Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 20 No. 2 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v20i02.42883

PAPER

Machine Learning System for the Effective Diagnosis and Survival Prediction of Breast Cancer Patients

Arturo Gago, Jean Marko Aguirre, Lenis Wong(✉)

Program Information Systems Engineering, Universidad Peruana de Ciencias Aplicadas, Lima, Perú

pcsilewo@upc.edu.pe

ABSTRACT

Breast cancer is one of the most significant global health challenges. Effective diagnosis and prognosis prediction are crucial for improving patient outcomes in the case of this disease. As machine learning (ML) has significantly improved prediction models in many disciplines, the goal of this study is to develop a ML system for medical specialists that can accurately predict tumor diagnosis and patient survival for breast cancer patients. For the training of diagnosis and survival prediction, five algorithmic models—decision tree (DT), random forest (RF), naive bayes (NB), support vector machines (SVMs), and gradient boosting—were trained with 569 records from the Breast Cancer Wisconsin dataset and 1,980 records from the Breast Cancer Gene Expression Profiles dataset. The results showed that the NB model exhibited better performance for tumor diagnosis, achieving an accuracy of 95.0%, while RF presented the best results for patient survival, with an accuracy of 76.0%. A survey of medical experts' experience with the resulting system showed high scores in reliability, performance, satisfaction, usability, and efficiency, confirming that ML systems have the potential to improve breast cancer patient outcomes.

KEYWORDS

breast cancer, diagnosis, treatment, machine learning (ML), random forest (RF), naive bayes (NB)

1 INTRODUCTION

Breast cancer is the fifth leading cause of death for women worldwide, with over two million cases recorded in 2020 [1]. Furthermore, according to the International Agency for Research on Cancer, there were 28 cases per 100,000 inhabitants, with a mortality rate of 8.5% [2]. Various factors contribute to breast cancer, with genetics, age, gender, or alcohol consumption being among the most significant [3]. Major symptoms can include significant breast inflammation, reddish skin tone, retraction of one or both nipples, constant breast pain, and, if left untreated, potentially

Gago, A., Aguirre, J.M., Wong, L. (2024). Machine Learning System for the Effective Diagnosis and Survival Prediction of Breast Cancer Patients. *International Journal of Online and Biomedical Engineering (iJOE)*, 20(2), pp. 95–113. https://doi.org/10.3991/ijoe.v20i02.42883

Article submitted 2023-07-07. Revision uploaded 2023-09-23. Final acceptance 2023-09-30.

leading to death [4]. The high number of cases and serious consequences continue to draw attention from the medical and research communities, prompting efforts to find effective solutions for detecting, treating, and improving the survival rates of cancer patients.

In any disease, previous cases often help inform current strategies, and computer analysis aims to expedite and streamline the analysis of these past cases. Studies have explored the use of machine learning (ML) to address critical interventions such as detection [5] and disease diagnosis [6], as well as to forecast patient survival rates. The effectiveness of any machine learning solution depends on the selection of appropriate algorithms and the implementation of the methods to ensure satisfactory results.

While the effectiveness of any ML study depends on selecting appropriate algorithms for the proposed solutions and outlining the methods used to ensure satisfactory results, many studies focus solely on aspects directly related to breast cancer, overlooking other important components, such as specific clinical factors. The aim of this study is to optimize tumor diagnosis prediction and patient survival prediction using ML. This will be achieved by comparing different algorithms and identifying the most effective ones that fit the system.

2 BACKGROUND

2.1 Machine learning

Machine learning is a technique that falls under the umbrella of artificial intelligence. Operating with various categories of algorithms, it enables the identification of patterns in large volumes of information, ultimately facilitating outcome prediction and enabling devices to perform tasks autonomously [7]. Often, these algorithms can be categorized as supervised, unsupervised, or reinforcement learning.

2.2 Supervised learning algorithms

Supervised learning algorithms are techniques in the field of artificial intelligence that enable machines to learn from labeled data sets. There are many models used for different situations, with the following having shown effectiveness, especially in medical prediction capacities [8].

Decision tree algorithm. The decision tree (DT) algorithm is a method based on a model that aims to identify potential outcomes by evaluating probabilities. This approach offers a more structured understanding of the problem and provides the opportunity to arrive at a logical solution [9].

Random forest algorithm. This algorithm enables the creation of regression and classification tasks. It is considered part of ensemble learning algorithms because it allows smaller or less powerful models to combine resources and become more [8].

Naive bayes algorithm. This algorithm is designed to differentiate between different objects based on specific characteristics. It is a probabilistic model used in classification tasks based on Bayes' theorem [10].

Support vector machine algorithms. This algorithm can achieve high accuracy on a large scale with minimal computational power. It operates by finding

the hyperplane within an N-dimensional space, where N represents the number of features, to effectively classify data points [11].

Gradient boosting algorithm. This algorithm can be used for both classification and regression tasks by combining various less powerful algorithms to achieve improved results. It works sequentially, aiming to create a structure that minimizes errors over multiple iterations [12].

3 RELATED WORKS

In the literature, numerous studies have been conducted on the topics of "risk factors" and "AI techniques" for detecting breast cancer. Three major risk factors for breast cancer have been identified, with race considered one of the most important indicators of risk for breast cancer [5], as well as being highly relevant for patients with breast cancer [13]. Age is also considered a high-risk factor, especially when considering the likelihood of successful treatment for diagnosed patients [5]. Interestingly, specific habits, such as night shift work, have been shown to significantly contribute to the development of the disease and are also considered risk factors [13].

Across a sample of the literature, twelve artificial intelligence algorithms were identified for the diagnosis and prognosis of breast cancer. [6] developed a system for diagnosing breast cancer using the SVM algorithm, achieving an accuracy of 99%, sensitivity of 98%, and specificity of 99%. In reference [14], the SV-NB-3-meta classifier algorithm achieved a diagnostic accuracy of 98.07%. [15] The study combined ML algorithms with artificial neural networks (ANN) for breast cancer diagnosis and prognosis. The best algorithm combination was found to be SVM + logistic regression (LR) + NB + DT, achieving an unsampled accuracy result of 97.67% and an up-sampled accuracy result of 98.83%. [16] presented a comparison of various ML classifiers, including DT, RF, LR, NB, K-nearest neighbors (KNN), and support vector machine (SVM), using the Wisconsin Prognosis Breast Cancer (WPB), Wisconsin Diagnosis Breast Cancer (WDB), and Wisconsin Breast Cancer and Mammographic Mass Dataset (WBM) datasets. The C-SVM algorithm with a radial basis function (RBF) kernel on the WDB dataset achieved the highest accuracy of 99.04%. In their study, [17] introduced a cloud-based framework for diagnosing breast cancer using the extreme learning machine (ELM) as the classifier. They achieved an accuracy of 98.68% with the WDB dataset. [18] aimed to enhance the accuracy of ML classification models for breast cancer prognosis by employing SVM, J48 (C4.5 DT algorithm), and multilayer perceptron (a feed-forward ANN) algorithms with the WDB dataset. The study found that the J48 algorithm achieved a Matthew's correlation coefficient (MCC) of 0.974, a sensitivity of 98.95%, a specificity of 98.58%, a Kappa statistic of 0.9735, and the highest accuracy of 98.83%. [19] aimed to predict breast cancer recurrence using ML algorithms with the medical records from King Abdullah University Hospital (KAUH). The study found that one algorithm achieved an accuracy of 90.14%. [20] utilized various classification methods, including NB, DT, LR, SVM, ANN, RF, and ML-based ontological models for breast cancer detection using the WDB dataset. The ML-based ontological model achieved the highest accuracy of 96.90%. [21] analyzed the use of ML algorithms to predict metastatic recurrence in early-stage breast cancer patients. They utilized data from patients at the Regional Oncology Center of Meknes with localized breast cancer and found that the SVM algorithm had the lowest error rate and achieved

the highest accuracy of 90.60%. [22] The researchers classified breast cancer tumors using numerical techniques based on image recognition software with the WDB dataset. They found that the ANN algorithm achieved the highest accuracy of 100%. [23] developed the stacked generalized ensemble (SGE) approach using the invasive ductal carcinoma dataset and compared it with other algorithms to classify invasive ductal carcinoma-positive and invasive ductal carcinoma-negative breast cancer. The best accuracy result of the SGE, using six learning models, was 87.80%. [24] proposed integrating multiple clinicopathological and genomic factors with dimensionality reduction using ML algorithms such as gradient boosting (GB), RF, SVM, and ANN with the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) database. The RF and SVM algorithms both achieved an accuracy of 72.0%. Finally, [25] conducted a study on the development of a breast cancer diagnosis system using genetic analysis with various ML algorithms. The study evaluated and compared the performance of classification models using the Area under the ROC Curve (AUC) metric. When using the XGBoost algorithm, the AUC was 0.76, and it achieved an accuracy of 77.0%.

4 PROPOSED MODEL

This study aims to establish an efficient machine learning system that provides accurate recommendations for breast cancer patients. Figure 1 depicts the workflow for training in both diagnosis prediction (DP) and survival prediction (SP).

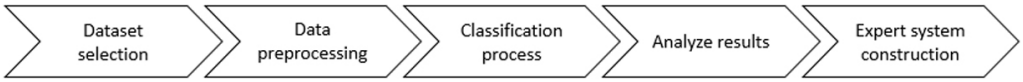

Fig. 1. Conceptual model of the proposed approach

After selecting the datasets, the data preprocessing phase involves performing data cleaning to eliminate inconsistencies in the selected datasets and defining training sets. ML algorithms are used for model training during the classification process. The results obtained predict the patient's diagnosis as either benign or malignant. Finally, in the last phase, the results are presented along with the corresponding metrics for each training session.

4.1 Dataset selection

For this study, patient data from two public datasets will be used: the WBD [26] and the breast cancer gene expression profiles from the METABRIC database [27], collected by [28]. Specifically, the sets will consist of (i) data from 569 instances of breast cancer tumors with 30 features [26] and (ii) data from 1980 clinical records of breast cancer patients composed of 31 features [28]. Table 1 displays some features for each dataset.

Table 1. Dataset 1 and dataset 2 features

Features Dataset 1	Features Dataset 2
radius	patient_id
texture	age_at_diagnosis
perimeter	type_of_breast_surgery
area	cancer_type
smoothness	cancer_type_detailed
compactness	cellularity
concavity	chemotherapy
concave points	pam50_+_claudin-low_subtype
symmetry	cohort
fractal dimension	er_status_measured_by_ihc
radius_mean	er_status
texture_mean	neoplasm_histologic_grade
perimeter_mean	her2_status
area_mean	hormone_therapy
smoothness_mean	mutation_count
concavity_mean	overall_survival_months
area_worst	overall_survival
concavity_worst	pr_status
smoothness_worst	tumor_stage

4.2 Data preprocessing

Data preprocessing techniques are crucial for ensuring data quality and improving the performance of machine learning models [29]. In the data preprocessing phase of each training, the following processes were carried out: data cleaning [30], feature selection, and variable encoding to a numerical representation using the "map" function in Python. For example, to encode the "ERStatus" data type, which initially has the values 'Negative' and 'Positive', the function is applied to convert them into numerical variables such as '1' and '0'.

4.3 Classification process

Figure 2 shows the training steps for predicting "diagnosis" in Python. Step 1 initiates model training by importing the function from "model_selection" that handles data splitting for testing and training. Next, variables are created: "X_train," "X_test," "y_train," and "y_test," to allocate 30% of the data for testing and the remainder for training. In step 2, the algorithm is invoked using the "ensemble" library. The variable "rfc" is created to manage training and predictions. "model5" utilizes "X_train" and "y_train" with the "fit" function to enable the model to learn from the data.

The variable “prediction5” is responsible for making predictions using the trained model, while the variable “cm5” is used to create the confusion matrix, which displays the correct and incorrect predictions for each output. In step 3, the code calculates the accuracy of the model. Finally, in step 4, the results are validated using real data. The variable “x_real_data” represents the authentic data and serves as the basis for the predicted outcome “prediction.”

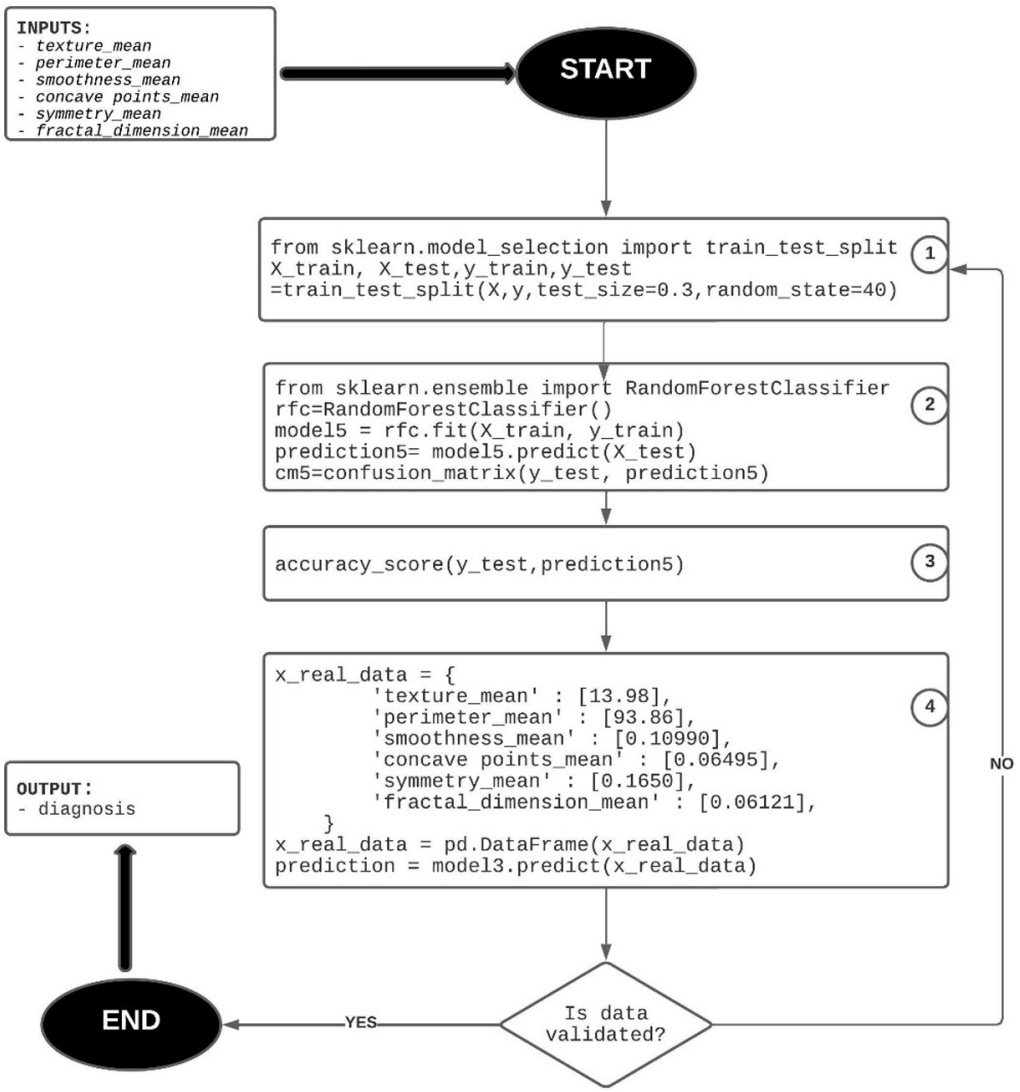

Fig. 2. Steps for training the random forest model in Python

This process is used for all algorithms. Additionally, the same process is followed for the training related to SP. For feature selection, a correlation matrix was generated to analyze and visualize the relationship of each column with the desired output. Jupyter is a tool that serves as a notebook for writing Python code and facilitates the management, processing, and training of data with machine learning models [31]. Figure 3 displays the Jupyter output, which offers a comprehensive overview of the essential model features. The correlation matrix enables analysis of the percentage of compatibility of each feature in the dataset with the “diagnosis” output.

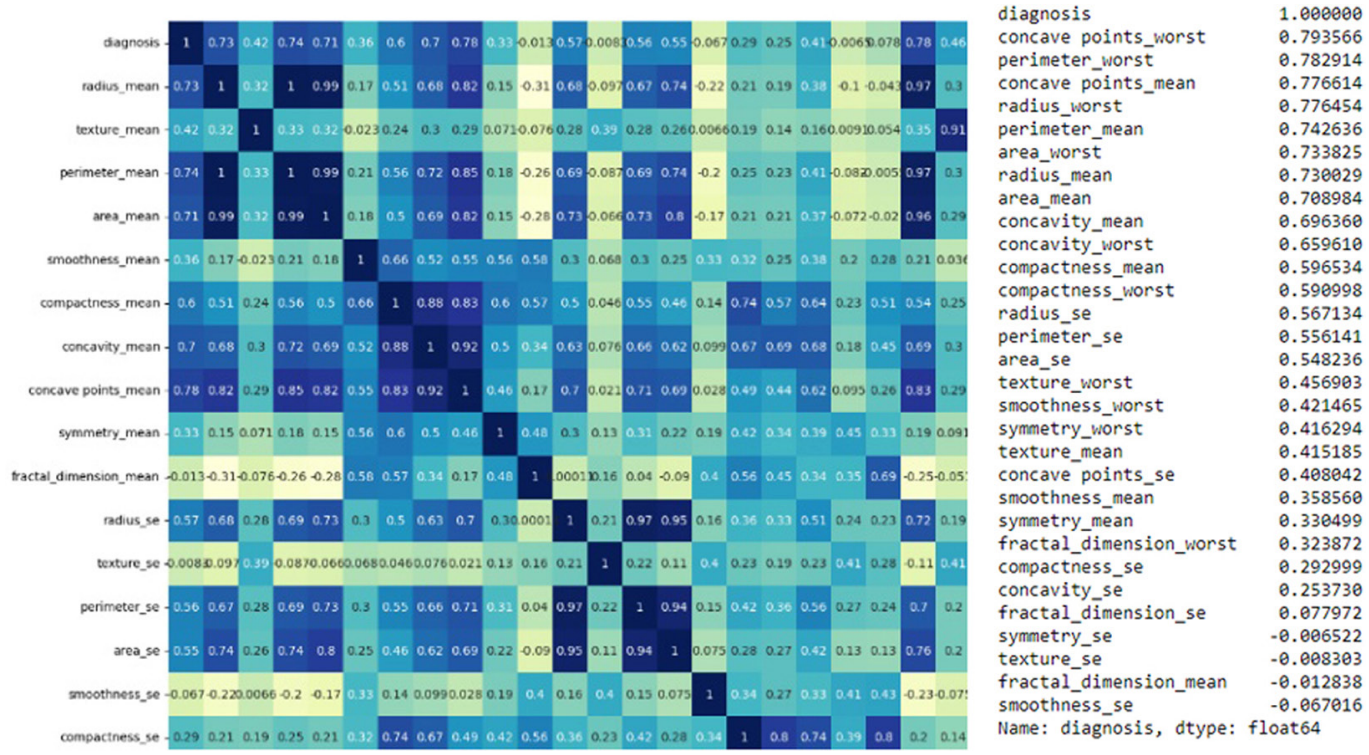

Fig. 3. Correlation matrix and WBC dataset features

Table 2 shows the six selected features from the WBC dataset for model training.

Table 2. Selected features for dataset 1

Features	Description
texture_mean	Standard deviation of grayscale values.
perimeter_mean	Mean tumor size.
smoothness_mean	Mean local variation of radius lengths.
concave points_mean	Mean number of concave portions of the contour.
symmetry_mean	Tumor symmetry
fractal_dimension_mean	Coastline approximation

For the second SP training, the same methodology was used for feature selection. A correlation matrix was constructed to analyze the compatibility of each feature with the "OverallSurvivalStatus" output, with the aim of achieving accurate prediction in the model. Table 3 shows the eleven selected features from the second dataset for model training.

Table 3. Selected features for dataset 2

Features	Description
AgeAtDiagnosis	Age of the patient at the date of diagnosis.
Cohort	Groups of people who share similar characteristics.
ERStatus	Cancer cells can be positive or negative for estrogen receptors.
NeoplasmHistologicGrade	The determination of whether cancer cells appear aggressive or not is made based on pathological observation of their nature.

(Continued)

Table 3. Selected features for dataset 2 *(Continued)*

Features	Description
LymphNodesExaminedPositive	During surgery, samples of the lymph nodes are taken and examined to determine if they are affected by cancer.
MutationCount	Number of genes with relevant mutations.
NottinghamPrognosticIndex	Prognosis after breast cancer surgery.
OncotreeCode	Diagnosis of cancer type from a clinical perspective assigning a unique OncoTree code to each diagnosis.
PRStatus	Cancer cells can be classified as positive or negative for progesterone receptors.
GeneClassifierSubtype	Three-gene subtype classifier.
TumorStage	Tumor status of the patient.

5 RESULTS AND DISCUSSION

5.1 Evaluation metrics

After the classification process, the results are compared to the original datasets to generate four variables: "True Positives" (TP), "False Positives" (FP), "False Negatives" (FN), and "True Negatives" (TN). These variables will, in turn, be used to generate metrics for evaluating the models (refer to Table 4).

Table 4. Description of variables for dataset 1 and dataset 2

Variable	Dataset 1	Dataset 2
TP	There is a correctly classified malignant tumor	The patient is correctly classified as a survivor
TN	The benign tumor is correctly classified	The patient's non-survival is correctly classified
FP	There is a malign tumor incorrectly classified	The patient survives incorrectly classified
FN	There is a benign tumor incorrectly classified	The patient's non-survival is incorrectly classified

These metrics are shown in equations (1), (2), (3), and (4).

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

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

$$F1\,score = \frac{2TP}{2TP + FP + FN} \tag{4}$$

"Precision" allows us to measure the quality of a positive prediction. "Recall" is a metric that enables us to measure the proportional value of the total most relevant requests that were correctly retrieved. "Accuracy" allows us to measure the value of predictions that were classified correctly or accurately. Lastly, the "F1 Score" is a metric that yields a value between 1 (perfect precision) and 0. It represents the harmonic mean of precision and recall for the trained instances.

5.2 Results of dataset 1 training

This study focused on gathering diverse metrics during the classification process, such as the confusion matrix, which offers a means to assess the performance of a classification problem and pinpoint error locations [32]. In Figure 4, the five confusion matrices for each algorithm are shown: DT (Figure 4a), NB (Figure 4b), Gradient Boosting (GB) (Figure 4c), RF (Figure 4d), and SVM (Figure 4e). Table 5 presents a summary of the correct and incorrect predictions for each trained model.

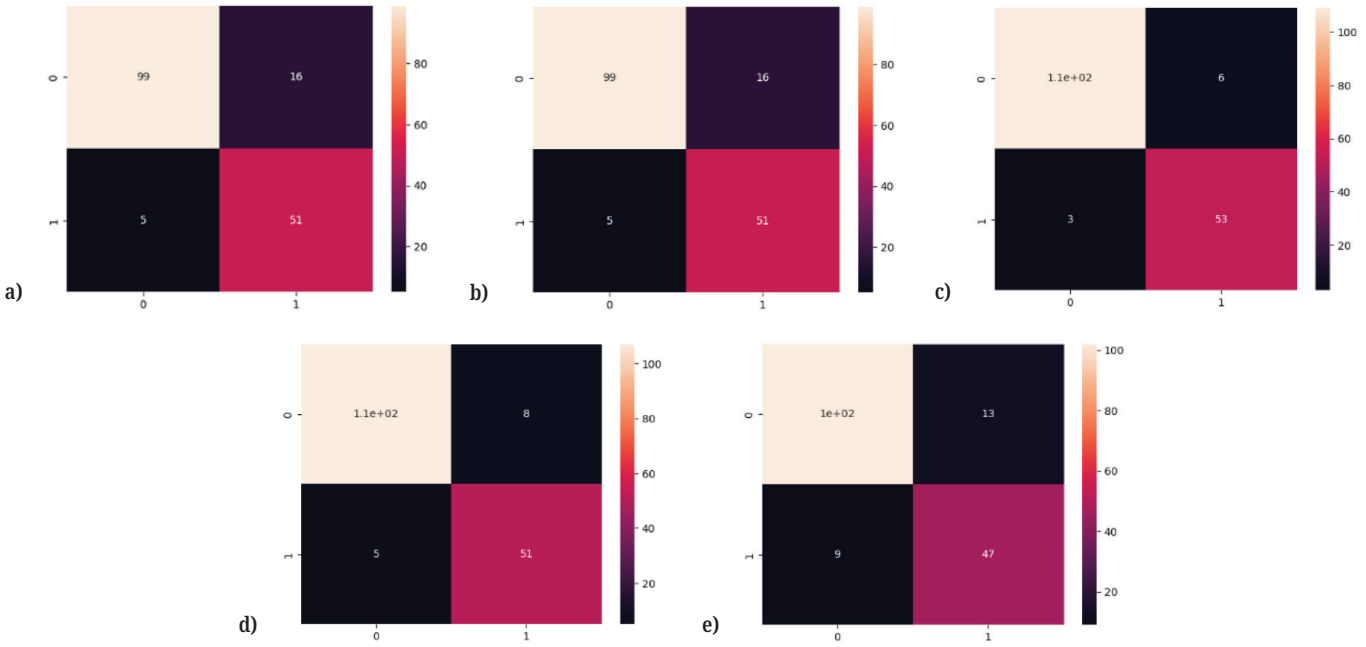

Fig. 4. Confusion matrices for DT (a), NB (b), GB (c), RF (d), and SVM (e) for dataset 1

Table 5. Metrics of the confusion matrix for dataset 1

Algorithm	Feature	Correct Predictions	Incorrect Predictions	Total
Decision tree	0 = 'Benign'	99	16	115
	1 = 'Malignant'	5	51	56
Naive Bayes	0 = 'Benign'	99	16	115
	1 = 'Malignant'	5	51	56
Gradient Boosting	0 = 'Benign'	109	6	115
	1 = 'Malignant'	3	53	56
Random Forest	0 = 'Benign'	107	8	115
	1 = 'Malignant'	5	51	56
SVM	0 = 'Benign'	102	13	115
	1 = 'Malignant'	9	47	56

The area under the curve (AUC) metric (see Figure 5) evaluates the classification performance of the model using the receiver operating characteristic (ROC) curve.

A higher AUC value, closer to 1, indicates a more effective classifier. Results show that the NB algorithm was the most effective classifier with an AUC value of 0.99 (see Figure 5b), followed by GB, RF, and finally SVM with AUC values of 0.98, 0.98, and 0.95, respectively. On the other hand, it is evident that the DT algorithm achieved the lowest AUC value with 0.89 (see Figure 5a).

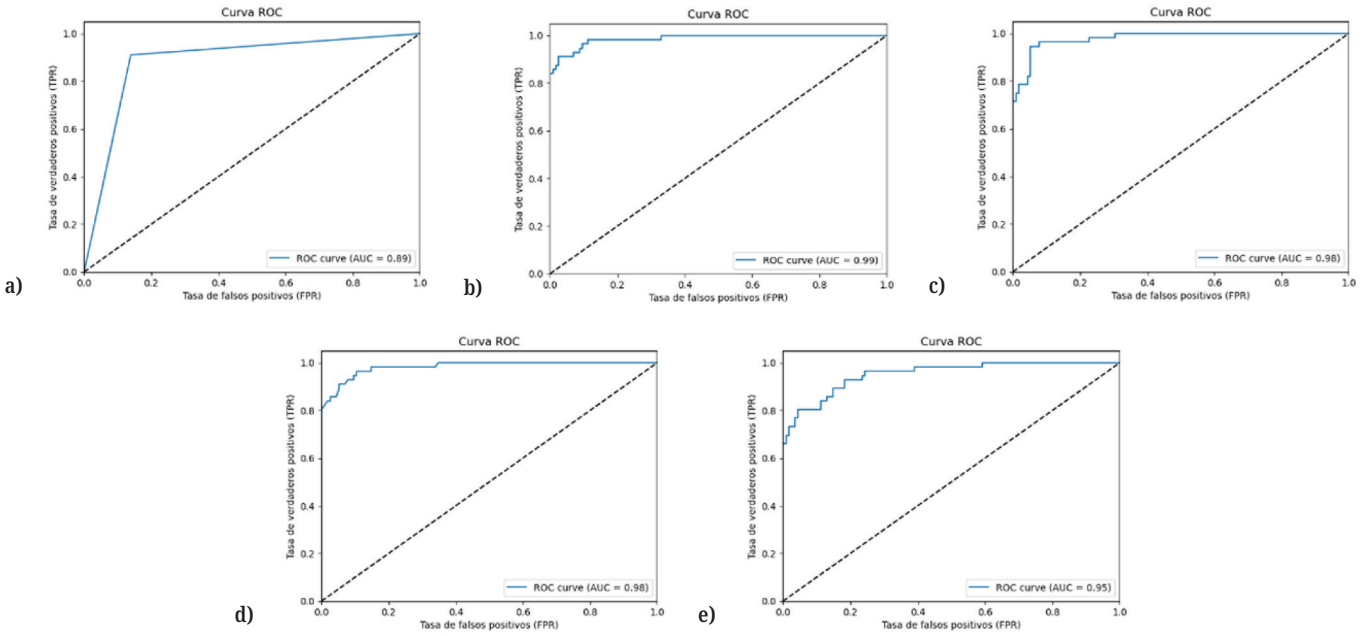

Fig. 5. ROC curves of the DT algorithm (a), NB (b), GB (c), RF (d), and SVM (e) for the dataset 1

During the initial training of the DP, tests were carried out using various algorithms to assess their respective evaluation metrics (see Table 6). Based on the confusion matrix, it was determined that the proposed model achieved the highest precision of 95% with NB, making it the preferred choice for diagnosing breast cancer within the proposed system.

Table 6. Results of metrics from the first training for dataset 1

Algorithm	Feature	Precision	Recall	F1 Score	Accuracy
Decision tree	0 = 'Benign'	0.95	0.86	0.90	0.88
	1 = 'Malignant'	0.76	0.91	0.83	
Random Forest	0 = 'Benign'	0.96	0.93	0.95	0.93
	1 = 'Malignant'	0.87	0.93	0.90	
Naive Bayes	0 = 'Benign'	0.96	0.97	0.97	0.95
	1 = 'Malignant'	0.94	0.91	0.93	
SVM	0 = 'Benign'	0.92	0.89	0.90	0.87
	1 = 'Malignant'	0.78	0.84	0.81	
Gradient Boosting	0 = 'Benign'	0.97	0.95	0.96	0.95
	1 = 'Malignant'	0.90	0.95	0.92	

Most studies focus on diagnosing by considering the majority of the features from the WBC dataset [26]. Additionally, they also considered data such as "standard

error" and "worst" for each index, which has been shown to improve accuracy in the studies [14] [15] [17]. This study suggests using "mean" fields because they offer the average value of the data, providing more concise information to optimize the accuracy of inputs in constructing the expert system for our research.

5.3 Results of dataset 2 training

Just as in the dynamic programming (DP) training, the same algorithms were used for the stochastic programming (SP) training. Figure 6 displays the confusion matrix for each of the trained models, and Table 7 presents a summary of the metrics for each trained model.

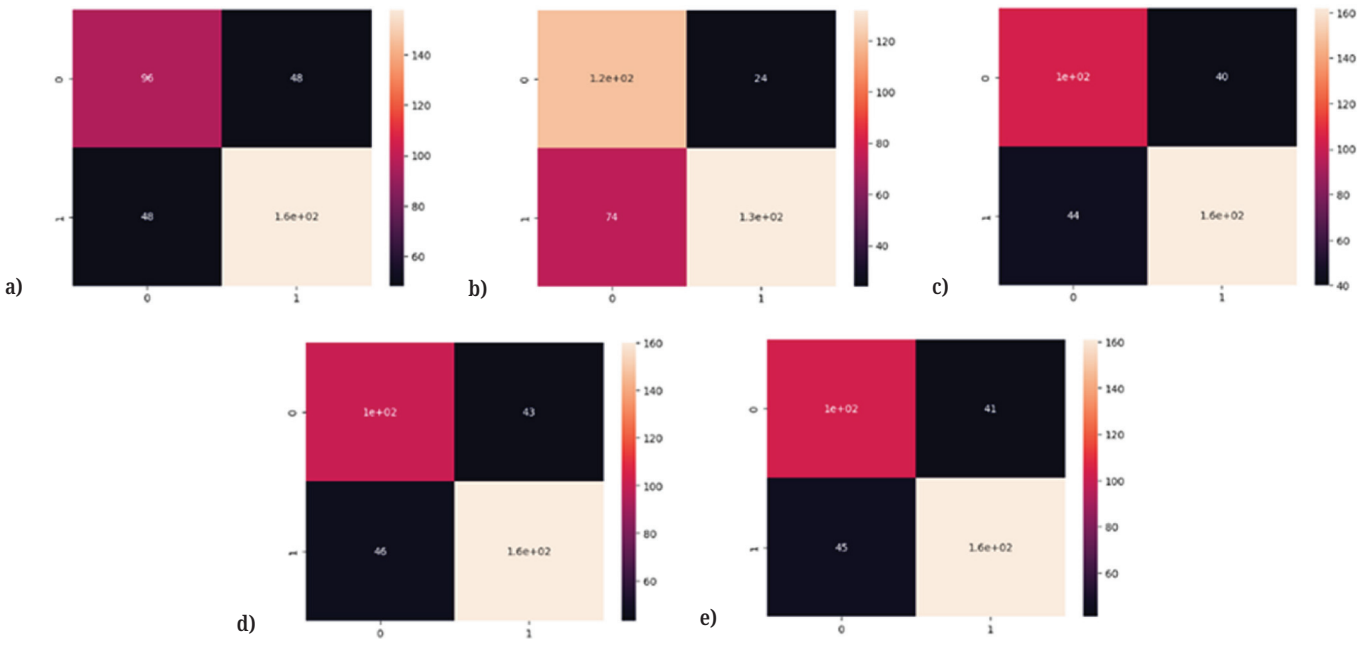

Fig. 6. Confusion matrices for DT (a), NB (b), GB (c), RF (d), and SVM (e) for dataset 2

Table 7. Metrics of the confusion matrix for dataset 2

Algorithm	Feature	Correct Predictions	Incorrect Predictions	Total
Decision tree	0 = 'LIVING'	96	48	144
	1 = 'DECEASED'	48	158	206
Naive Bayes	0 = 'LIVING'	120	24	144
	1 = 'DECEASED'	74	132	206
Gradient Boosting	0 = 'LIVING'	104	40	144
	1 = 'DECEASED'	44	162	206
Random Forest	0 = 'LIVING'	101	43	144
	1 = 'DECEASED'	46	160	206
SVM	0 = 'LIVING'	103	41	144
	1 = 'DECEASED'	45	161	206

In Figure 6, the ROC curves display the AUC metrics for the SP training. The GB and SVM algorithms (see Figures 7c and 7e) exhibit the highest AUC value of 0.84. However, the second-highest value, although slightly lower, is achieved by the RF algorithm (see Figure 7d) with a score of 0.83, indicating that the models can effectively differentiate between positive and negative classes. Results show that the DT model (see Figure 7a) and the NB model (see Figure 7b) displayed the lowest values, indicating that they may not be effective classifiers for stochastic programming.

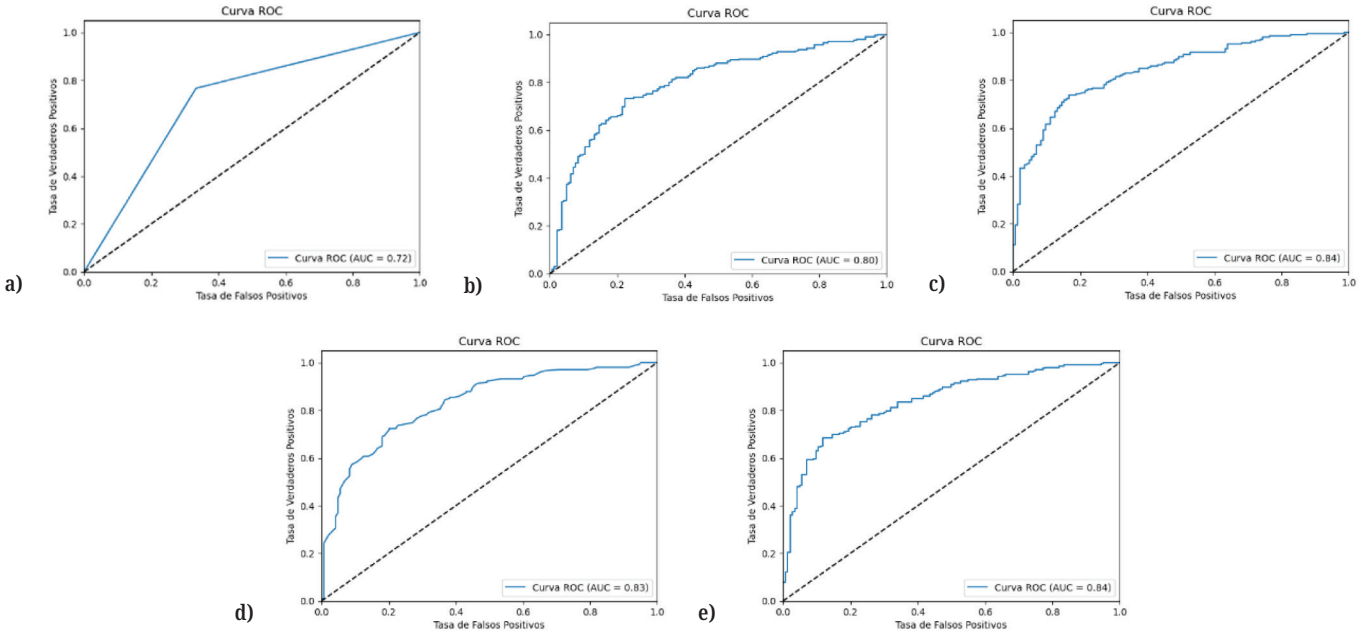

Fig. 7. ROC curves of the DT algorithm (a), NB algorithm (b), GB algorithm (c), RF algorithm (d), and SVM algorithm (e) for the dataset 2

During the SP training, a model was developed using dataset 2. The model was trained with five different algorithms to identify the one that would yield the highest prediction accuracy based on the input values. Table 8 displays the results of the metrics for each algorithm. RF and GB achieved the best metrics, with 76% accuracy.

Table 8. Metrics result from the second training for dataset 2

Algorithm	Feature	Precision	Recall	F1 Score	Accuracy
Decision tree	0 = 'LIVING'	0.67	0.67	0.67	0.73
	1 = 'DECEASED'	0.77	0.77	0.77	
Random Forest	0 = 'LIVING'	0.71	0.71	0.71	0.76
	1 = 'DECEASED'	0.80	0.80	0.80	
Naive Bayes	0 = 'LIVING'	0.62	0.83	0.71	0.72
	1 = 'DECEASED'	0.85	0.64	0.73	
SVM	0 = 'LIVING'	0.70	0.72	0.71	0.75
	1 = 'DECEASED'	0.80	0.78	0.78	
Gradient Boosting	0 = 'LIVING'	0.70	0.72	0.71	0.76
	1 = 'DECEASED'	0.80	0.79	0.79	

As a fundamental part of the study, the SP training considered two resulting values: '0' indicating that the patient survives breast cancer, and '1' indicating that the patient will not survive. This information is crucial for our system's development. Thanks to dataset 2, we were able to develop a function that identifies patients with a higher survival rate based on the number of months lived by the patients and the variable indicating their status as alive or deceased. Therefore, based on the METABRIC database [27], [24] achieved an accuracy of less than 75% by selecting data for all fields complete and considering HER2 as a negative prognostic factor for their research. In this study, we considered specific and relevant clinical data, including the treatments administered to each patient, in order to develop a system capable of diagnosing and recommending treatments to medical specialists.

5.4 Expert system construction

For the development of the expert system, we utilized the models that demonstrated the best performance. In the case of tumor DP, we employed the NB algorithm. The RF algorithm was selected for patient SP to gain insight into treatments such as radiotherapy, chemotherapy, and hormonal therapy. In Figure 8, all the components for the system development are presented, including the Flutter mobile framework, AWS services, and PostgreSQL as the database. It includes the security layer, database, back-end, and the ML model training module.

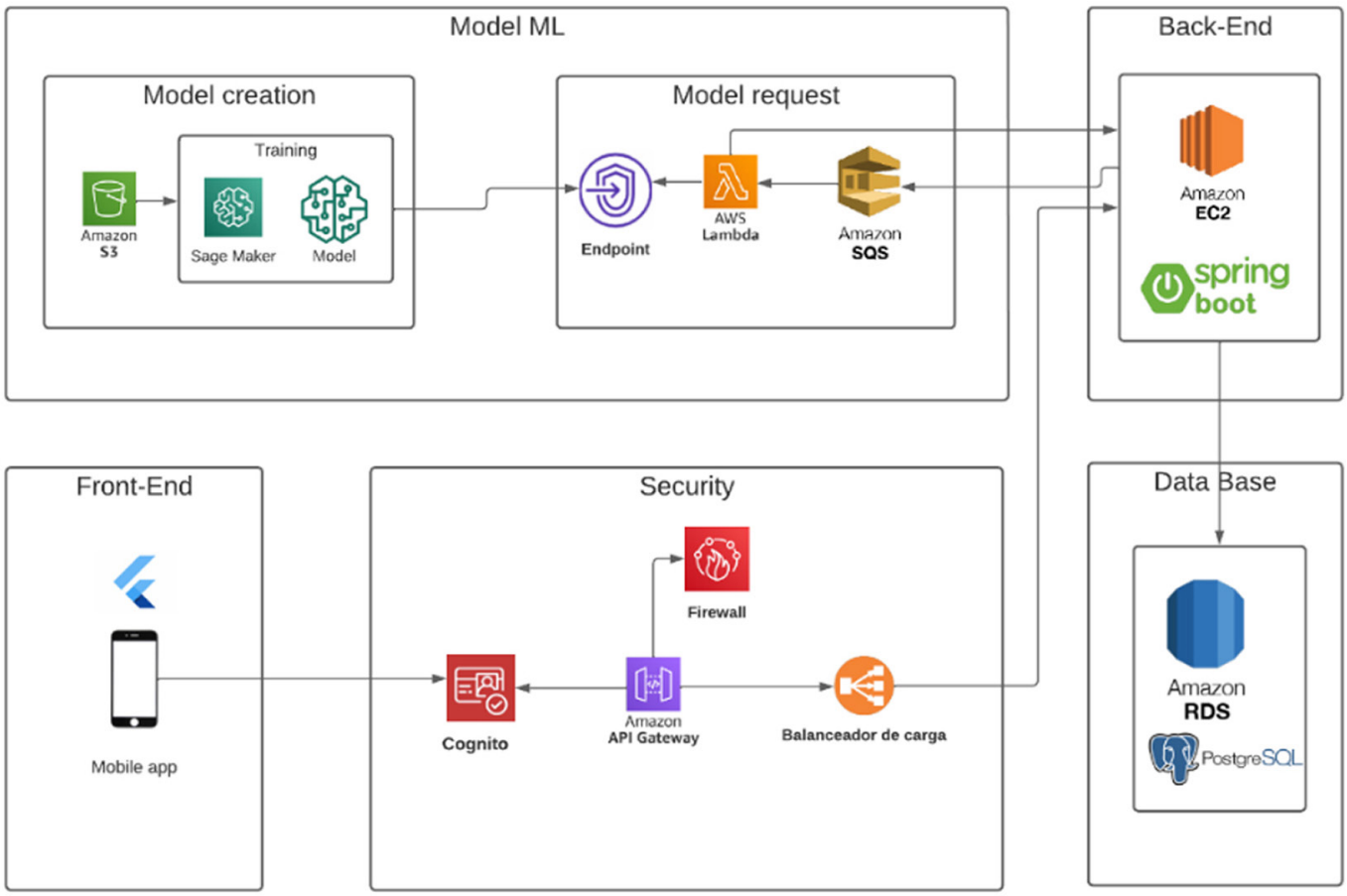

Fig. 8. Physical architecture of the proposed system

The expert system was validated by a group of 18 medical specialists who utilized the system demonstration and provided their feedback.

The validation process involved the following steps: (1) an in-person explanation of the system; (2) experts using the system on a mobile device; and (3) the creation of a survey. This experiment was conducted individually by each expert and lasted approximately 40 minutes.

For the system simulation, the following steps were followed:

- Experts were required to create an account to access the system.
- A medical institution and a test patient were added for the expert user.
- The main functionalities of the system (tumor diagnosis and treatment recommendation) were simulated using the test data provided.

The survey was conducted according to the criteria outlined in the ISO/IEC 25000 standard [33], which encompass functionality, performance, usability, reliability, efficiency, and maintainability. This study was conducted using Google Forms, which included five closed-ended questions. A Likert scale was applied to 2 of the questions (0 = very bad, 1 = bad, 2 = normal, 3 = good, 4 = very good), and 1 open-ended question was included. Table 9 presents the survey questions, question types, and their categorization according to quality characteristics.

Table 9. Survey developed for experts

ID	Question	Class	Features
Q1	After using HealthApp, what was the level of response from the application to the query made?	Closed (Multiple-choice)	Performance
Q2	Did you find the application very easy to use?	Closed (dichotomous)	Usability
Q3	How would you evaluate your satisfaction with the HealthApp application?	Closed (Mul-tiple-choice)	Reliability
Q4	Would you recommend this application?	Closed (dichotomous)	
Q5	Do you think HealthApp helps specialists make decisions regarding treatment recommendations for breast cancer patients?	Closed (dichotomous)	Efficiency
Q6	From your experience, what would you improve in the HealthApp application?	Open	Maintainability

Regarding the “performance” (Q1) and “reliability” (Q3) of the system, an average rating of 3.0 was obtained. This indicates that the application has a “good” level of response to queries (see Figure 9), and the doctors are satisfied with its use (see Figure 10). Additionally, 100% of the doctors indicated that they would recommend the system (see Figure 11b).

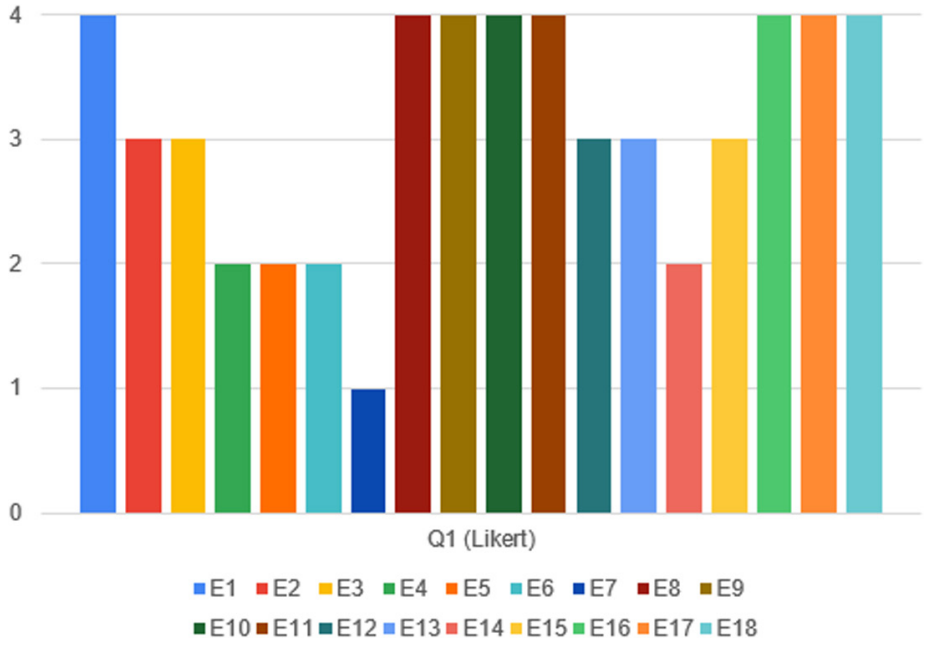

Fig. 9. Results of the questionnaire "Performance" (Q1)

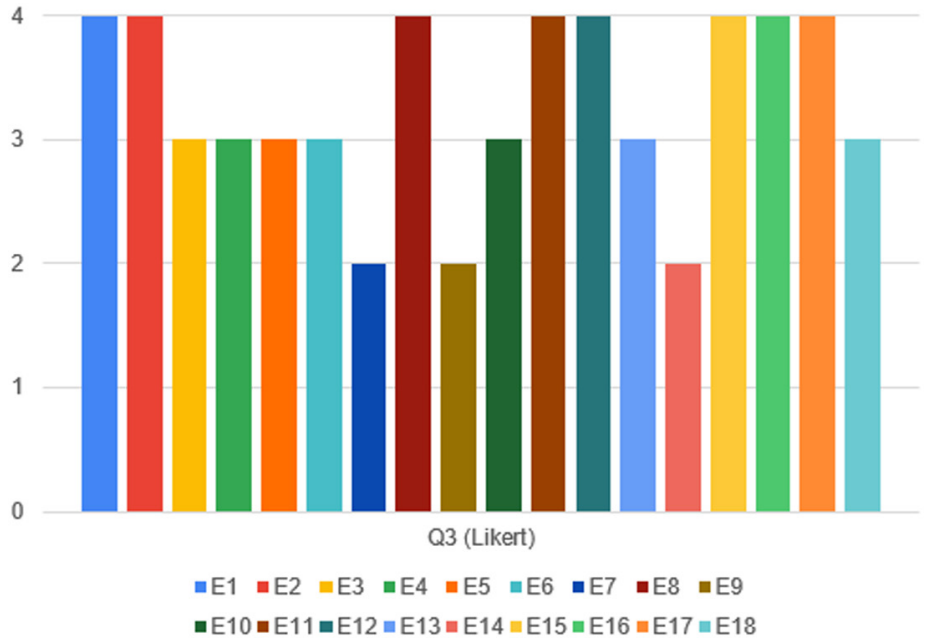

Fig. 10. Results of the questionnaire "Reliability" (Q3)

Regarding the "usability" (Q2) of the system, it was found that 89% of the doctors indicated that the application is easy to use (see Figure 11a). As for "efficiency" (Q5), 100% of the doctors stated that the proposed system would greatly help in making decisions regarding treatment recommendations for their patients (see Figure 11c). Lastly, concerning "maintainability" (Q6), the doctors offered their recommendations for system improvements, including modifying the UI, optimizing the flow of functionalities, and adding more features to enhance the diagnosis.

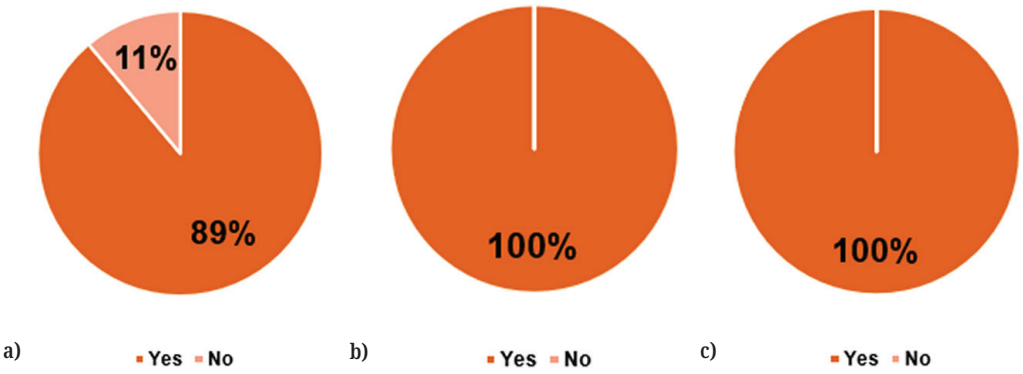

Fig. 11. Results of the questionnaire Q2 (a), Q4 (b), Q5 (c)

6 CONCLUSIONS

In this study, we propose to develop a system for medical experts to predict breast cancer tumor diagnosis and patient survival using machine learning algorithms. Two datasets were created using the publicly available Wisconsin Breast Cancer Dataset Diagnostic and the Breast Cancer Gene Expression Profiles from the METABRIC database. These datasets were trained using five different ML algorithms: DT, NB, GB, RF, and SVM. Each training process involved tuning the hyperparameters to achieve optimal results. Evaluation of each model was conducted using metrics such as precision, recall, accuracy, and F1-score. The NB model achieved the highest precision and a balanced performance in terms of both precision and recall for predicting tumor diagnosis. The RF model showed higher precision in predicting patient survival. A system utilizing the NB model for tumor diagnosis and the RF model for patient survival was developed and validated through a survey conducted among medical specialists. The survey aimed to gather their perceptions of the system's characteristics, including performance, usability, reliability, efficiency, and maintainability.

7 ACKNOWLEDGMENT

We express our gratitude to the medical specialists who supported us and provided valuable technical guidance and information for this research.

8 REFERENCES

[1] Cancer.Net, "Cáncer de mama: Estadísticas | Cancer.Net." [Online]. Available: https://www.cancer.net/es/tipos-de-c%C3%A1ncer/c%C3%A1ncer-de-mama/estad%C3%ADsticas

[2] J. Ferlay *et al.*, "Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods," *Int. J. Cancer*, vol. 144, no. 8, pp. 1941–1953, 2019. https://doi.org/10.1002/ijc.31937

[3] Oncosalud, "Síntomas de cáncer de mama: ¿cómo detectarlos? | Blog Oncosalud," [Online]. Available: https://blog.oncosalud.pe/cancer-de-mama-sintomas-y-causas.

[4] MayoClinic, "Cáncer de mama – Síntomas y causas – Mayo Clinic," [Online]. Available: https://www.mayoclinic.org/es-es/diseases-conditions/breast-cancer/symptoms-causes/syc-20352470.

[5] X. Jiang, A. Wells, A. Brufsky, D. Shetty, K. Shajihan, and R. E. Neapolitan, "Leveraging Bayesian networks and information theory to learn risk factors for breast cancer metastasis," *BMC Bioinformatics*, vol. 21, no. 1, pp. 1–17, 2020. https://doi.org/10.1186/s12859-020-03638-8

[6] M. H. Memon, J. P. Li, A. U. Haq, M. H. Memon, W. Zhou, and R. Lacuesta, "Breast cancer detection in the IOT health environment using modified recursive feature selection," *Wirel. Commun. Mob. Comput.*, vol. 2019, 2019. https://doi.org/10.1155/2019/5176705

[7] P. P. Torralba, "¿Qué es el machine learning? Aprendizaje supervisado vs no supervisado," *Thinking for Innovation*, 2023. [Online]. Available: https://www.iebschool.com/blog/que-machine-learning-big-data/.

[8] Gonzalez L, "Aprendizaje supervisado: Random forest classification," Aprende IA. [Online]. Available: https://aprendeia.com/aprendizaje-supervisado-random-forest-classification/.

[9] Raona, "Los 10 algoritmos esenciales en machine learning," [Online]. Available: https://www.raona.com/los-10-algoritmos-esenciales-machine-learning/.

[10] R. Gandhi, "Naive bayes classifier. What is a classifier?" *Towards Data Science*. [Online]. Available: https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c.

[11] R. Gandhi, "Support vector machine – introduction to machine learning algorithms," *Towards Data Science*. [Online]. Available: https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47.

[12] Amat Joaquín, "Gradient boosting con Python," cienciadedatos.net. [Online]. Available: https://www.cienciadedatos.net/documentos/py09_gradient_boosting_python.html.

[13] M. Szkiela, E. Kusideł, T. Makowiec-Dabrowska, and D. Kaleta, "Night shift work—a risk factor for breast cancer," *International Journal of Environmental Research and Public Health 2020,* vol. 17, no. 2, p. 659, 2020. https://doi.org/10.3390/ijerph17020659

[14] M. Abdar *et al.*, "A new nested ensemble technique for automated diagnosis of breast cancer," *Pattern Recognit. Lett.*, vol. 132, pp. 123–131, 2020. https://doi.org/10.1016/j.patrec.2018.11.004

[15] U. Naseem *et al.*, "An automatic detection of breast cancer diagnosis and prognosis based on machine learning using ensemble of classifiers," *IEEE Access*, vol. 10, pp. 78242–78252, 2022. https://doi.org/10.1109/ACCESS.2022.3174599

[16] S. A. El_Rahman, "Predicting breast cancer survivability based on machine learning and features selection algorithms: A comparative study," *J. Ambient. Intell. Humaniz. Comput.*, vol. 12, no. 8, pp. 8585–8623, 2021. https://doi.org/10.1007/s12652-020-02590-y

[17] V. Lahoura *et al.*, "Cloud computing-based framework for breast cancer diagnosis using extreme learning machine," *Diagnostics 2021,* vol. 11, no. 2, p. 241, 2021. https://doi.org/10.3390/diagnostics11020241

[18] Y. S. Solanki *et al.*, "A hybrid supervised machine learning classifier system for breast cancer prognosis using feature selection and data imbalance handling approaches," *Electronics 2021,* vol. 10, no. 6, p. 699, 2021. https://doi.org/10.3390/electronics10060699

[19] A. Alzu'bi, H. Najadat, W. Doulat, O. Al-Shari, and L. Zhou, "Predicting the recurrence of breast cancer using machine learning algorithms," *Multimed. Tools Appl.*, vol. 80, no. 9, pp. 13787–13800, 2021. https://doi.org/10.1007/s11042-020-10448-w

[20] H. El Massari, N. Gherabi, S. Mhammedi, H. Ghandi, F. Qanouni, and M. Bahaj, "An ontological model based on machine learning for predicting breast cancer," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 7, pp. 108–115, 2022. https://doi.org/10.14569/IJACSA.2022.0130715

[21] E. Merouane, A. Said, and E. F. Noureddine, "Prediction of metastatic relapse in breast cancer using machine learning classifiers," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 2, pp. 176–181, 2022. https://doi.org/10.14569/IJACSA.2022.0130222

[22] F. Zahedi and M. K. Moridani, "Classification of breast cancer tumors using mammography images processing based on machine learning," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 18, no. 5, pp. 31–42, 2022. https://doi.org/10.3991/ijoe.v18i05.29197

[23] D. Kumar and U. Batra, "Classification of invasive ductal carcinoma from histopathology breast cancer images using stacked generalized ensemble," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 3, pp. 4919–4934, 2021. https://doi.org/10.3233/JIFS-201702

[24] M. Zhao, Y. Tang, H. Kim, and K. Hasegawa, "Machine learning with K-means dimensional reduction for predicting survival outcomes in patients with breast cancer," *Cancer Inform*, vol. 17, 2018. https://doi.org/10.1177/1176935118810215

[25] D. Santos, "Breast cancer survival prediction using machine learning and gene expression profiles," *medRxiv*, 2022. https://doi.org/10.1101/2022.01.22.22269470

[26] W. Wolberg, O. Mangasarian, N. Street, and W. Street, "Breast cancer wisconsin (Diagnostic)," *UCI Machine Learning Repository*, 1995. https://doi.org/10.24432/C5DW2B

[27] "Breast cancer dataset (METABRIC, nature 2012 & nat commun 2016) in CBioPortal increasing the resolution on breast cancer – the METABRIC study," [Online]. Available: https://www.cbioportal.org/study/summary?id=brca_metabric

[28] B. Pereira *et al.*, "The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes," *Nature Communications 2016*, vol. 7, no. 1, pp. 1–16, 2016. https://doi.org/10.1038/ncomms11479

[29] A. Géron, "Hands-on machine learning with scikit-learn, keras, and tensorflow," *O'Reilly Media, Inc.*, 2019. [Online]. Available: https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/.

[30] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, "Multivariate data analysis," *Hampshire, UK : Cengage Learning, EMEA*, vol. 8th Ed, p. 834, 2019. [Online]. Available: https://www.worldcat.org/title/1082318713.

[31] AWS, "Set up a Jupyter notebook server – deep learning AMI," [Online]. Available: https://docs.aws.amazon.com/dlami/latest/devguide/setup-jupyter.html. [Accessed: Jun. 12, 2023].

[32] S. Pandey, B. Schlomann, K. Fender, and R. Reynolds-Haertle, "Resultados de los modelos de aprendizaje automático," *Finance | Dynamics 365* | Microsoft Learn. [Online]. Available: https://learn.microsoft.com/es-es/dynamics365/finance/finance-insights/confusion-matrix.

[33] P. A. Roa, C. Morales, and P. Gutiérrez, "Norma ISO/IEC 25000," *Tecnología Investigación y Academia*, vol. 3, no. 2, pp. 27–33, 2015. [Online]. Available: https://revistas.udistrital.edu.co/index.php/tia/article/view/8373. [Accessed: Jun. 11, 2023].

9 AUTHORS

Arturo Gago is a graduate in Information Systems Engineering from Peruvian University of Applied Sciences in 2023. He is working as a Software Quality Assurance Engineer at a Cloud Solutions Consultancy. He has knowledge about software development, test automation, business processes, and project management using agile frameworks (E-mail: u201724018@upc.edu.pe).

Jean Marko Aguirre is a Mobile developer with a degree in Information Systems Engineering from Peruvian University of Applied Sciences in 2023. Currently specializing in mobile development with a focus on Flutter and IoT (Internet of Things) and possessing a strong background in cloud services, automation, and project management (E-mail: u201819426@upc.edu.pe).

Lenis Wong is a Professor and Researcher of Software Engineering and Information Systems Engineering at the National University of San Marcos and Peruvian University of Applied Sciences, Peru. She holds a PhD in Systems Engineering and Computer Science. MSc. in Systems and Computer Engineering with mention in Software Engineering. She is a member of the AI research group and has carried out different multidisciplinary projects where Artificial Intelligence, Software Engineering and Information Systems Engineering have been applied in education, health, medicine and healthcare. She has published several international peer-reviewed scientific articles in different multidisciplinary areas such as: ML, DL, IoT, e-Health, Software Engineering, Requirements Engineering, Cloud Computing, E-Learning, Gamification, Cyberattacks, Natural Language Processing, Networks and Blockchain Technologies (E-mail: pcsilewo@upc.edu.pe).