Vol. 8, No. 3, June 2026

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Development and Evaluation of a Multi-Algorithm Application for Predicting Breast Cancer Patient Survival

Zulkifli[1], Kraugusteeliana [2], Sukarni[3], Ikna Awaliyani[4], Nur Asini[5], Fitriana[6]

[1]Department of Informatics Engineering, Faculty of Technology and Informatics, Aisyah University, Indonesia

[2]Information System Department, Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

[3,6]Department of Midwifery, Faculty of Health, Aisyah University, Indonesia

[4]Department of Information Technology Education, Faculty of Teacher Training and Education, Aisyah University, Indonesia

[5]Department of Medicine, Faculty of Medicine, Aisyah University, Indonesia

Received:
October 15, 2025
Revised:
May 17, 2026
Accepted:
June 1, 2026
Published:
June 22, 2026

Corresponding Author:
Author Name*:
Zulkifli
Email*:
zulkifli@aisyahuniversity.ac.id

DOI:
10.63158/journalisi.v8i3.1665

Abstract. This study developed a multi-algorithm machine learning prototype for multiclass breast cancer survival prediction using 1,980 patient records, classifying patients as Living, Died of Disease, or Died of Other Causes. The framework integrated NN, SVM, RF, NB, and KNN algorithms within a decision-support monitoring application, with preprocessing steps including data cleaning, normalization, feature preparation, and dataset partitioning. To prevent target leakage, survival-related variables were excluded from the predictor set. The revised evaluation results indicated that NB and KNN delivered the strongest performance, achieving weighted average F1-scores of 0.93 and 0.92, respectively, while NN and RF showed comparatively lower results. These findings highlight the potential of machine learning for breast cancer survival status monitoring, although the proposed system is designed as a decision-support prototype rather than a clinical diagnostic tool. Therefore, before actual healthcare deployment, more research incorporating explainable AI techniques, external validation, and real-world clinical testing is needed.

Keywords: Machine Learning, Survival Prediction, Decision Support System, Breast Cancer, Multi-class Classification

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

1. INTRODUCTION

One of the most prevalent diseases in the world, breast cancer continues to be a serious public health issue, especially for women [1]. Uncontrolled cell proliferation in breast tissue is the disease's hallmark, and it has a high risk of spreading to organs like the brain, liver, lungs, and bones [2]. With over 2.3 million new cases and 685,000 related deaths recorded globally in 2020, breast cancer overtook lung cancer as the most common cancer [3]. Early identification and prompt treatment are crucial to improving patient outcomes and lowering mortality rates because of its high incidence and variety of clinical presentations [4], [5].

Lifestyle factors like obesity and alcohol use, particularly after menopause, are the primary causes of this malignancy and raise the risk of breast cancer [6]. Reducing breast cancer mortality rates requires prompt screening procedures [7]. Furthermore, breast cancer still poses a serious threat to effective treatment and prevention due to its complicated origin and wide range of clinical manifestations. In order to create successful treatment plans, it is essential to understand the complex nature of breast cancer [8].

Different clinical and technological goals are represented by patient monitoring, survival prediction, and breast cancer diagnosis. The goal of breast cancer diagnosis is to determine the kind or presence of the disease via imaging, histological investigation, or medical tests. On the other hand, survival prediction uses machine learning algorithms to estimate patient outcomes or survival status based on clinical and demographic factors. In the meantime, the study's suggested monitoring application is a decision-support [9] prototype that combines many machine learning algorithms to assess and forecast patient survival status on a regular basis using patient data from the past. Therefore, the creation of a multi-algorithm application for survival prediction and patient monitoring in clinical support settings is the main contribution of this research rather than cancer diagnosis itself [10].

Using three outcome categories—Living, Died of Disease, and Died of Other Causes—this study suggests a multi-algorithm machine learning prototype application for forecasting the survival status of breast cancer patients. The suggested system compares prediction performance using defined evaluation metrics by integrating NN, SVM, RF, NB, and KNN

algorithms into a single monitoring and decision-support framework. Instead than serving as a major clinical decision-making tool or an initial diagnosis, the program is meant to support survival tracking and analytical review. Furthermore, because the technology has not yet received clinical validation or been implemented in actual medical settings, it is still regarded as a prototype. Therefore, more validation and physician review are necessary before practical healthcare use can be considered.

An Application Using Multiple Algorithms to Track and Forecast Breast Cancer Survival The experimental inquiry will result in the development of patients. Several algorithms will be used, including the neural network method, naive bayes, random forest, KNN, and SVM. These five algorithms will be used to determine the degree of accuracy in predicting the survival status of patients with breast cancer [11]. This tool can also be used to periodically estimate the survival rate of patients with breast cancer. Machine learning techniques have been used in a number of prior research to analyze survival and forecast breast cancer. Nevertheless, without combining several algorithms into a single monitoring application, the majority of earlier research concentrated on binary classification, single-algorithm evaluation, or diagnostic prediction tasks. Furthermore, explainability, clinical validation, and application implementation restrictions were not well addressed in several investigations. In order to determine current research gaps and support the suggested contribution of this work, a comparative analysis of earlier research is crucial. A comparison of relevant papers on survival analysis and breast cancer prediction is shown in Table 1.

Table 1. Comparison of related studies on breast cancer prediction and survival analysis

Study	Dataset	Algorithm	Task	Evaluation Metric	Limitation
Ma et al. [4]	Breast cancer dataset	RIME-SVM	Diagnostic classification	Accuracy	No multiclass survival prediction
Balkenende et al. [12]	Breast cancer imaging dataset	Deep Learning	Breast cancer imaging analysis and diagnosis	Accuracy, imaging performance	Focused on imaging interpretation rather than survival status prediction or

					application implementation
Nemade & Fegade [13]	Breast cancer prediction dataset	Machine Learning Techniques	Breast cancer prediction	Accuracy	Primarily focused on prediction performance without multiclass survival analysis or integrated monitoring application
Supriya & Chengoden [14]	Federated learning breast cancer dataset	SHAP-based Federated Learning	Breast cancer prediction and explainability	Accuracy, explainability analysis	Focused on explainable federated learning without comparative multi-algorithm survival prediction
Proposed Study	METABRIC secondary/public dataset	NN, SVM, RF, NB, KNN	Multiclass survival status classification	Accuracy, Precision, Recall, F1-score	Prototype application not yet clinically validated

This study suggests a multi-algorithm machine learning prototype application for forecasting the survival status of breast cancer patients utilizing three outcome categories: Living, Died of Disease, and Died of Other Causes, based on the comparison mentioned above. The suggested methodology compares prediction performance using defined evaluation metrics by integrating NN, SVM, RF, NB, and KNN algorithms into a single monitoring and decision-support system. In contrast to other research, this study concentrates on comparative method evaluation and multiclass survival prediction inside an integrated application framework. However, because it has not yet been clinically

validated or implemented in actual healthcare settings, the developed application is currently regarded as a prototype [15].

Particularly for applications involving survival prediction and categorization, machine learning techniques have demonstrated considerable promise in aiding the interpretation of breast cancer data. However, rather than being a major clinical diagnosis tool, the suggested system in this study should be viewed as a decision-support prototype. By offering prediction insights based on past patient data and comparative algorithm analysis, the program is intended to support healthcare practitioners. The system's predictions should be viewed as helpful analytical outputs rather than conclusive medical diagnoses or treatment choices because it has not undergone prospective clinical validation or actual hospital implementation [16].

2. METHODS

These are the steps involved in creating a multi-algorithm-based application that uses the Neural Network method to track and forecast breast cancer patients' survival. The accuracy of the algorithm is then assessed using NB, RF, KNN, SVM, and NN.

The steps involved in creating a multi-algorithm-based application to track and forecast breast cancer patients' survival using neural network algorithms are depicted in Figure 1. Identification of the issues is the initial step, which is followed by analysis, design, and coding. A secondary public dataset from the METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) breast cancer dataset, which is accessible on the Kaggle platform, was used in this investigation. Raghad Alharbi's "Breast Cancer Gene Expression Profiles (METABRIC)" repository provided access to the dataset [17]. The dataset, which has been extensively utilized for machine learning and survival prediction research, includes clinical and survival-related data from patients with breast cancer. In May 2026, the dataset was obtained via the Kaggle repository and utilized in compliance with the platform's public dataset usage policy. The dataset utilized in this study matches the version that was accessible on the Kaggle repository at the time of access. There was no direct patient involvement during the research procedure because the dataset came from secondary publicly available data. Furthermore, no personally identifying information was handled during the trial because all patient records were anonymised

prior to publication. The analysis did not include records with missing essential features, duplicate entries, incorrect labeling, or inadequate survival outcome information.

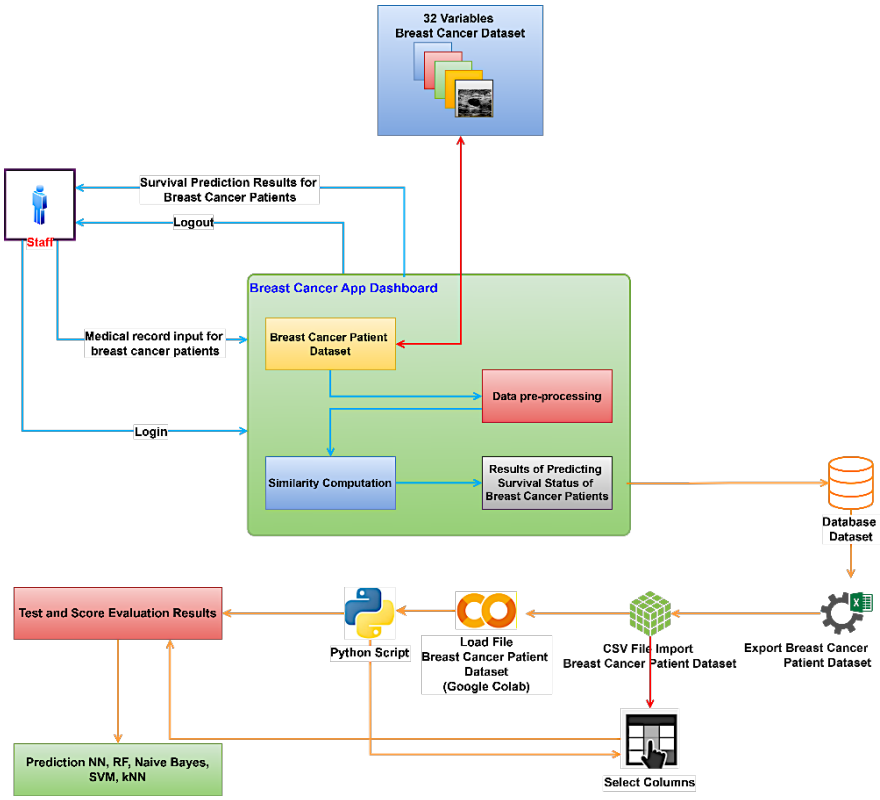

Figure 1. Machine learning evaluation method for monitoring and predicting breast cancer patient survival

1,980 breast cancer patient records from the publicly accessible METABRIC dataset made up the dataset used in this investigation. To guarantee data completeness, consistency, and appropriateness for multiclass survival prediction analysis, inclusion and exclusion criteria were used prior to model creation. The inclusion criteria included valid numerical or categorical values that could be processed during preprocessing and normalization stages, clearly defined survival outcome labels classified as "Living," "Died of Disease," or "Died of Other Causes," and patient records with all the clinical and demographic information needed for machine learning analysis. Furthermore, the analysis only included entries from the final, cleaned version of the METABRIC dataset. In the meantime, duplicate patient entries found during data cleaning, records with excessive missing values in significant predictor variables that could not be processed reliably, and records with missing, incomplete, or inconsistent survival status information were all

eliminated as part of the exclusion criteria. Additionally, in order to lower the risk of target leakage during model training and evaluation, variables directly related to survival outcomes, such as "Overall Survival Status," "Overall Survival (Months)," "Relapse Free Status," and "Relapse Free Status (Months)," were removed from the predictor feature set. Additionally, records that failed preprocessing or normalization methods and included corrupted or invalid values were not included. 1,980 eligible patient records were kept for preprocessing, dataset splitting, model training, validation, and testing after these criteria were applied.

To guarantee compliance with the machine learning techniques employed in this investigation, categorical variables were converted into numerical representations during the preprocessing phase. The encoding procedure was carried out according to each variable's properties. Only categorical variables with naturally ordered relationships—such as tumor stage or severity level where the categories represent significant ranking structures were subjected to ordinal encoding. Because the encoded numerical values maintained the natural sequence of the clinical data, this method was deemed suitable. However, encoding techniques that avoid artificial ranking assumptions such as one-hot encoding or label encoding, depending on the algorithm requirements and data structure should be chosen for nominal categorical variables that lacked ordinal relationships, such as gender, treatment category, or mutation classification. In order to minimize bias and stop machine learning models from misinterpreting nominal categories as ordered numerical values, encoding techniques were chosen. Following encoding, the dataset was normalized and features were chosen before RF [18], KNN [2], SVM [19], and NN [20] models were trained and assessed.

Figure 2 shows the research framework for tracking and predicting the survival of patients with breast cancer. This approach shows how medical professionals engage with a predictive system by illuminating the workflow of a Breast Cancer Application Dashboard. Staff and doctors are the two main user roles served by the application; both can create accounts, log in, enter data, and input medical records for patients with breast cancer. The user then inputs the patient's details and the breast cancer variable's values. Chemotherapy, tumor stage, ER status measured by IHC, mutation count, inferred menopausal state, tumor size, overall survival status, hormone therapy, cellularity, integrative cluster, HER2 status, cohort, primary tumor laterality, 3-gene classifier

subtype, lymph nodes examined positively, cancer type detailed, radio therapy, sex, Nottingham prognostic index, neoplasm histologic grade, overall survival (months), cancer type, HER2 status measured by SNP6, PAM50 + claudin-low subtype, tumor other histologic subtype, ER status, and type of breast surgery. The system uses a multi-algorithm to evaluate data and determine the survival status of patients with breast cancer based on 32 input variables. This method's output data generates a dataset on breast cancer, which is then stored in a database. Python was used in the Google Colab environment for the machine learning research. The SVM, RF, NB, and KNN algorithms were implemented using Scikit-learn, and the Neural Network (NN) model was developed using TensorFlow/Keras. Additionally, comparative model evaluation and visual process analysis were supported by Orange Data Mining. For data preprocessing, analysis, and visualization, other libraries including NumPy, Pandas, and Matplotlib were utilized. To enable a fair comparative analysis, all models were trained and assessed using the same preprocessing techniques and dataset splitting setup.

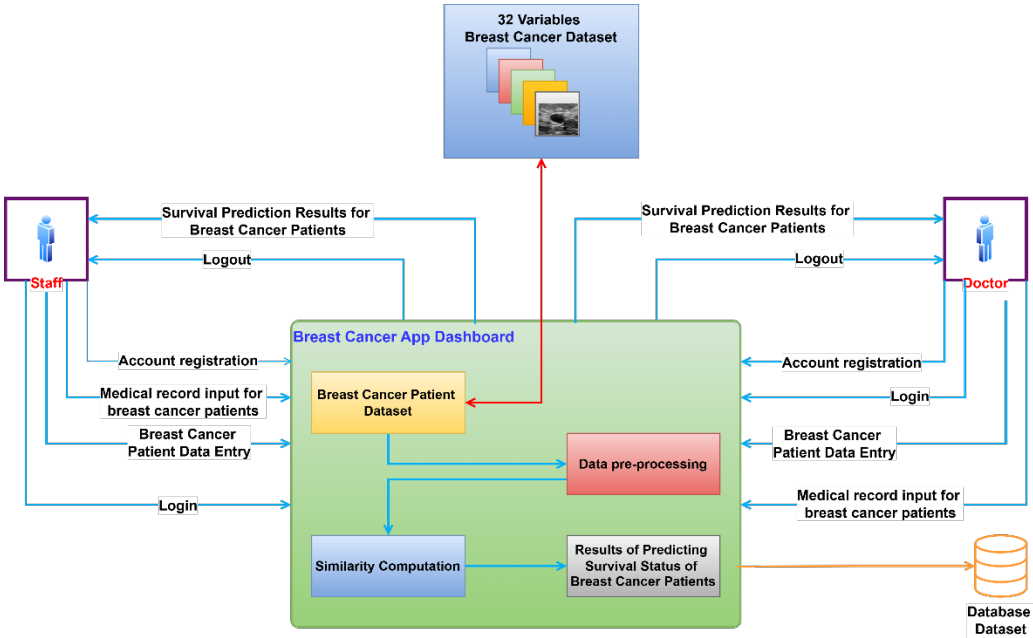

Figure 2. Application development framework for monitoring and predicting breast cancer patient survival

In order to facilitate comparative machine learning evaluation across the NN, SVM, KNN, RF, and NB algorithms using the same input configuration, 32 variables were used. However, during preprocessing and evaluation, variables that can contribute target

leakage such as characteristics closely linked to survival outcomes were thoroughly examined. Data preprocessing was done prior to model training in order to enhance data quality and lower the possibility of biased prediction outcomes. Data cleaning, addressing missing values, categorical data encoding, normalization, and feature selection were among the preprocessing steps. Before rerunning the models, a number of variables that were directly or substantially connected with patient survival outcomes were removed from the predictor set in order to avoid target leakage during machine learning evaluation. "Overall Survival Status," "Overall Survival (Months)," "Relapse Free Status," and "Relapse Free Status (Months)" were among the variables eliminated because they contain information that is closely related to the target variable and could artificially inflate prediction performance if used as input features. "Living," "Died of Disease," and "Died of Other Causes" were the three multiclass survival categories that made up the study's aim variable. The evaluation procedure was created to better represent realistic survival prediction scenarios and enhance the model's capacity for generalization by eliminating leakage-prone variables.

The 1,980 patient records in the breast cancer dataset were extracted in CSV format from a database and processed in Google Colab for survival prediction using machine learning. To facilitate model training, parameter adjustment, optimization, and final evaluation while minimizing overfitting, the dataset was split into subsets for training, validation, and testing. Using NumPy and Pandas for preprocessing, Scikit-learn for SVM, RF, KNN, and NB, and TensorFlow/Keras for the Neural Network model, the study constructed five algorithms: [21], SVM [22], KNN [23], RF [24] and NB [25] models). Three survival classes made up the dataset utilized in this study: Living (837 cases), Died of Disease (646 instances), and Died of Other Causes (497 occurrences). All classes were adequately represented for multi-class classification analysis, despite the fact that the class distribution was not exactly balanced. The disparity between classes was deemed modest. The dataset was split proportionally into training, validation, and testing subsets using an 80:10:10 ratio while preserving the original class distribution in each subset to lower the risk of bias during model training and evaluation. Every survival category was uniformly represented in the training, validation, and testing datasets thanks to this proportionate distribution. In order to identify the best algorithm for predicting breast cancer survival, the evaluation process comprised preprocessing, dataset splitting, model training with the same configuration, confusion matrix analysis, and performance

evaluation using accuracy, precision, recall, and F1-score metrics. Nevertheless, this study did not use any particular resampling methods like oversampling, undersampling, or SMOTE.

3. RESULTS AND DISCUSSION

Phases of software development that use neural network algorithms to track and forecast breast cancer patients' survival include building a dataset from patient input characteristics, which is subsequently utilized to forecast breast cancer patients' survival. Naive Bayes, Random Forest, KNN, SVM, and Neural Network techniques are used in this procedure. At this point in the analysis, problem identification is done. The software for the monitoring and prediction system of breast cancer patient survival must be developed in order to measure the accuracy of survival forecasts. 33 patient input data variables were used in the software's development.

3.1. Problem Identification

Breast cancer survival prediction remains challenging due to the complexity of factors influencing patient outcomes. Unlike diagnosis, which focuses on detecting cancer, survival prediction aims to estimate patient outcomes after diagnosis and treatment. This study addresses limitations in previous research by proposing a leakage-aware, multi-algorithm machine learning framework using the METABRIC dataset to classify patients into three categories: Living, Died of Disease, and Died of Other Causes. The framework compares the performance of NN, SVM, RF, NB, and KNN algorithms and implements a prototype decision-support monitoring application to support realistic evaluation of multiclass survival prediction models.

3.2. Analysis Problem

This study addresses the problem of predicting breast cancer patient survival status using clinical and demographic data from the METABRIC dataset. The task was formulated as a multiclass classification problem with three outcomes: Living, Died of Disease, and Died of Other Causes. Five machine learning algorithms NN, SVM, RF, NB, and KNN were compared using a leakage-controlled framework that included data preprocessing, feature selection, encoding, normalization, training, validation, and testing. Survival related variables were removed to prevent target leakage, and model performance was

evaluated using accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC metrics. The results were integrated into a prototype decision-support application to provide a realistic assessment of algorithm performance and identify the strengths and limitations of each model for breast cancer survival prediction.

3.3. Design of a Software Prototype for Predicting Breast Cancer Patient Survival Based on Multi-Algorithm

This is the software prototype for the Breast Cancer Patient Survival Prediction Monitoring System, as shown in Figure 3. Figure 3 shown the user interface (UI) design and interactive user workflow of a software prototype for predicting breast cancer patient survival based on a multi-algorithm approach. The diagram connects the user roles of Doctor and Staff (on the far right) to various application screens via directional blue data flow lines, mapping out how administrative and clinical tasks translate into digital interfaces. When users register accounts or log in via the main authentication page, they are directed to the primary Breast Cancer App Dashboard interface in the center, which features an "All Actions" hub where medical professionals can tap options to add new patients, run diagnostic computations, or manage profiles. From this center console, navigating to "Add Patient" routes the user to a "New Patient" profile form (bottom right), while opting for clinical analysis opens an extensive sequence of multi-page "Diagnosis Input" screens (distributed across the top and bottom center rows) designed to collect detailed medical factors such as tumor size, chemotherapy usage, and histologic grades. Once this clinical data is processed by the underlying multi-algorithm prediction engine, the final outputs are routed back to specialized user screens such as the "AI Check" status page on the mid-left or the "Survival Prediction Results for Breast Cancer Patients" profile view on the bottom-left enabling healthcare workers to efficiently interpret automated prognostic insights directly from their personalized accounts.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

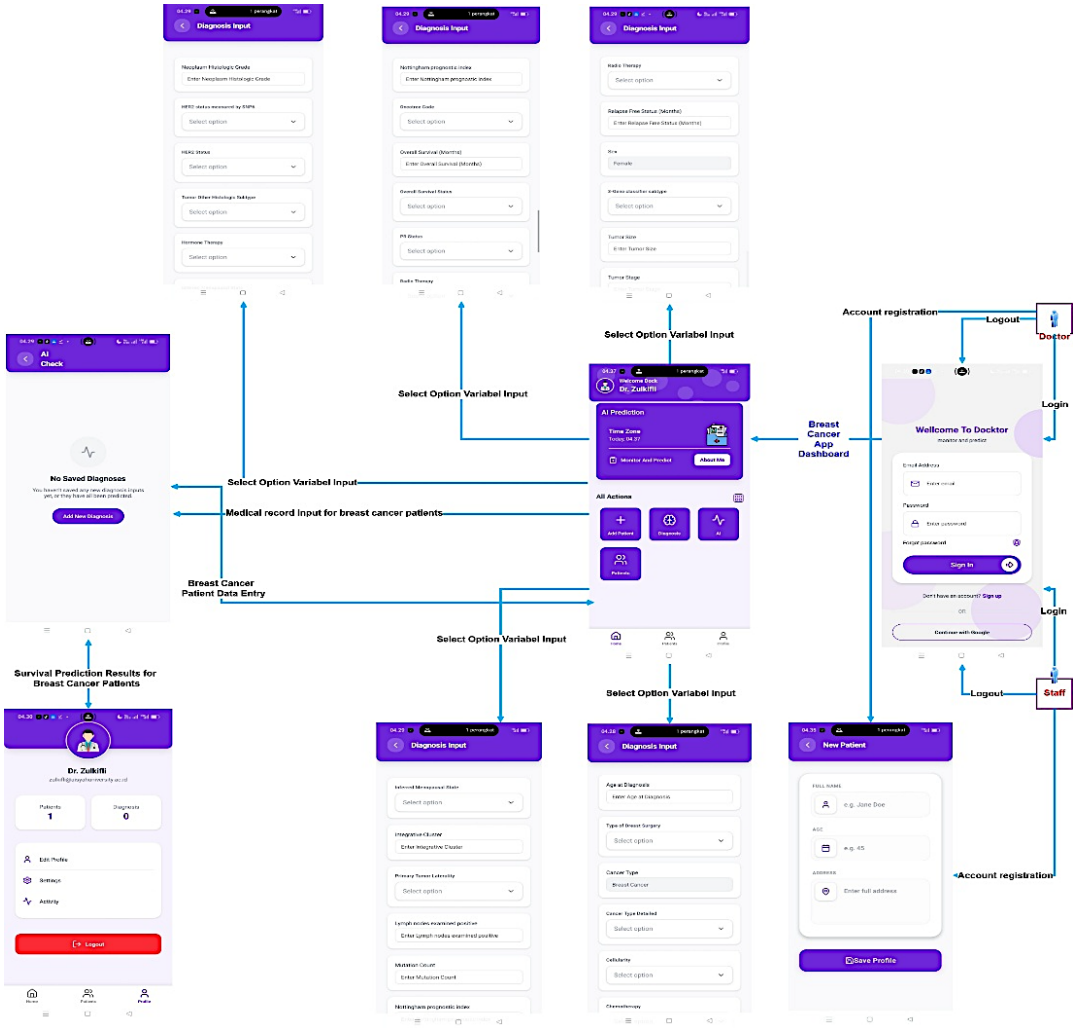

Figure 3. Design of a software prototype for predicting the survival of breast cancer patients based on a multi-algorithm approach

3.4. Complexity of Software Development for Monitoring and Predicting Breast Cancer Patient Survival

The suggested prototype system's application development effort, working time, and implementation cost were all roughly estimated using the Use Case Point (UCP) technique. According to the estimation results, a three-person development team will need to put in between 687 and 1,716 working hours during the development phase, with an anticipated implementation cost of Rp 1,098,547,200. In order to preserve attention on the predictive assessment framework and enhance methodological clarity, the complete UCP calculation methodology is described separately inside the application

development portion, as the primary focus of this study is machine learning-based survival prediction performance.

3.5. Dataset Database

The dataset utilized to gauge the degree of accuracy in predicting breast cancer patient survival using the NB, RF, KNN, SVM, and NN algorithms is displayed in Table 2 [26]. There were 1,980 data instances in the dataset, which was divided into three classes: Living (837 data instances), Died of Disease (646 data instances), and Died of Other Causes (497 data instances). After then, the dataset was split into three subsets: 10% for testing, 10% for validation, and 80% for training. There were 669 training, 84 validation, and 84 testing sets of data in the Living class. There were 516 training, 65 validation, and 65 testing sets of data in the Died of Disease class. In the meantime, there were 399 training, 49 validation, and 49 testing sets of data in the Died of Other Causes class. In order to retain the overall number of 1,980 data instances, the final dataset distribution included 1,584 training data, 198 validation data, and 198 testing data. Ten of the 1980 datasets are included in this analysis.

3.6. Dataset Normalization

A normalized dataset is shown in Table 3. The dataset was preprocessed, including data cleansing, addressing missing values, categorical data transformation, and numerical normalization, prior to model training. To increase model stability and convergence during training, the normalization method was used to scale numerical attributes into a particular interval [27]. The following formula was used to apply Min-Max normalization in this study:

$$X' = 0.1 + \frac{(X - X_{min})(0.9 - 0.1)}{X_{max} - X_{min}} \quad (1)$$

where X represents the original value, $X_{\min}$ is the minimum value of the attribute, and $X_{\max}$ is the maximum value of the attribute. The data is scaled into the range of 0.1 to 0.9 via this procedure. Nonetheless, a few of the preprocessing results' values fall outside of the specified range, which could be a sign of unnormalized attributes, inconsistent rounding, or different preprocessing methods for numerical and categorical variables. To guarantee consistency between the applied formula and the provided results, the preprocessing process and normalizing outputs should be examined and explained.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Table 2. Dataset Database

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Py-001	75,65	1	1	3	3	0	2	1	2	2	3	3	1	1	1	2	4ER+	2	10	0	6	3	140,5	2	1	1	138,65	0	1	1	22	2	1
Py-002	43,19	2	1	3	3	0	4	1	2	2	3	3	1	1	1	1	4ER+	2	0	2	4	3	84,63	2	2	1	83,52	0	1	2	10	1	1
Py-003	47,68	1	1	6	2	1	5	1	2	2	2	3	1	5	1	1	9	2	3	1	4	6	164,93	2	2	1	162,76	0	1	0	25	2	1
Py-004	56,45	2	1	3	2	1	5	1	2	2	2	2	1	1	1	2	3	2	1	4	4	3	164,33	2	2	1	162,17	0	1	0	10	2	1
Py-005	70,91	2	1	3	3	0	5	1	2	2	1	1	1	1	1	2	4ER+	1	0	3	2	3	163,53	2	2	1	161,38	0	1	0	21	1	1
Py-006	45,27	1	1	3	3	1	2	1	1	1	3	3	1	1	0	1	4ER-	2	3	0	5	3	164,9	2	2	1	105,99	1	1	0	19	2	1
Py-007	51,46	2	1	3	1	1	2	1	2	2	2	1	2	1	1	2	4ER+	1	1	0	4	3	103,83	2	2	1	102,47	0	1	0	25	2	1
Py-008	44,64	2	1	6	2	1	7	1	2	2	2	3	1	5	1	1	8	2	3	0	4	6	75,33	2	2	1	74,34	0	1	3	33	2	1
Py-844	76,89	1	1	4	2	0	5	1	2	2	3	3	1	1	1	2	1	2	0	8	4	3	114,23	1	2	0	25,86	1	1	2	25	2	2
Py-1980	62,9	2	1	3	3	0	5	4	2	2	3	3	1	1	1	2	1	1	45	4	6	3	175,96	1	2	1	121,18	1	1	0	25	0	3

Note: Column 1: Patient ID; 2: Age at Diagnosis; 3: Type of Breast Surgery (Select Option: 1: Mastectomy; 2: Breast Conserving); 4: Cancer Type(Select Option: 1: Breast Cancer; 2: Breast Sarcoma) ; 5: Cancer Type Detailed (Select Option: 1: Breast; 2: Breast Angiosarcoma; 3: Breast Invasive Ductal Carcinoma; 4: Breast Invasive Lobular Carcinoma; 6: Breast Mixed Ductal and Lobular Carcinoma); 6: Cellularity (Select Option: 1: Low; 2: Moderate; 3: High); 7: Chemotherapy (Select Option: 1: Yes; 0: No); 8: Pam50 + Claudin-low subtype(Select Option: 1: Basal; 2: Claudin-low; 3: Her2; 4: LumA; 5: LumB; 6: NC; 7: Normal); 9: Cohorts; 10: ER status measured by IHC (Select Option: 1: Negative; 2: Positive); 11: ER Status (Select Option: 1: Negative; 2: Positive); 12: Neoplasm Histologic Grade; 13: HER2 status measured by SNP6 (Select Option: 1: Gain; 2: Loss; 3: Neutral; 4: Undef); 14: HER2 Status (Select Option: 1: Negative; 2: Positive); 15: Tumor Other Histologic Subtype (Select Option: 1: Ductal/NST; 2: Lobular; 3: Medullary; 4: Metaplastic; 5: Mixed; 6: Mucinous; 7: Other; 8: Tubular/ cribriform); 16: Hormone Therapy (Select Option: 1: Yes; 0: No); 17: Inferred Menopausal State (Select Option: 1: Pre; 2: Post); 18: Integrative Cluster; 19: Primary Tumor Laterality (Select Option: 1: left; 2: Right); 20: Lymph nodes examined positive; 21: Mutation Count; 22: Nottingham prognostic index; 23: Oncotree Code (Select Option: 1: BRCA; 2: BREAST; 3: IDC; 4: ILC; 5: IMMC; 6: MBC; 6: MDLC; 7: PBS); 24: Overall Survival (Months); 25: Overall Survival Status (Select Option: 1: Deceased; 2: Living); 26: PR Status (Select Option: 1: Negative; 2: Positive); 27: Radio Therapy (Select Option: 1: Yes; 0: No); 28: Relapse Free Status (Months); 29: Relapse Free Status (Select Option: 0: Not Recurred; 1: Recurred); 30: Sex (Select Option: 1: Female); 31: 3-Gene classifier subtype (Select Option: 0: Null; 1: ER-/HER2-; 2: ER+/HER2- High Prolif; 3: ER+/HER2- Low Prolif; 4: HER2+): 32: Tumor Size; 33: Tumor Stage and Column 34 namely: Patient's Vital Status (Result Option: 1: Living; 2: Died of Other Causes; 3: Died of Disease)

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Table 3. Dataset Normalization

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
Py-0001		0.96	0	0	0	1	0	0	0	1	1	1	1	0	0	1	1	1	0.22	0	1	0	0.64	1	0	1	0.82	0	0	0.33	0.52	1
Py-0002		0	1	0	0	1	0	0.4	0	1	1	1	1	0	0	1	0	1	0	0.25	0.5	0	0.09	1	1	1	0.42	0	0	0.66	0	0.5
Py-0003		0.13	0	0	1	0.5	1	0.6	0	1	1	0.5	1	0	1	1	0	1	0.06	0.125	0.5	1	0.89	1	1	1	1	0	0	0	0.65	1
Py-0004		0.39	1	0	0	0.5	1	0.6	0	1	1	0.5	0.5	0	0	1	1	1	0.02	0.5	0.5	0	0.88	1	1	1	0.99	0	0	0	0	1
Py-0005		0.82	1	0	0	1	0	0.6	0	1	1	0	0	0	0	1	1	0	0	0.375	0	0	0.87	1	1	1	0.98	0	0	0	0.47	0.5
Py-0006		0.06	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	1	0.06	0	0.75	0	0.89	1	1	1	0.58	1	0	0	0.39	1
Py-0007		0.24	1	0	0	0	1	0	0	1	1	0.5	0	1	0	1	1	0	0.02	0	0.5	0	0.28	1	1	1	0.55	0	0	0	0.65	1
Py-0008		0.04	1	0	1	0.5	1	1	0	1	1	0.5	1	0	1	1	0	1	0.06	0	0.5	1	0	1	1	1	0.35	0	0	1	1	1
Py-0844		1	0	0	0.33	0.5	0	0.6	0	1	1	1	1	0	0	1	1	1	0	1	0.5	0	0.38	0	1	0	0	1	0	0.66	0.65	1
Py-1980		0.58	1	0	0	1	0	0.6	1	1	1	1	1	0	0	1	1	0	1	0.5	1	0	1	0	1	1	0.69	1	0	0	0.65	0

Note: Column 1: Patient ID; 2: Age at Diagnosis; 3: Types of Breast Surgery; 4: Cancer Type; 5: Cancer Type 6: Cellularity; 7: Chemotherapy; 8: Pam50 + Claudin-low subtype; 9: Cohorts; 10: ER status measured by IHC; 11: ER Status; 12: Neoplasm Histologic Grade; 13: HER2 status measured by SNP6; 14: HER2 Status; 15: Tumor Other Histologic Subtype; 16: Hormone Therapy; 17: Inferred Menopausal State; 18: Integrative Cluster; 19: Primary Tumor Laterality; 20: Lymph nodes examined positive; 21: Mutation Count; 22: Nottingham prognostic index; 23: Oncotree Code; 24: Overall Survival (Months); 25: Overall Survival Status; 26: PR Status; 27: Radio Therapy; 28: Relapse Free Status (Months); 29: Relapse Free Status; 30: Sex; 31: 3-Gene classifier subtype; 32: Tumor Size and; 33: Tumor Stage

3.7. Model Training and Methodology

The sequentially structured NN architecture used to optimize the feature extraction and classification procedures is shown in Figure 4. Two completely linked (Dense) hidden layers (Hidden Layers) of 512 and 256 neurons, respectively, come after the input layer, which gets the dimensions of the raw data. Each Dense layer is merged with the Batch Normalization mechanism, which stabilizes the input distribution between layers to enable faster convergence, and the ReLU activation function to solve non-linearity issues in order to improve learning efficiency [28]. In order to prevent overfitting, a Dropout layer with a rate of 0.3 is also applied as a regularization strategy by randomly deactivating neurons during the training phase. The output layer, which generates the final classification probabilities using the softmax activation function, is where this flow comes to a conclusion.

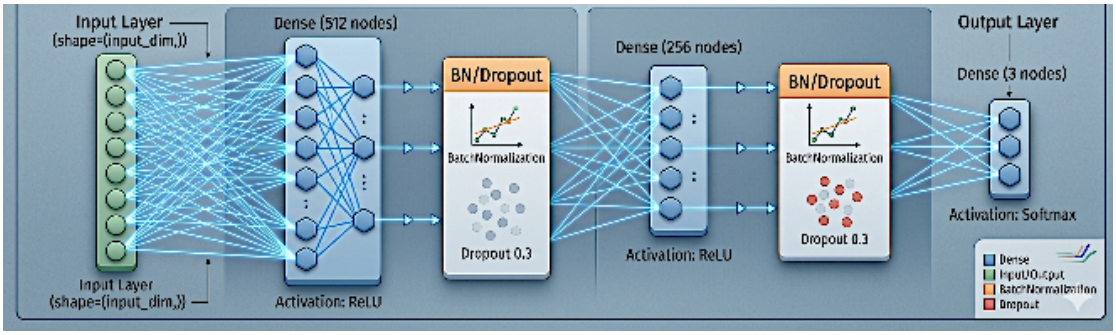

Figure 4. Architecture of NN algorithm

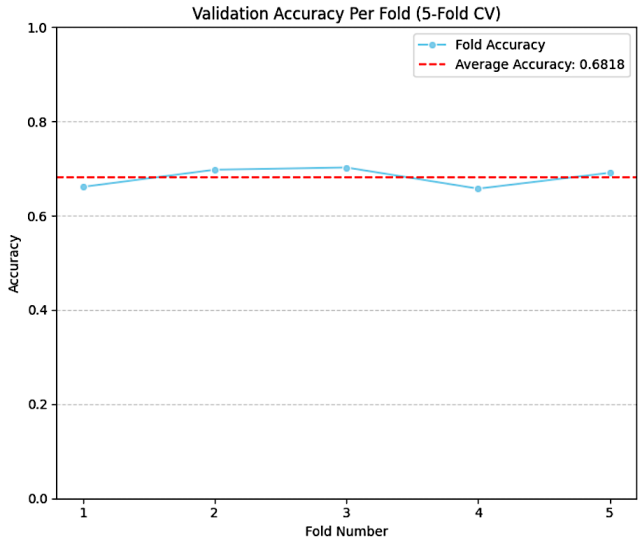

Figure 5. Performance of loss training and validation NN algorithm

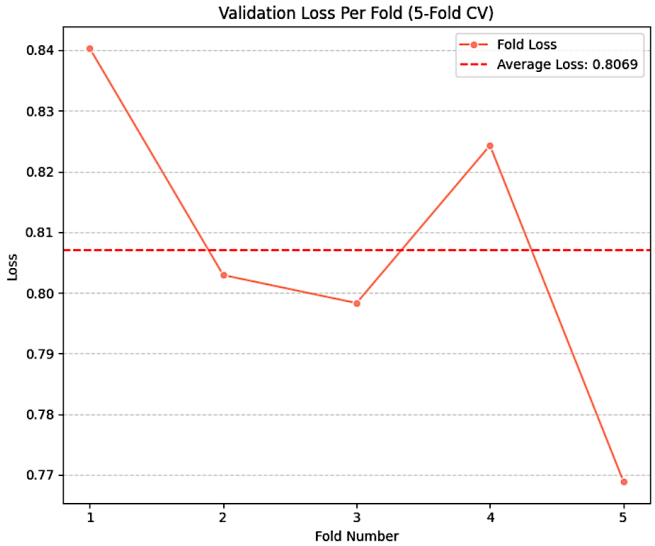

Figure 6. Performance of accuracy training and validation NN algorithm

The 5-Fold Cross Validation results are displayed in Figures 5 and 6, where the Neural Network (NN) method showed reasonably consistent performance across all folds. The average accuracy on the validation accuracy graph was 68.18%, with the second and third folds showing the highest accuracy values. The validation loss graph, on the other hand, displayed an average loss of 0.8069, with the fifth fold having the lowest loss value. Overall, these findings show that the model can make rather accurate predictions, while model improvement and improved data processing can still enhance its performance. Table 4 presents the test results for the SVM algorithm [29].

The 5-fold cross-validation approach was used to assess the Support Vector Machine (SVM), Random Forest (RF), Naive Bayes (NB), and K-Nearest Neighbor (KNN) algorithms with various parameter configurations. Polynomial (poly), radial basis function (rbf), and linear kernels with C = 1 and scale parameters were used to test the SVM method. The results showed a perfect average accuracy of 1.00 (100%) across all folds, demonstrating very consistent classification performance [30]. With n_estimators=500, random_state=10, max_depth=5, criterion='entropy', and class_weight=None, the RF method produced an average accuracy of 0.77 (77%), demonstrating high predictive power but falling short of SVM. When the NB algorithm was tested with class_prior=None, fit_prior=False, alpha=1.0, and force_alpha=False, it demonstrated good and steady classification performance with an accuracy of 0.95 (95%). In a similar vein, the KNN method with n_neighbors=5, weights='uniform', and metric='minkowski' likewise attained

an accuracy of 0.95 (95%) during testing, demonstrating good prediction performance and efficient classification capabilities [31].

Table 4. Performance training of SVM, RF, NB, and KNN

Model	Fold	Parameter Settings	Avg Accuracy
SVM	5	(C=1; poly; scale; degree=2)	1.0
		(C=1; rbf; scale; degree=4)	1.0
		(C=1; linear; scale; degree=3)	1.0
RF	5	(n_estimators=500, random_state=10, max_depth=5, criterion='entropy', class_weight = None, verbose = 0)	0.77
NB	5	(class_prior=None, fit_prior=False, alpha=1.0, force_alpha=False)	0.95
KNN	5	(n_neighbors=5, weights='uniform', metric='minkowski', metric_params=None, n_jobs=None)	0.95

3.8. Evaluation Confusion Matrix and Performance Evaluation

A sample of 198 of the 1,980 breast cancer patient records included in this investigation were designated as testing data for the confusion matrix analysis and final model evaluation. In order to guarantee that the evaluation process was carried out on previously unseen data, the testing subset, which comprised roughly 10% of the entire dataset, was divided at the data splitting stage. In order to ensure that each survival class—"Living," "Died of Disease," and "Died of Other Causes"—was represented in the testing dataset, the selection of these test samples was done methodically following preprocessing and data cleaning methods. The classification performance of the NN, SVM, RF, NB, and KNN models on the identical 198 independent test samples is thus reflected in the confusion matrices provided in this paper.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

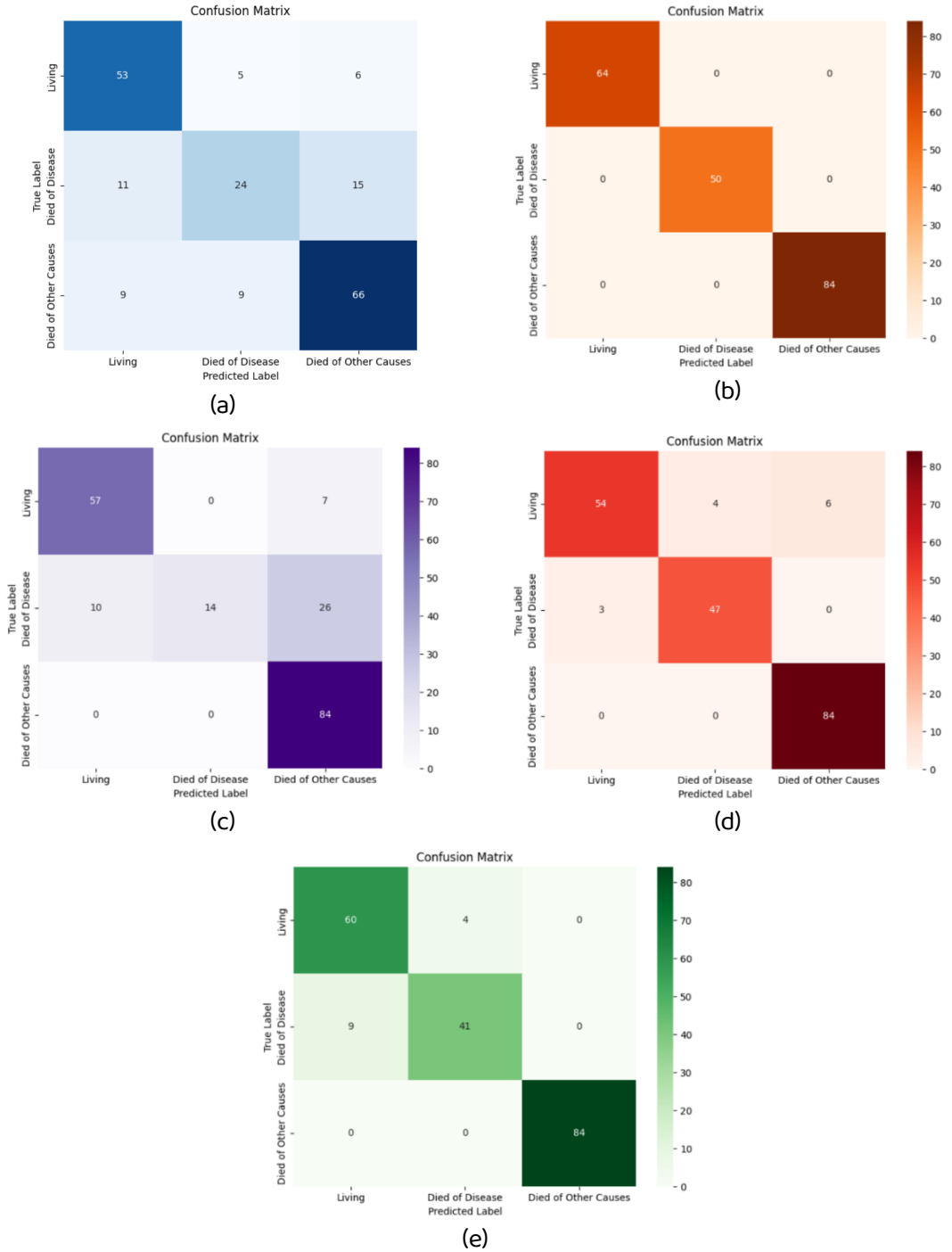

Figure 7. Confusion matrix (a) NN, (b) SVM, (c) RF, (d) NB, (e) KNN performance

The classification performance of the NN, SVM, RF, NB, and KNN models in predicting the three classes—Living, Died of Disease, and Died of Other Causes is shown by the confusion matrix findings. Although 55 data points were incorrectly classified, especially when it came to identifying patients who died from disease, the NN model accurately

predicted 143 individuals, including 53 Living, 24 Died of Disease, and 66 Died of Other Causes cases [32]. Excellent accuracy and consistency were demonstrated by the SVM model's flawless classification performance, which correctly predicted all 198 samples with no misclassification across all classes [33]. The RF model performed well in the Living and Died of Other Causes classes, properly identifying 57 and 84 data points, respectively. However, it had trouble differentiating the Died of Disease class because of many misclassifications into other categories [34]. Additionally, the NB model performed exceptionally well, accurately identifying all Died of Other Causes cases and correctly classifying the majority of samples with only small errors between the Living and Died of Disease classes [35]. Although there was still considerable misclassification between the Living and Died of Disease categories, the KNN model also performed well, accurately predicting the majority of data points and attaining 100% classification in the Died of Other Causes class [23]. Overall, the results suggest that SVM outperformed NN and RF in classification, although NB and KNN also shown significant predictive abilities. The performance evaluation of NN, SVM, RF, NB, and KNN is displayed in Table 5.

Table 5. Performance evaluation of NN, SVM, RF, NB, and KNN

	Precision	**Recall**	**F1-Score**	**Class**
NN Model	0.76	0.75	0.75	Living
	0.80	0.67	0.73	Died of Disease
	0.54	0.66	0.59	Died of Other Causes
Weighted Avg.	0.72	0.70	0.71	
Macro Avg.	0.70	0.69	0.69	
SVM Model	Precision	Recall	F1-Score	Class
	1.0	1.0	1.0	Living
	1.0	1.0	1.0	Died of Disease
	1.0	1.0	1.0	Died of Other Causes
Weighted Avg.	1.0	1.0	1.0	
Macro Avg.	1.0	1.0	1.0	
RF Model	Precision	Recall	F1-Score	Class
	0.88	0.85	0.86	Living
	1.0	0.31	0.47	Died of Disease
	0.69	1.0	0.82	Died of Other Causes

Weighted Avg.	0.83	0.78	0.75	
Macro Avg.	0.86	0.72	0.72	
NB Model	Precision	Recall	F1-Score	Class
	0.95	0.84	0.89	Living
	0.92	0.94	0.93	Died of Disease
	0.93	1.0	0.97	Died of Other Causes
Weighted Avg.	0.93	0.93	0.93	
Macro Avg.	0.93	0.93	0.93	
KNN Model	Precision	Recall	F1-Score	Class
	0.87	0.94	0.90	Living
	0.91	0.82	0.86	Died of Disease
	1.0	1.0	1.0	Died of Other Causes
Weighted Avg.	0.94	0.93	0.93	
Macro Avg.	0.93	0.92	0.92	

The classification performance of five different machine learning models NN, SVM, RF, N), and KNN is assessed and contrasted in this extensive version of Table 5 for the patient outcome classes "Living," "Died of Disease," and "Died of Other Causes." With an absolute score of 1.0 in precision, recall, and F1-score for each category and average, the SVM Model exhibits mathematically perfect, albeit suspiciously perfect, metrics. The NB Model, which exhibits extremely stable and robust performance overall and results in identical Weighted and Macro Averages of 0.93, comes next in terms of reliability. With a Weighted Average F1-score of 0.93 and a Macro Average of 0.92, as well as a perfect 1.0 across all measures, particularly for the "Died of Other Causes" class, the KNN Model competes at a similar tier. The RF Model, on the other hand, exhibits a highly skewed performance; although it maintains strong metrics for the "Living" class and a perfect 1.0 precision for "Died of Disease," its terrible recall of 0.31 for disease-related deaths significantly lowers its corresponding F1-score to 0.47, though it recovers enough to post a Macro Average F1-score of 0.72. Lastly, the NN Model performs poorly in respect to the rest of the cohort; it struggles especially with the "Died of Other Causes" class (0.59 F1-score) and anchors the bottom of the comparison with a Macro Average F1-score of only 0.69.

3.9. ROC Analysis

The Receiver Operating Characteristic (ROC) curve is used to evaluate the efficacy of classification models, particularly in binary and multi-class classification. The connection between the model's true positive rate (TPR) and false positive rate (FPR) at different decision thresholds is displayed by the ROC curve. The performance of each class and algorithm is displayed by multiple lines in Figure 8. Each class's ROC is displayed in the other lines; an AUC (Area Under the Curve) near 1.00 denotes excellent performance in class differentiation. The diagonal line represents a random classifier's performance. The ideal AUCs for the "Died of Disease" and "Died of Other Causes" classes in this plot show how successfully the model classified them. The multi-class Receiver Operating Characteristic (ROC) curves assessing the effectiveness of five distinct machine learning algorithms across three patient outcome classes—"Died of Disease" (blue), "Died of Other Causes" (orange), and "Living" (green)—are displayed in Figure 8. Plotting (a) NN, (b) SVM, (c) RF, (d) NB, and (e) KNN against a dashed diagonal baseline that represents a random classifier, each sub-graph represents a distinct method. The Area Under the Curve (AUC) measures each model's performance; a value nearer 1.0 denotes better classification abilities. With an AUC score of 0.89 for the class that died of illness and living and 0.83 for the class that died of other causes, Graph (a) NN shows predictive power. With a faultless AUC score of 1.00 for all three classes, Graph (b) SVM exhibits immaculate predicting performance. With an AUC of 1.00 for the "Living" class and 0.99 for both mortality classes, models (c) RF and (d) NB exhibit nearly flawless performance. With extremely reliable metrics, Graph (e) KNN closely follows, matching the flawless 1.00 for "Living" and hitting 0.98 for the other two classes. With an AUC of 0.89 for both "Living" and "Died of Disease," and an AUC of 0.83 for "Died of Other Causes," graph (a) NN, on the other hand, performs the worst and has the most gradual curves in the group.

The ROC-AUC evaluation was carried out utilizing a one-vs-rest (OvR) multiclass technique because this study involves multiclass survival status classification with three target categories ("Living," "Died of Disease," and "Died of Other Causes"). Using this method, each class was assessed independently, with one class being the positive class and the others being the negative class. In order to provide an overall assessment of the model's capacity for discriminating across all classes, the final ROC-AUC performance was then provided using macro-average AUC values. To see the classification performance for each survival category separately, class-specific ROC curves and AUC

values can be examined in addition to the macro-average ROC-AUC score. To guarantee fair comparison analysis in the multiclass prediction context, this assessment methodology was uniformly applied to all machine learning models, including NN, SVM, RF, NB, and KNN.

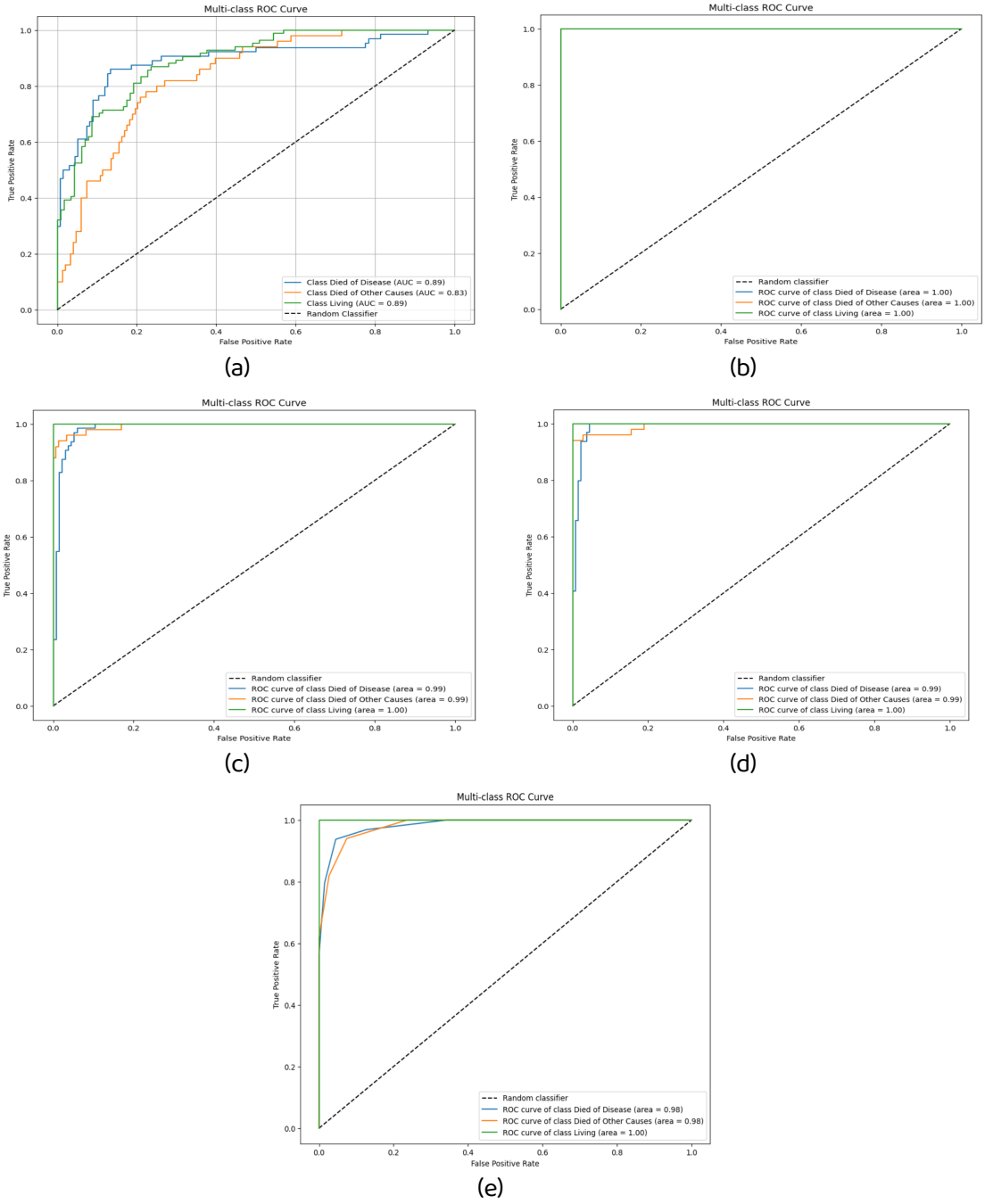

Figure 8. ROC curve each class's performance and each algorithm's (a) NN, (b), SVM, (c) RF, (d) NB, (e) KNN

3.10. Comparison of the use of Multi-Algorithms in Predicting the Survival Rate of Breast Cancer Patients

Figure 9 shows a comparative analysis of five machine learning algorithms NN, RF, NB, SVM, and KNN against the Actual ground truth data in predicting the survival rates of ten breast cancer patients. The horizontal axis represents the ten sampled patients, while the vertical axis maps categorical numerical codes assigned to survival outcomes, where a value of 1 corresponds to "Living," 2 corresponds to "Died of Disease," and 3 corresponds to "Died of Other Causes."

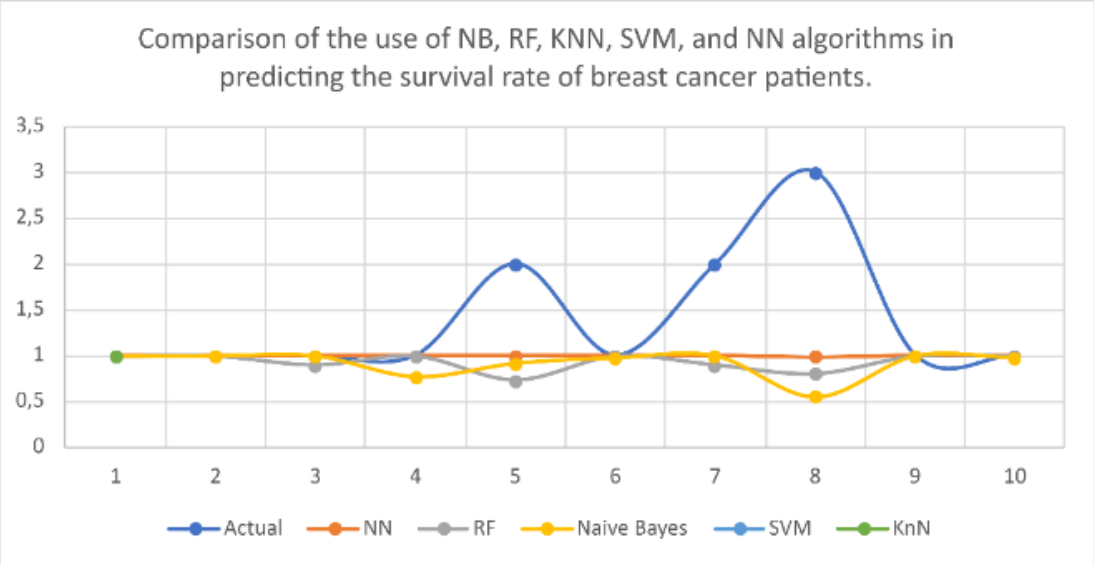

Figure 9. Results of a comparison of the accuracy levels in predicting the survival of breast cancer patients using the multi-algorithms

The validation results of a comparison of the NB, RF, KNN, SVM, and NN algorithms' predictions for breast cancer patient survival are displayed in Table 6. Each algorithm's validation column compares the expected and actual outcomes. They receive a "Valid" score if the expected outcomes are the same, and a "Invalid" score if they differ.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Table 6. Validation results of breast cancer patient survival prediction using multi-algorithm

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Py-0001	Living	0,9	Living	Valid	1	Living	Valid	0,9	Living	Valid	0,9	Living	Valid	0,8	Living	Valid
Py-0002	Living	1	Living	Valid	1	Living	Valid	0,9	Living	Valid	0,9	Living	Valid	0,8	Living	Valid
Py-0005	Died of Disease	0,9	Died of Disease	Valid	0,3	Died of Disease	Valid	0,9	Died of Disease	Valid	0,6	Died of Disease	Valid	0,8	Living	Invalid
Py-0006	Living	0,9	Living	Valid	0,9	Living	Valid	0,9	Living	Valid	0,9	Living	Valid	0,8	Living	Valid
Py-0014	Living	0,9	Living	Valid	0,9	Living	Valid	0,9	Living	Valid	0,9	Living	Valid	1	Living	Valid
Py-0056	Living	0,9	Living	Valid	1	Living	Valid	0,7	Living	Valid	0,9	Living	Valid	0,6	Living	Valid
Py-0083	Died of Disease	0,9	Died of Disease	Valid	0,8	Died of Disease	Valid	0,5	Died of Other Causes	Invalid	0,9	Died of Disease	Valid	0,7	Living	Invalid
Py-0149	Died of Other Causes	0,9	Died of Other Causes	Valid	0,8	Died of Other Causes	Valid	0,5	Died of Other Causes	Valid	0,9	Died of Other Causes	Valid	0,8	Died of Disease	Invalid
Py-2823	Living	0,9	Living	Valid	1	Living	Valid	0,9	Living	Valid	0,9	Living	Valid	1	Living	Valid
Py-4968	Living	0,9	Living	Valid	1	Living	Valid	0,9	Living	Valid	0,9	Living	Valid	0,8	Living	Valid

Note: 1= Patient Code, 2= Actual, 3= NN,4= Prediction,5= Validation,6= RF,7= Prediction,8= Validation,9= NB,10=Prediction ,11= Validation,12= SVM,13= Prediction,14= Validation,15= KNN,16= Prediction,17= Validation

The validation results of breast cancer patient survival forecasts utilizing a multi-algorithm method across 10 sample patient codes are displayed in Table 6. The footnote states that the table structure compares the predictions and validation results of five different machine learning models with the actual patient status (Column 2). Columns 3–5 are NN, Columns 6–8 are RF, Columns 9–11 are NB, Columns 12–14 are SVM, and Columns 15–17 are KNN. The precise status prediction (e.g., "Living," "Died of Disease," or "Died of Other Causes"), each model's similarity or probability score, and the resulting "Valid" or "Invalid" label that indicates whether the prediction matched the actual truth data are used to analyze each model's performance. According to the data, more complicated examples reveal algorithmic inconsistencies, whereas standard "Living" cases (such as Py-0001, Py-0002, Py-0006) enjoy unanimous, true predictions across all five algorithms. For example, patient Py-0083 is misclassified by both the NB model (predicting "Died of Other Causes") and the KNN model (predicting "Living"), demonstrating how these multi-algorithm validation comparisons help identify specific edge cases where individual models fail, while patient Py-0005 (actually "Died of Disease") is incorrectly classified as "Living" by the KNN model, resulting in a "Invalid" marker.

3.11. Discussion

With accuracy, precision, recall, and F1-score all hitting 1.00 on the testing dataset, the early studies demonstrated that the Support Vector Machine (SVM) model attained flawless performance values. Such near-perfect performance in medical prediction tasks required careful interpretation, even though these results showed excellent classification capability. This is because highly separable data patterns, limited dataset variability, overfitting, or possible data leakage between training and testing subsets could all have an impact. A number of survival-related variables, such as "Overall Survival Status," "Overall Survival (Months)," "Relapse Free Status," and "Relapse Free Status (Months)," were found to have a strong correlation with the target variable and may have caused target leakage by giving the model direct outcome-related information. All leakage-prone variables were eliminated from the predictor set in order to solve this problem, and the dataset was reprocessed using the same preprocessing, normalization, training, validation, and testing techniques to guarantee equitable comparison between methods. The SVM model no longer performed flawlessly after excluding these survival-related predictors, suggesting that target leakage or excessively informative factors probably had an impact on the initial results. As a result, the reevaluated results are thought to be more accurate

and offer a more trustworthy evaluation of the model's capacity for generalization and usefulness in healthcare prediction tasks.

To avoid target leakage during model training and evaluation, extra care must be taken during the feature selection phase. A number of the dataset's variables, including "Overall Survival Status," "Overall Survival (Months)," "Relapse Free Status," and "Relapse Free Status (Months)," have a strong correlation with the survival outcome and may provide direct insights into the target class. The models might learn outcome-related information that would not often be available in real-world prediction situations if these characteristics were added as predictor attributes while simultaneously predicting patient survival status ("Living," "Died of Disease," or "Died of Other Causes"). Evaluation measures, such as the flawless accuracy seen in the SVM models, may be falsely exaggerated as a result of this situation. To guarantee that the evaluation represents realistic prediction performance rather than information leakage, these factors should either be removed from the predictor set or explicitly justified within the experimental design.

When compared to the other assessed algorithms, the RF model performed worse, especially when it came to recall for the "Died of Disease" class. The lower recall value suggests that some patients in this survival category were not accurately identified by the RF model, leading to more false-negative predictions. This restriction could be caused by a number of things. First, the RF model may have been increasingly biased toward majority classes due to class imbalance in the dataset, which would have decreased sensitivity to the "Died of Disease" category. Second, the model's capacity to produce highly discriminative decision boundaries for minority or clinically difficult cases may have been hampered by the overlap in clinical characteristics among the three survival groups. Third, the model had fewer directly informative outcome features when survival-related leakage-prone variables were eliminated, making the classification job more difficult but more feasible. Additionally, when predictor variables have noisy or weakly linked characteristics that don't significantly aid in multiclass separation, Random Forest models may perform worse. RF performance can be influenced by feature significance and class distribution, in contrast to SVM and KNN, which may be better able to capture complex decision boundaries in normalized datasets. In order to enhance RF sensitivity and overall multiclass survival prediction performance, future research should take into

account using feature selection, class balancing strategies like SMOTE, hyperparameter optimization, and explainability analysis, especially for the "Died of Disease" category. The current work has not yet incorporated direct validation from doctors or medical professionals, despite the fact that the suggested application showed good prediction performance based on machine learning evaluation metrics. Future studies should include oncologists, doctors, and other healthcare professionals in the review process to determine the clinical relevance, interpretability, and usability of the prediction results if the system is meant to be used in hospital or healthcare settings. To guarantee that the application supports actual clinical workflows, offers valuable decision-support data, and complies with medical standards and patient care criteria, expert validation is crucial. Clinician input can also help to improve the system's general usefulness in actual healthcare settings, as well as its feature selection, interface design, and explanation procedures.

To evaluate the dependability and usability of the suggested monitoring system, application-level testing is crucial in addition to algorithm performance evaluation. Future iterations of the study should incorporate software testing techniques like user acceptance testing (UAT) to gauge system acceptance among healthcare users, usability testing to assess user interaction and interface efficacy, and black-box testing to confirm functional correctness of application features. Clinician input should be included if the application is meant for use in clinical settings in order to evaluate the prediction outputs' applicability, interpretability, and significance in aiding medical decision-making. The validity of the suggested application would be strengthened by these extra review phases from both a technical machine learning perspective and a real-world healthcare implementation one.

The suggested multi-algorithm framework's performance can be compared to a number of earlier works on survival analysis and breast cancer prediction. Previous research has mostly concentrated on binary classification tasks, including differentiating between benign and malignant cancers or forecasting overall survival using constrained outcome categories. This study, on the other hand, used a multiclass survival prediction approach with three target classes: "Living," "Died of Disease," and "Died of Other Causes." By combining NN, SVM, RF, NB, and KNN algorithms within the same experimental setting, dataset, and evaluation process, the suggested methodology offers a more

comprehensive comparative comparison than earlier studies that assessed only individual algorithms or standalone predictive models. The findings indicated that while RF, NB, and KNN exhibited differing degrees of efficacy for multiclass survival classification, NN and SVM produced the best prediction performance. The performance of several machine learning algorithms on breast cancer survival prediction tasks with increasingly complicated outcome categories is better understood thanks to this comparison approach.

The current work does not yet offer explainability analysis about the factors influencing prediction outcomes, despite the fact that the suggested machine learning models demonstrated good predictive performance. Interpretability is crucial in clinical decision-support systems because medical practitioners must comprehend the rationale behind a patient's classification into a specific survival group. In order to determine the most significant clinical variables influencing the predictions, future research should use explainable artificial intelligence (XAI) [36] techniques like feature importance analysis, SHAP (SHapley Additive exPlanations) [37], or LIME (Local Interpretable Model-Agnostic Explanations) [38]. Explainability study can boost confidence in the suggested application for practical healthcare use, evaluate whether the predictions are consistent with medical knowledge, and assist doctors in interpreting model behavior. Additionally, the display of significant features may enhance transparency in machine learning-based clinical decision-support systems and help discover crucial survival-related parameters in patients with breast cancer.

The results of earlier breast cancer survival prediction research using comparable datasets, such as the METABRIC and SEER datasets, were also compared with the performance of the suggested multiclass survival prediction framework. Previous research utilizing the METABRIC dataset mostly concentrated on employing individual machine learning or deep learning models to estimate particular outcomes or predict binary survival. Strong classification performance utilizing SVM, Neural Network, and ensemble-based techniques was reported in a number of research; however, many of these studies did not assess multiclass survival categories or look at how leakage-prone survival-related factors affected model performance. Similarly, using conventional machine learning algorithms and statistical techniques, studies based on the SEER dataset frequently focused on long-term survival prediction, mortality risk estimation, or

recurrence analysis. In contrast to these other studies, the current study used NN, SVM, RF, NB, and KNN algorithms in a comparative multiclass classification framework to predict three survival outcome categories: "Living," "Died of Disease," and "Died of Other Causes." The evaluation results became more realistic and comparable to earlier survival prediction studies once survival-related factors were eliminated to decrease target leakage. The results showed that while RF had a lower recall for the "Died of Disease" class, NB and KNN performed more steadily under leakage-controlled conditions. This comparison emphasizes the significance of leakage avoidance, fair evaluation, and open reporting in breast cancer survival prediction research and implies that multiclass survival prediction using clinically realistic predictors is still a difficult challenge.

4. CONCLUSION

Using 1,980 patient records from the METABRIC dataset, this study created a multi-algorithm machine learning prototype for multiclass breast cancer survival status categorization. Patients were divided into three groups in the prediction task: Living, Died of Disease, and Died of Other Causes. After eliminating leakage-prone variables linked to survival outcomes, a comparative analysis was carried out utilizing NN, SVM, RF, NB, and KNN algorithms. The NN model obtained weighted average precision, recall, and F1-score values of 0.72, 0.70, and 0.71, respectively, based on the revised evaluation findings. While NB and KNN obtained weighted average F1-score values of 0.93 and 0.92, respectively, the RF model obtained 0.83 precision, 0.78 recall, and 0.75 F1-score. These results show that under leakage-controlled evaluation settings, NB and KNN performed more consistently. The developed system has not yet undergone clinical validation, external validation, clinician evaluation, or usability testing in actual healthcare settings; it is simply meant to be a prototype decision-support and monitoring application. Furthermore, the use of a single public dataset and the possible impact of target leakage in previous trials continue to limit the current findings. Consequently, more stringent feature selection techniques, external datasets, and clinically realistic validation settings should be used to reassess the presented results. Before considering actual healthcare application, future research should concentrate on external validation, explainability analysis, larger and more diversified datasets, and clinician-centered usability evaluation.

ACKNOWLEDGMENT

The author would like to express their sincere gratitude to Aisyah University for the support and facilities provided during the completion of this research.

REFERENCES

[1] J. Potsangbam and S. Shuleenda Devi, "Classification of Breast Cancer Histopathological Images Using Transfer Learning with DenseNet121," *Procedia Comput. Sci.*, vol. 235, pp. 1990–1997, 2024, doi: 10.1016/j.procs.2024.04.188.

[2] K. Mallikharjuna Rao, G. Saikrishna, and K. Supriya, "Data preprocessing techniques: emergence and selection towards machine learning models - a practical review using HPA dataset," *Multimed. Tools Appl.*, vol. 82, no. 24, pp. 37177–37196, 2023, doi: 10.1007/s11042-023-15087-5.

[3] A. Chandra, A. S. Kadam, P. Jain, M. Kumawat, H. Y. Patil, and M. A. Gawas, "Breast Cancer Classification using Metaheuristic Optimization and Machine Learning," *2024 3rd Int. Conf. Artif. Intell. Internet Things, AIIoT 2024*, no. AIIoT, pp. 1–4, 2024, doi: 10.1109/AIIoT58432.2024.10574626.

[4] J. Ma, S. Chen, and J. Sun, "Breast Cancer Diagnostic Model Based on RIME-SVM," *2024 IEEE 6th Int. Conf. Power, Intell. Comput. Syst. ICPICS 2024*, pp. 402–406, 2024, doi: 10.1109/ICPICS62053.2024.10796222.

[5] S. Benghazouani, S. Nouh, and A. Zakrani, "Optimizing breast cancer diagnosis: Harnessing the power of nature-inspired metaheuristics for feature selection with soft voting classifiers," *Int. J. Cogn. Comput. Eng.*, vol. 6, no. February 2024, pp. 1–20, 2025, doi: 10.1016/j.ijcce.2024.09.005.

[6] A. Kaur and S. Gupta, "Unveiling Precision in Breast Cancer Prediction with Random Forest and Decision Trees," *Proc. 5th Int. Conf. Smart Electron. Commun. ICOSEC 2024*, no. Icosec, pp. 1232–1236, 2024, doi: 10.1109/ICOSEC61587.2024.10722493.

[7] T. H. Rafi, R. M. Shubair, F. Farhan, M. Z. Hoque, and F. M. Quayyum, "Recent Advances in Computer-Aided Medical Diagnosis Using Machine Learning Algorithms with Optimization Techniques," *IEEE Access*, vol. 9, pp. 137847–137868, 2021, doi: 10.1109/ACCESS.2021.3108892.

[8] X. Xiong *et al.*, "Breast cancer: pathogenesis and treatments," *Signal Transduct. Target. Ther.*, vol. 10, no. 1, 2025, doi: 10.1038/s41392-024-02108-4.

[9] A. Choi *et al.*, "Development of a machine learning-based clinical decision support system to predict clinical deterioration in patients visiting the emergency department," *Sci. Rep.*, vol. 13, no. 1, pp. 1–10, 2023, doi: 10.1038/s41598-023-35617-3.

[10] M. Umer, S. Sadiq, H. Karamti, W. Karamti, R. Majeed, and M. Nappi, "IoT Based Smart Monitoring of Patients' with Acute Heart Failure," *Sensors*, vol. 22, no. 7, pp. 1–18, 2022, doi: 10.3390/s22072431.

[11] Z. Zulkifli, F. L. Gaol, T. Agung, and B. Widodo, "Software Testing Integration-Based Model (I-BM) Framework for Recognizing Measure Fault Output Accuracy Using Machine Learning Approach," *Int. J. Softw. Eng. Knowl. Eng.*, 2023.

[12] L. Balkenende, J. Teuwen, and R. M. Mann, "Application of Deep Learning in Breast Cancer Imaging," *Semin. Nucl. Med.*, vol. 52, no. 5, pp. 584–596, 2022, doi: 10.1053/j.semnuclmed.2022.02.003.

[13] V. Nemade and V. Fegade, "Machine Learning Techniques for Breast Cancer Prediction," *Procedia Comput. Sci.*, vol. 218, no. 2022, pp. 1314–1320, 2022, doi: 10.1016/j.procs.2023.01.110.

[14] Y. Supriya and R. Chengoden, "Breast cancer prediction using Shapely and Game theory in Federated Learning environment," *IEEE Access*, vol. 12, no. June, pp. 123018–123037, 2024, doi: 10.1109/ACCESS.2024.3424934.

[15] Y. Du, C. McNestry, L. Wei, A. M. Antoniadi, F. M. McAuliffe, and C. Mooney, "Machine learning-based clinical decision support systems for pregnancy care: A systematic review," *Int. J. Med. Inform.*, vol. 173, no. January, p. 105040, 2023, doi: 10.1016/j.ijmedinf.2023.105040.

[16] V. Nimbalkar and R. Kalra, "A Human-Centered Approach to Interpretable Machine Learning in Clinical Decision Support Systems," *2025 2nd Int. Conf. Artif. Intell. Innov. Healthc. Ind.*, pp. 1–5, doi: 10.1109/ICAIIHI67124.2025.11403473.

[17] R. Alharbi, "Breast Cancer Gene Expression Profiles (METABRIC)." [Online]. Available: https://www.kaggle.com/datasets/raghadalharbi/breast-cancer-gene-expression-profiles-metabric

[18] P. Singh and S. Verma, "Multi-classifier model for software fault prediction," *Int. Arab J. Inf. Technol.*, vol. 15, no. 5, pp. 912–919, 2018.

[19] A. M. Ismael and A. Şengür, "Deep learning approaches for COVID-19 detection based on chest X-ray images," *Expert Syst. Appl.*, vol. 164, no. March 2020, 2021, doi: 10.1016/j.eswa.2020.114054.

[20] S. Goyal, *Handling Class-Imbalance with KNN (Neighbourhood) Under-Sampling for Software Defect Prediction*, vol. 55, no. 3. Springer Netherlands, 2022. doi: 10.1007/s10462-021-10044-w.

[21] S. S. Cross, R. F. Harrison, and R. L. Kennedy, *Introduction to neural networks*, vol. 346, no. 8982. 1995. doi: 10.1016/S0140-6736(95)91746-2.

[22] I. Aydin, M. Karaköse, and E. Akin, "Artificial immune based support vector machine algorithm for fault diagnosis of induction motors," *Int. Aegean Conf. Electr. Mach. Power Electron. Electromotion ACEMP'07 Electromotion'07 Jt. Conf.*, pp. 217–221, 2007, doi: 10.1109/ACEMP.2007.4510505.

[23] S. Goyal, "Handling Class-Imbalance with KNN (Neighbourhood) Under-Sampling for Software Defect Prediction," *Artif. Intell. Rev.*, vol. 55, no. 3, pp. 2023–2064, 2022, doi: 10.1007/s10462-021-10044-w.

[24] E. Dritsas and M. Trigka, "Data-Driven Machine-Learning Methods for Diabetes Risk Prediction," *Sensors*, vol. 22, no. 14, 2022, doi: 10.3390/s22145304.

[25] S. A. Ali, N. R. Roy, and D. Raj, "Software Defect Prediction using Machine Learning," *Proc. 17th INDIACom; 2023 10th Int. Conf. Comput. Sustain. Glob. Dev. INDIACom 2023*, no. Icoei, pp. 639–642, 2023.

[26] O. Esteban, C. J. Markiewicz, R. W. Blair, C. A. Moodie, and ..., "fMRIPrep: a robust preprocessing pipeline for functional MRI," *Nat Methods*, vol. 16, no. 1, pp. 111–116, 2019, doi: https://doi.org/10.1038/s41592-018-0235-4.

[27] T. Byun, V. Sharma, A. Vijayakumar, and ..., "Input prioritization for testing neural networks," *IEEE Int. Conf. Artif. Intell. Test.*, 2019.

[28] H. Wang, R. Czerminski, and A. C. Jamieson, "Neural Networks and Deep Learning," *Mach. Age Cust. Insight*, pp. 91–101, 2021, doi: 10.1108/978-1-83909-694-520211010.

[29] Y. Niu, "Adaptive two-SVM multi-objective cuckoo search algorithm for software defect prediction," *Int. J. Comput. Sci. Math.*, vol. 9, no. 6, pp. 547–554, 2018, doi: 10.1504/IJCSM.2018.096327.

[30] T. Meeradevi, S. Sasikala, L. Murali, N. Manikandan, and K. Ramaswamy, "Lung cancer detection with machine learning classifiers with multi-attribute decision-making system and deep learning model," *Sci. Rep.*, vol. 15, no. 1, pp. 1–19, 2025, doi: 10.1038/s41598-025-88188-w.

[31] Y. Xu, C. Peng, M. Li, Y. Li, and S. Du, "Pyramid Feature Attention Network for Monocular Depth Prediction," *Proc. - IEEE Int. Conf. Multimed. Expo*, 2021, doi: 10.1109/ICME51207.2021.9428446.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[32] S. Ruuska, W. Hämäläinen, S. Kajava, M. Mughal, P. Matilainen, and J. Mononen, "Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle," *Behav. Processes*, vol. 148, pp. 56–62, 2018, doi: 10.1016/j.beproc.2018.01.004.

[33] M. Mustaqeem, "Principal component based support vector machine (PC-SVM): a hybrid technique for software defect detection," *Cluster Comput.*, vol. 24, no. 3, pp. 2581–2595, 2021, doi: 10.1007/s10586-021-03282-8.

[34] Y. Soe, "Software defect prediction using random forest algorithm," 2018. doi: 10.1109/SEATUC.2018.8788881.

[35] Z. Zulkifli, F. A. Makkiyah, D. Antoni, F. Fitriana, T. Jamaan, and A. Taufik, "Multi-Algorithm to Measure the Accuracy Level of Diabetes Status Prediction," *J. Appl. Data Sci.*, vol. 5, no. 2, pp. 736–746, 2024, doi: 10.47738/jads.v5i2.250.

[36] K. Kalasampath, K. N. Spoorthi, S. Sajeev, S. S. Kuppa, K. Ajay, and A. Maruthamuthu, "A Literature Review on Applications of Explainable Artificial Intelligence (XAI)," *IEEE Access*, vol. 13, no. February, pp. 41111–41140, 2025, doi: 10.1109/ACCESS.2025.3546681.

[37] M. Sharma Timilsina, S. Sen, B. Uprety, V. B. Patel, P. Sharma, and P. N. Sheth, "Prediction of HHV of fuel by Machine learning Algorithm: Interpretability analysis using Shapley Additive Explanations (SHAP)," *Fuel*, vol. 357, no. PA, p. 129573, 2024, doi: 10.1016/j.fuel.2023.129573.

[38] K. Davagdorj, M. Li, and K. H. Ryu, "Local Interpretable Model-Agnostic Explanations of Predictive Models for Hypertension BT - Advances in Intelligent Information Hiding and Multimedia Signal Processing," J.-S. Pan, J. Li, K. H. Ryu, Z. Meng, and A. Klasnja-Milicevic, Eds., Singapore: Springer Singapore, 2021, pp. 426–433.