in labelling a real patient with breast cancer with the wrong healthy label. For future work, this research may explore and investigate the capability of deep learning approaches and implement some data science techniques in handling imbalanced data for better improvement. This research also aims to influence and encourage health professionals to implement machine learning in the health field.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Karli Eka Setiawan: Conceptualization, Methodology, Software, Validation, Formal analysis, Data Curation, Visualization, Supervision, Project Administration, and Funding acquisition. **Hayyun Lisdiana**: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, writing – Original Draft, Writing - Review & Editing.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] M. Arnold, E. Morgan, H. Rumgay, et al. Current and Future Burden of Breast Cancer: Global Statistics for 2020 and 2040, Breast 66 (2022), 15-23. https://doi.org/10.1016/j.breast.2022.08.010.

[2] WHO, Breast Cancer, World Health Organization, Apr. 20, 2025. https://www.who.int/news-room/fact-sheets/detail/breast-cancer.

[3] A. Mashekova, Y. Zhao, E.Y. Ng, et al. Early Detection of the Breast Cancer Using Infrared Technology – A Comprehensive Review, Therm. Sci. Eng. Prog. 27 (2022), 101142. https://doi.org/10.1016/j.tsep.2021.101142.

[4] J. Valencia, E. Gutiérrez López, J.Á. González Fraga, H.A. Cantero Ronquillo, Breast Cancer Risk Factors in Cuban Women, Preprint, (2024). https://doi.org/10.17632/7JHDDNPZ2P.1.

[5] J.M. Valencia-Moreno, J.A. Gonzalez-Fraga, E. Gutierrez-Lopez, V. Estrada-Senti, H.A. Cantero-Ronquillo, V. Kober, Breast Cancer Risk Estimation with Intelligent Algorithms and Risk Factors for Cuban Women, Comput. Biol. Med. 179 (2024), 108818. https://doi.org/10.1016/j.compbiomed.2024.108818.

[6] J. Wu, C. Hicks, Breast Cancer Type Classification Using Machine Learning, J. Pers. Med. 11 (2021), 61. https://doi.org/10.3390/jpm11020061.

[7] R. Rabiei, Prediction of Breast Cancer Using Machine Learning Approaches, J. Biomed. Phys. Eng. 12 (2022), 297–308. https://doi.org/10.31661/jbpe.v0i0.2109-1403.

[8] M.A. Naji, S.E. Filali, K. Aarika, E.H. Benlahmar, R.A. Abdelouhahid, O. Debauche, Machine Learning Algorithms for Breast Cancer Prediction and Diagnosis, Procedia Comput. Sci. 191 (2021), 487-492. https://doi.org/10.1016/j.procs.2021.07.062.

[9] Y. Amethiya, P. Pipariya, S. Patel, M. Shah, Comparative Analysis of Breast Cancer Detection Using Machine Learning and Biosensors, Intell. Med. 2 (2022), 69-81. https://doi.org/10.1016/j.imed.2021.08.004.

[10] C. Cortes, V. Vapnik, Support-vector Networks, Mach. Learn. 20 (1995), 273-297. https://doi.org/10.1007/bf00994018.

[11] B. Gaye, D. Zhang, A. Wulamu, Improvement of Support Vector Machine Algorithm in Big Data Background, Math. Probl. Eng. 2021 (2021), 5594899. https://doi.org/10.1155/2021/5594899.

[12] E.Y. Boateng, J. Otoo, D.A. Abaye, Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A Review, J. Data Anal. Inf. Process. 08 (2020), 341-357. https://doi.org/10.4236/jdaip.2020.84020.

[13] H.A. Shiddiqi, K.E. Setiawan, R. Fredyan, Leveraging Support Vector Machines and Ensemble Learning for Early Diabetes Risk Assessment: A Comparative Study, Eng. Math. Comput. Sci. J. 7 (2025), 1-6. https://doi.org/10.21512/emacsjournal.v7i1.12846.

[14] H.A. Salman, A. Kalakech, A. Steiti, Random Forest Algorithm Overview, Babylon. J. Mach. Learn. 2024 (2024), 69-79. https://doi.org/10.58496/bjml/2024/007.

[15] K.E. Setiawan, A. Kurniawan, S.Y. Prasetyo, Comparative Analysis of Machine Learning Decision Tree-Based Models for Predicting Maternal Health Risks, Procedia Comput. Sci. 245 (2024), 57-64. https://doi.org/10.1016/j.procs.2024.10.229.

[16] G. Ngo, R. Beard, R. Chandra, Evolutionary Bagging for Ensemble Learning, Neurocomputing 510 (2022), 1-14. https://doi.org/10.1016/j.neucom.2022.08.055.

[17] S. González, S. García, J. Del Ser, L. Rokach, F. Herrera, A Practical Tutorial on Bagging and Boosting Based Ensembles for Machine Learning: Algorithms, Software Tools, Performance Study, Practical Perspectives and Opportunities, Inf. Fusion 64 (2020), 205-237. https://doi.org/10.1016/j.inffus.2020.07.007.

[18] K.E. Setiawan, Predicting Recurrence in Differentiated Thyroid Cancer: A Comparative Analysis of Various Machine Learning Models Including Ensemble Methods with Chi-Squared Feature Selection, Commun. Math. Biol. Neurosci. 2024 (2024), 55. https://doi.org/10.28919/cmbn/8506.

[19] P. Schober, T.R. Vetter, Logistic Regression in Medical Research, Anesth. Analg. 132 (2021), 365-366. https://doi.org/10.1213/ane.0000000000005247.

[20] S.Y. Prasetyo, A. Kurniawan, E.F.A. Sihotang, R. Puspita, K.E. Setiawan, Heart Disease Risk Prediction Using K-Nearest Neighbor: A Study of Euclidean and Cosine Distance Metrics, in: 2023 3rd International Conference on Smart Cities, Automation & Intelligent Computing Systems (ICON-SONICS), IEEE, Bali, Indonesia, 2023: pp. 236–240. https://doi.org/10.1109/icon-sonics59898.2023.10435299.

[21] M. Shirdel, M. Di Mauro, A. Liotta, Worthiness Benchmark: A Novel Concept for Analyzing Binary Classification Evaluation Metrics, Inf. Sci. 678 (2024), 120882. https://doi.org/10.1016/j.ins.2024.120882.

[22] Ž.Đ. Vujovic, Classification Model Evaluation Metrics, Int. J. Adv. Comput. Sci. Appl. 12 (2021), 599–606. https://doi.org/10.14569/ijacsa.2021.0120670.

[23] S. Riyanto, I.S. Sitanggang, T. Djatna, T.D. Atikah, Comparative Analysis Using Various Performance Metrics in Imbalanced Data for Multi-Class Text Classification, Int. J. Adv. Comput. Sci. Appl. 14 (2023), 1082-1090. https://doi.org/10.14569/ijacsa.2023.01406116.