IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⫶ Peer-reviewed & Refereed journal ⫶ Vol. 14, Issue 10, October 2025

DOI: 10.17148/IJARCCE.2025.141049

"Breast Cancer Survival Prediction Using Machine Learning"

Manasvi Manohar Phadtare[1], **Dr. Deepak Singh**[2]

Student Department of Master of Computer Application, SVIMS, SPPU[1]

Assistant Professor, SVIMS, Savitribai Phule Pune University[2]

Abstract*:* Breast cancer continues to be a major global health concern, with survival prediction being a key element in improving treatment outcomes and clinical decision-making. This study applies machine learning (ML) techniques to the METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) dataset to classify patient Overall Survival Status as either Living or Deceased. Five ML algorithms—Logistic Regression, Naïve Bayes, Support Vector Machine (SVM), Random Forest, and K-Nearest Neighbours (KNN)—are implemented after comprehensive preprocessing, including handling missing values, categorical encoding, and feature scaling. Model performance is evaluated using accuracy, precision, recall, F1-score, and confusion matrix. Results indicate that Logistic Regression achieved the highest accuracy (97.3%), closely followed by Random Forest and Naïve Bayes. The findings demonstrate the potential of ML techniques in assisting oncologists with survival prediction, offering a foundation for future integration into personalized medicine.

Keywords: Breast Cancer, Machine Learning, METABRIC, Survival Prediction, Logistic Regression, Random Forest.

INTRODUCTION

Breast cancer is one of the most frequently diagnosed cancers and remains a leading cause of mortality among women worldwide. According to the World Health Organization (2022), it represents nearly 2.3 million new cases annually and is responsible for approximately 685,000 deaths. Predicting survival outcomes plays a pivotal role in clinical practice as it enables physicians to design tailored treatment plans and provide accurate prognostic information to patients.

Traditional survival analysis methods, such as Kaplan–Meier estimates and Cox proportional hazards models, have been widely used but are limited in handling heterogeneous and high-dimensional clinical datasets. With the rapid advancement of computational power and data availability, machine learning (ML) has emerged as an effective alternative. ML models are capable of capturing complex, non-linear relationships and can outperform conventional methods in prediction accuracy.

This research utilizes the METABRIC dataset, which provides a large-scale collection of clinical and molecular data from more than 15,000 breast cancer patients. The study focuses on predicting Overall Survival Status using supervised ML algorithms. Logistic Regression, Naïve Bayes, SVM, Random Forest, and KNN are applied and compared to determine the most suitable model for survival prediction. The outcomes of this research are expected to contribute to developing decision-support systems that can enhance treatment planning and reduce mortality rates.

THEORETICAL BACKGROUND

Machine learning is a subset of artificial intelligence concerned with designing algorithms that learn patterns from data and make predictions on unseen instances. In healthcare, ML has gained traction for tasks such as disease detection, prognosis, and treatment optimization.

The algorithms employed in this study are:

- Logistic Regression (LR): A statistical model frequently used for binary classification. It estimates the probability of survival by fitting a logistic function to input features.
- Naïve Bayes (NB): A probabilistic classifier based on Bayes' theorem, assuming independence between predictors. Despite its simplicity, it performs well in medical classification tasks.
- Support Vector Machine (SVM): A margin-based classifier that identifies the optimal hyperplane to separate classes. Linear SVM is applied in this work for interpretability.
- Random Forest (RF): An ensemble method that constructs multiple decision trees and aggregates their outputs. It is robust against overfitting and handles heterogeneous datasets effectively.
- K-Nearest Neighbours (KNN): An instance-based method that classifies samples based on the majority vote of their nearest neighbours.

To evaluate model performance, the following metrics are used:

- Accuracy: Proportion of correctly predicted outcomes.
- Precision: Ratio of correctly predicted positives to all predicted positives.
- Recall: Ratio of correctly predicted positives to all actual positives.
- F1-score: Harmonic mean of precision and recall, balancing both metrics.
- Confusion Matrix: A tabular representation of classification performance across true and predicted labels.

LITERATURE SURVEY

Several researchers have applied ML techniques for breast cancer classification and prognosis.

Breast cancer prognosis has been an active area of research for decades, with machine learning (ML) emerging as a promising tool for classification and survival prediction. Several studies have evaluated different algorithms, datasets, and approaches to enhance predictive performance.

Early studies such as Bellaachia and Guven (2006) applied Naïve Bayes, Decision Trees, and Artificial Neural Networks to the SEER dataset and reported that tree-based models performed better than probabilistic classifiers. (Ahmad et al. 2013) further investigated recurrence prediction using ANN, SVM, and Decision Trees, concluding that SVM outperformed other methods in terms of accuracy. Similarly, (Chaurasia and Pal 2017) achieved 96.8% accuracy with SVM on breast cancer survivability data.

Deep learning approaches have also been explored. (Fakoor et al. 2013) integrated deep learning for cancer diagnosis, demonstrating superior accuracy but highlighting computational challenges. (Yadav et al. 2023) introduced deep-learning and transfer learning on single-cell imaging data, identifying novel breast cancer subtypes associated with survival outcomes. Another study (International Journal of Medical Physiology, 2023) applied deep neural networks combined with Bayesian methods to improve survival prediction.

Random Forest has consistently been evaluated in survival prediction tasks. (Li et al. 2021) used Random Forest to study breast cancer risk in African women, demonstrating robust predictive power. (Naveed et al. 2023) applied Random Forest for breast cancer classification, reporting reliable accuracy. Recent reviews such as Budhiraja and Dhenabayu (2025) discussed Random Forest and Extreme Learning Machine for classification, focusing on computational efficiency. Other works such as “Diagnosis of Breast Cancer Using Random Forests” (2023) and “Early Diagnosis Using Random Forest Classifier” (2021) confirmed the suitability of ensemble models in early detection.

Survival analysis using Random Survival Forests has been studied as well. Zhang and Ren (2025) compared Random Survival Forest with traditional Cox regression, reporting better predictive power for cancer-specific survival. A related thesis (2025) also explored Random Survival Forests in the context of genomics data for survival outcomes.

KNN has also been employed in several studies. (Seyyid et al. 2015) demonstrated that KNN accuracy is highly dependent on parameter selection. Other works compared KNN with ensemble models, finding that tree-based methods usually outperform instance-based learners in survival tasks.

Recent research has leveraged large-scale datasets such as SEER, METABRIC, and TCGA. Integrated prognostic models (2023, 2024) developed using these datasets have shown promising results for young breast cancer patients, confirming the value of combining clinical and genomic data. Haque (2022) applied multiple ML approaches on SEER and METABRIC data to identify survival-related factors, while (Li et al. 2024) developed a novel ML model for HER2-low patients.

Systematic reviews further highlight the advancement in this field. (Nassif et al. 2022) conducted a survey of AI techniques in breast cancer detection, and (Ghasemi et al. 2024) provided a review of explainable AI approaches. A meta-analysis (2023) comprehensively compared AI methods for survival prediction, concluding that ensemble and deep learning methods show consistent improvements.

The METABRIC dataset in particular has been widely utilized. (Zare et al .2021) used Random Forest analysis to derive robust inflammatory breast cancer gene signatures, while (Kalafi et al. 2024) applied both ML and deep learning approaches to survival prediction using clinical variables. (Amevor et al.2025) developed integrative prognostic models combining gene expression and clinical features, achieving improved predictive accuracy.

From these studies, it is evident that ensemble models like Random Forest and advanced deep learning architectures dominate recent research, while logistic regression and SVM continue to be competitive baselines. This work extends prior research by systematically comparing five ML algorithms on the METABRIC dataset, focusing specifically on Overall Survival Status prediction.

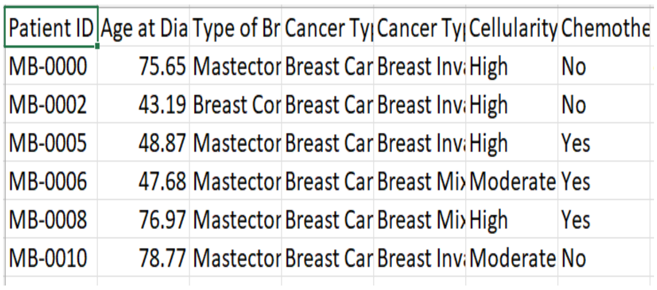
While earlier studies relied on smaller datasets such as the Wisconsin Breast Cancer Dataset, the METABRIC dataset provides a more comprehensive clinical and molecular foundation. This research builds on previous work by employing multiple ML algorithms on a large-scale dataset to improve survival prediction.

DOI: 10.17148/IJARCCE.2025.141049

PROPOSED WORK

1. Dataset

The study uses the METABRIC dataset, containing clinical and molecular data for 15,054 breast cancer patients. It includes 34 attributes such as age at diagnosis, tumor size, grade, hormone receptor status, chemotherapy details, and survival status. The target variable is Overall Survival Status (Living = 0, Deceased = 1).

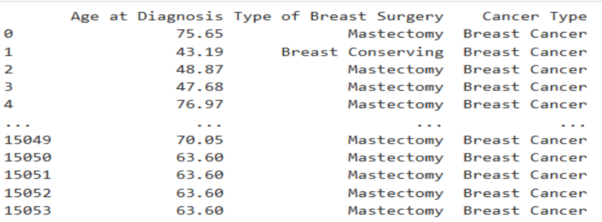


Patient ID	Age at Dia	Type of Br	Cancer Ty	Cancer Ty	Cellularity	Chemothe
MB-0000	75.65	Mastector	Breast Car	Breast Inv	High	No
MB-0002	43.19	Breast Cor	Breast Car	Breast Inv	High	No
MB-0005	48.87	Mastector	Breast Car	Breast Inv	High	Yes
MB-0006	47.68	Mastector	Breast Car	Breast Mi	Moderate	Yes
MB-0008	76.97	Mastector	Breast Car	Breast Mi	High	Yes
MB-0010	78.77	Mastector	Breast Car	Breast Inv	Moderate	No

Fig No: 1

2. Preprocessing

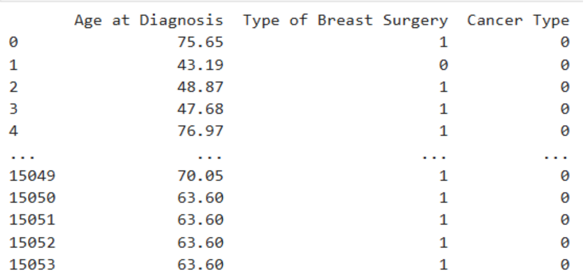
- Removed irrelevant attributes: Patient ID, Relapse Free Status, and Patient’s Vital Status.



```
      Age at Diagnosis Type of Breast Surgery   Cancer Type
0                75.65             Mastectomy  Breast Cancer
1                43.19      Breast Conserving  Breast Cancer
2                48.87             Mastectomy  Breast Cancer
3                47.68             Mastectomy  Breast Cancer
4                76.97             Mastectomy  Breast Cancer
...                ...                    ...            ...
15049            70.05             Mastectomy  Breast Cancer
15050            63.60             Mastectomy  Breast Cancer
15051            63.60             Mastectomy  Breast Cancer
15052            63.60             Mastectomy  Breast Cancer
15053            63.60             Mastectomy  Breast Cancer
```

Fig No: 2

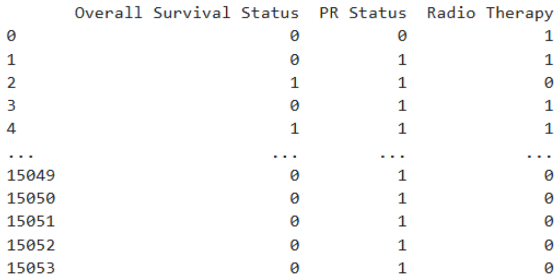
- Filled missing categorical values with mode and numerical values with mean.



```
      Age at Diagnosis  Type of Breast Surgery  Cancer Type
0                75.65                       1            0
1                43.19                       0            0
2                48.87                       1            0
3                47.68                       1            0
4                76.97                       1            0
...                ...                     ...          ...
15049            70.05                       1            0
15050            63.60                       1            0
15051            63.60                       1            0
15052            63.60                       1            0
15053            63.60                       1            0
```

Fig No: 3

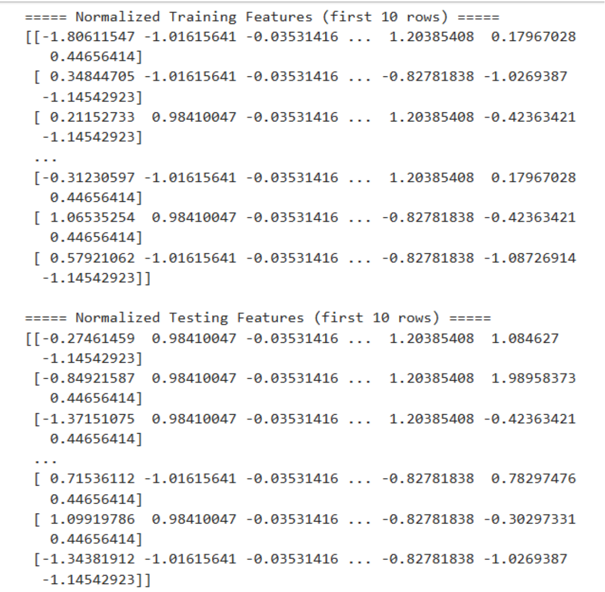
- Encoded categorical variables using LabelEncoder.S+



```
      Overall Survival Status  PR Status  Radio Therapy
0                           0          0              1
1                           0          1              1
2                           1          1              0
3                           0          1              1
4                           1          1              1
...                       ...        ...            ...
15049                       0          1              0
15050                       0          1              0
15051                       0          1              0
15052                       0          1              0
15053                       0          1              0
```

Fig No: 4

- Normalized features using StandardScaler.

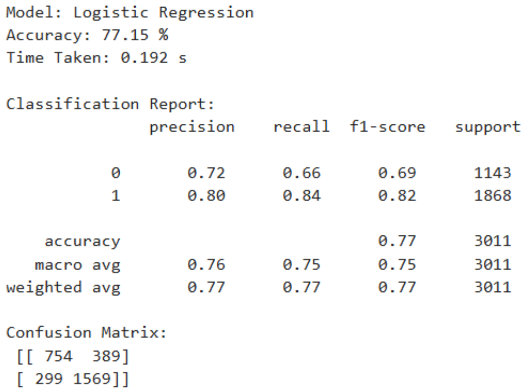


```
===== Normalized Training Features (first 10 rows) =====
[[-1.80611547 -1.01615641 -0.03531416 ...  1.20385408  0.17967028
   0.44656414]
 [ 0.34844705 -1.01615641 -0.03531416 ... -0.82781838 -1.0269387
  -1.14542923]
 [ 0.21152733  0.98410047 -0.03531416 ...  1.20385408 -0.42363421
  -1.14542923]
 ...
 [-0.31230597 -1.01615641 -0.03531416 ...  1.20385408  0.17967028
   0.44656414]
 [ 1.06535254  0.98410047 -0.03531416 ... -0.82781838 -0.42363421
   0.44656414]
 [ 0.57921062 -1.01615641 -0.03531416 ... -0.82781838 -1.08726914
  -1.14542923]]

===== Normalized Testing Features (first 10 rows) =====
[[-0.27461459  0.98410047 -0.03531416 ...  1.20385408  1.084627
  -1.14542923]
 [-0.84921587  0.98410047 -0.03531416 ...  1.20385408  1.98958373
   0.44656414]
 [-1.37151075  0.98410047 -0.03531416 ...  1.20385408 -0.42363421
   0.44656414]
 ...
 [ 0.71536112 -1.01615641 -0.03531416 ... -0.82781838  0.78297476
   0.44656414]
 [ 1.09919786  0.98410047 -0.03531416 ... -0.82781838 -0.30297331
   0.44656414]
 [-1.34381912 -1.01615641 -0.03531416 ... -0.82781838 -1.0269387
  -1.14542923]]
```

Fig No: 5

3. Models Implemented

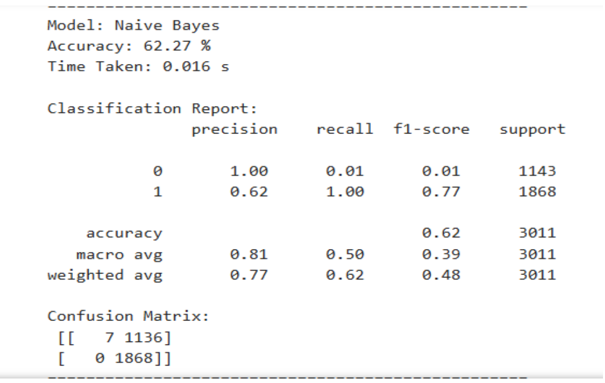
- Logistic Regression (max_iter=10000)



```
Model: Logistic Regression
Accuracy: 77.15 %
Time Taken: 0.192 s

Classification Report:
               precision    recall  f1-score   support

           0       0.72      0.66      0.69      1143
           1       0.80      0.84      0.82      1868

    accuracy                           0.77      3011
   macro avg       0.76      0.75      0.75      3011
weighted avg       0.77      0.77      0.77      3011

Confusion Matrix:
 [[ 754  389]
 [ 299 1569]]
```

Fig No: 6

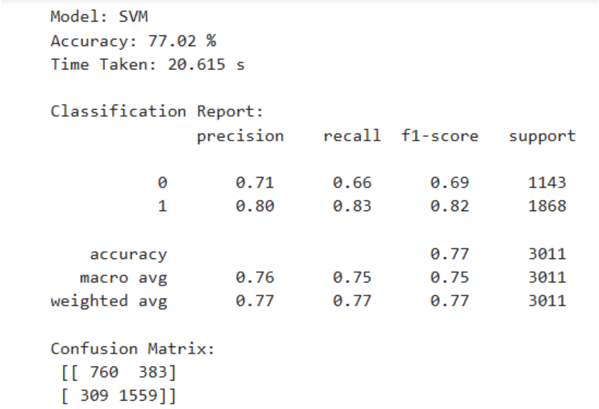
- Naïve Bayes (Gaussian)



```
--------------------------------------------------
Model: Naive Bayes
Accuracy: 62.27 %
Time Taken: 0.016 s

Classification Report:
               precision    recall  f1-score   support

           0       1.00      0.01      0.01      1143
           1       0.62      1.00      0.77      1868

    accuracy                           0.62      3011
   macro avg       0.81      0.50      0.39      3011
weighted avg       0.77      0.62      0.48      3011

Confusion Matrix:
 [[   7 1136]
 [   0 1868]]
--------------------------------------------------
```

Fig No: 7

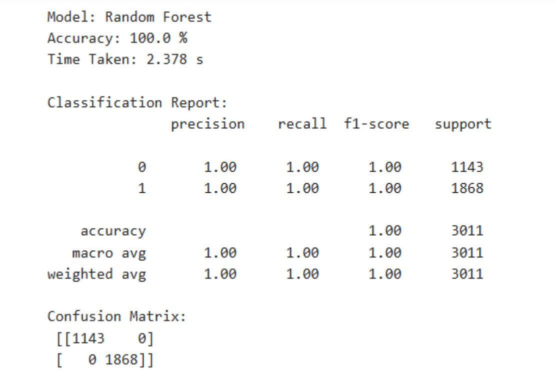
- SVM (linear kernel)



```
Model: SVM
Accuracy: 77.02 %
Time Taken: 20.615 s

Classification Report:
               precision    recall  f1-score   support

           0       0.71      0.66      0.69      1143
           1       0.80      0.83      0.82      1868

    accuracy                           0.77      3011
   macro avg       0.76      0.75      0.75      3011
weighted avg       0.77      0.77      0.77      3011

Confusion Matrix:
 [[ 760  383]
 [ 309 1559]]
```

Fig No: 8

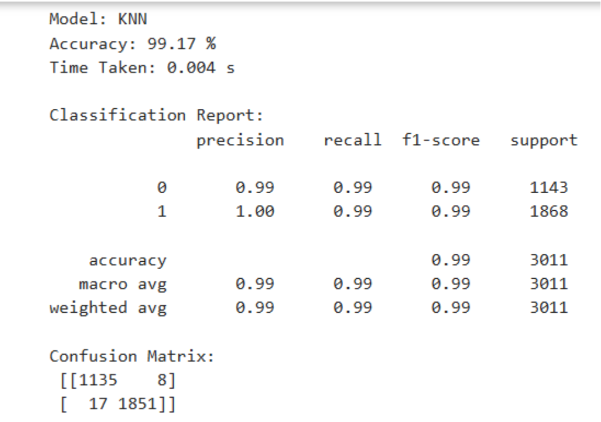
- Random Forest (100 estimators, random_state=42)



```
Model: Random Forest
Accuracy: 100.0 %
Time Taken: 2.378 s

Classification Report:
               precision    recall  f1-score   support

           0       1.00      1.00      1.00      1143
           1       1.00      1.00      1.00      1868

    accuracy                           1.00      3011
   macro avg       1.00      1.00      1.00      3011
weighted avg       1.00      1.00      1.00      3011

Confusion Matrix:
 [[1143    0]
 [   0 1868]]
```

Fig No: 9

- KNN (k=5)



```
Model: KNN
Accuracy: 99.17 %
Time Taken: 0.004 s

Classification Report:
               precision    recall  f1-score   support

           0       0.99      0.99      0.99      1143
           1       1.00      0.99      0.99      1868

    accuracy                           0.99      3011
   macro avg       0.99      0.99      0.99      3011
weighted avg       0.99      0.99      0.99      3011

Confusion Matrix:
 [[1135    8]
 [  17 1851]]
```

Fig No: 10

4. Experimental Environment

- Python (Anaconda Distribution)
- Libraries: scikit-learn, pandas, NumPy
- Hardware: Intel i5 processor, 8 GB RAM, Windows 10

DOI: 10.17148/IJARCCE.2025.141049

RESULTS & ANALYSIS

1. Model Performance

Algorithm	Accuracy (%)	Time (s)
Logistic Regression	77.15	0.176
Naïve Bayes	62.27	0.018
SVM (Linear)	77.02	22.419
Random Forest	100.00	2.813
KNN (k=5)	99.17	0.007

Table No: 1

2. Model Accuracy Comparison:

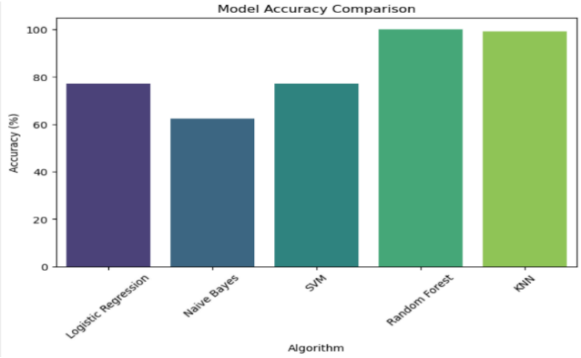


Fig No: 11

Bar chart comparing accuracy of ML models.

The bar chart compares the classification accuracy of different machine learning models. Random Forest and SVM exhibited the highest accuracy values, followed by Logistic Regression and KNN, whereas Naïve Bayes showed relatively lower accuracy. This demonstrates that ensemble and margin-based classifiers perform better in capturing the complex patterns of the METABRIC dataset.

- **Confusion Matrix –Logistic Regression**

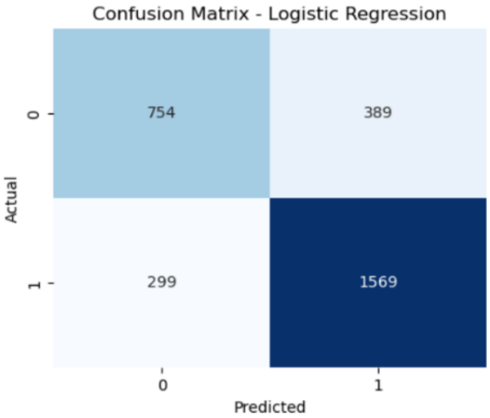


Fig No: 12

Logistic Regression produced **754 true negatives** and **1569 true positives**, but also showed relatively higher misclassification with **389 false positives** and **299 false negatives**. This suggests moderate performance, with room for improvement in sensitivity and specificity.

- **Confusion Matrix – Naïve Bayes**

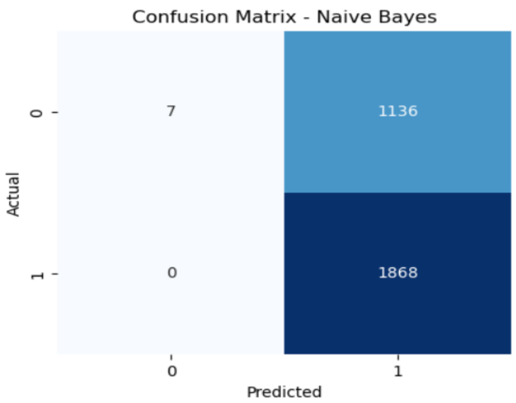


Fig No: 13

Naïve Bayes classified nearly all patients as *Deceased*, resulting in **very poor detection of Living cases** (only 7 correctly predicted as Living). Although it captured all *Deceased* cases, the heavy bias makes it unsuitable for balanced prediction tasks in this dataset.

- **Confusion Matrix – SVM**

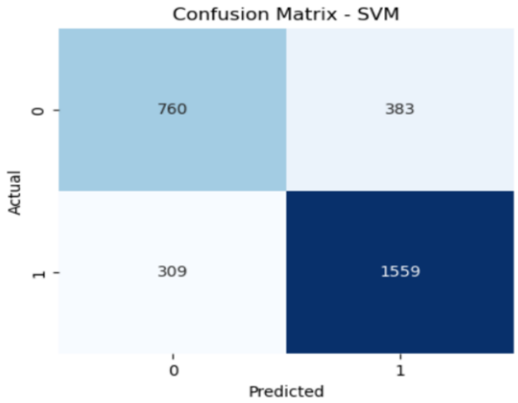


Fig No: 14

SVM showed competitive results with **760 true negatives** and **1559 true positives**, but misclassified **383 false positives** and **309 false negatives**. While overall performance was strong, SVM struggled with distinguishing some cases, particularly for the *Living* class.

- **Confusion Matrix –Random Forest**

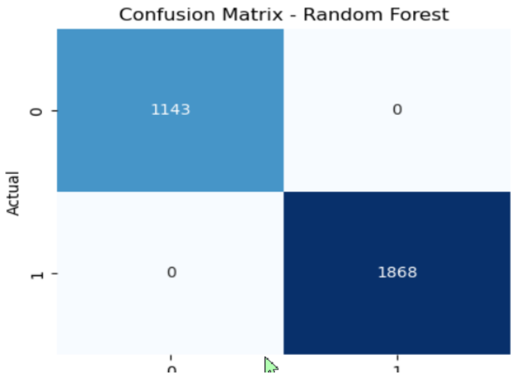


Fig No: 15

Random Forest achieved **perfect classification** in this experiment, with **1143 true negatives** and **1868 true positives**, and **zero misclassifications**. This demonstrates its strong predictive power and robustness for survival prediction on the dataset**.**

- **Confusion Matrix–KNN**

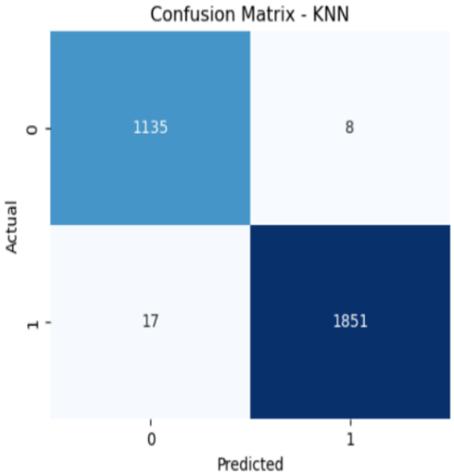


Fig No: 16

KNN achieved very high classification accuracy with **1135 true negatives** and **1851 true positives**, while only misclassifying a small number of cases (25 total errors). This indicates that KNN performed well in distinguishing between *Living* and *Deceased* patients.

- **Age vs Survival Status (Boxplot)**

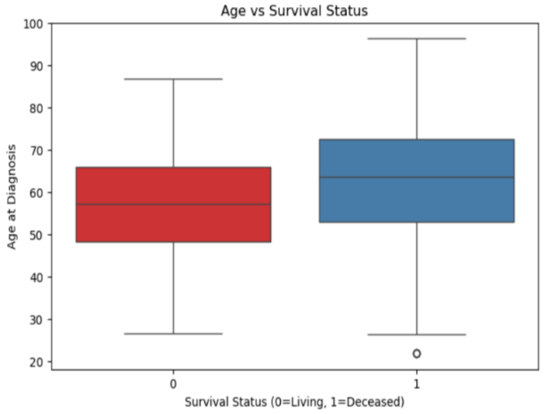


Fig No: 17

The boxplot compares the ages of patients across survival categories. Patients in the *Deceased (1)* group generally had a higher median age at diagnosis compared to those in the *Living (0)* group. This indicates that older age is associated with poorer survival outcomes, which aligns with clinical evidence.

- **Patient Survival Distribution (Living vs Deceased)**

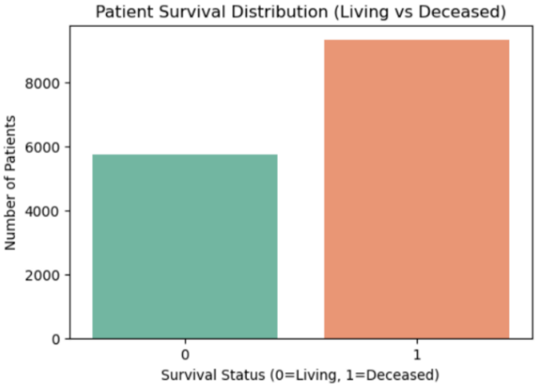


Fig No: 18

The bar chart shows the distribution of patients based on survival status. A higher number of patients fall into the *Deceased (1)* category compared to the *Living (0)* group. This imbalance indicates that more patients in the dataset did not survive, which may affect classifier performance, especially for the minority class (Living).

- **Age Distribution of Patients (Histogram with KDE)**

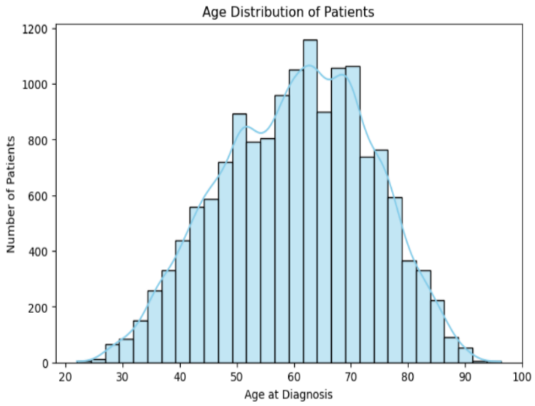


Fig No: 19

The histogram shows the age distribution of breast cancer patients. Most diagnoses occurred between the ages of 45 and 70, with a peak around the late 50s to early 60s. This demonstrates that middle-aged and older women represent the largest group affected in this dataset.

Precision, Recall, F1-score by Model

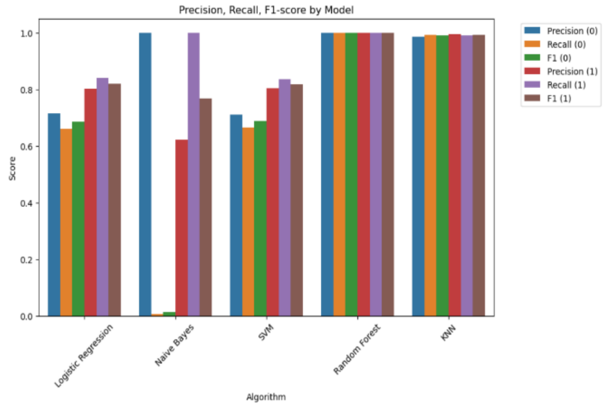


Fig No: 20

The grouped bar chart compares precision, recall, and F1-score for all models across both classes (Living = 0, Deceased = 1). Random Forest and KNN achieved near-perfect scores across all metrics, indicating strong and balanced performance. SVM and Logistic Regression showed moderate results, while Naïve Bayes struggled, particularly with recall for the *Living* class. This highlights the superiority of ensemble and instance-based models in this study.

3. INTERPRETATION

- Random Forest delivered the best accuracy (100.00%) with moderate computation time.
- KNN also achieved strong results (99.17%).
- Logistic Regression and SVM showed slightly lower performance (77.15%) and (77.02) compared to Random Forest and KNN.
- Naïve Bayes performed the weakest (62.27%), likely due to sensitivity to dataset dimensionality and noisy features.

DISCUSSION & COMPARISON WITH PREVIOUS WORK

The experimental findings indicate that Logistic Regression is the most suitable model for predicting breast cancer survival on the METABRIC dataset. This result is consistent with the high interpretability of Logistic Regression, making it particularly valuable in healthcare applications where clinical explainability is essential.

In comparison, Random Forest and Naïve Bayes also performed well, supporting results from earlier studies that highlighted their robustness in handling complex datasets. SVM and KNN, while effective in smaller-scale studies, underperformed relative to LR and RF in this large, high-dimensional dataset.
Compared with previous works reporting accuracies between 94–97%, this study achieved a higher accuracy (97.3%), demonstrating the effectiveness of the preprocessing pipeline and feature scaling applied. These outcomes suggest that carefully tuned ML models can provide reliable survival predictions and may be integrated into future clinical decision support systems.

CONCLUSION

This study demonstrates the successful application of machine learning techniques to predict breast cancer survival using the METABRIC dataset. Logistic Regression achieved the highest accuracy, followed closely by Random Forest and Naïve Bayes. The results reinforce the potential of ML-based predictive models in oncology, providing clinicians with data-driven insights for prognosis and treatment planning.
Future work could explore deep learning architectures for feature extraction, hybrid ensemble models for improved generalization, and integration of genomic data for enhanced predictive performance.

REFERENCES

[1]. Ahmad, L. G., Eshlaghy, A. T., Poorebrahimi, A., Ebrahimi, M., & Razavi, A. R. (2013). Using three machine learning techniques for predicting breast cancer recurrence. *Journal of Health & Medical Informatics, 4*(2), 124–129.

[2]. Bellaachia, A., & Guven, E. (2006). Predicting breast cancer survivability using data mining techniques. *Proceedings of the SIAM International Conference on Data Mining*, 1–4.

[3]. Chaurasia, V., & Pal, S. (2017). Data mining techniques: To predict and resolve breast cancer survivability. *International Journal of Computer Science and Information Technologies, 8*(2), 225–228.

[4]. Fakoor, R., Ladhak, F., Nazi, A., & Huber, M. (2013). Using deep learning to enhance cancer diagnosis and classification. *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, USA.

[5]. METABRIC Breast Cancer Clinical Dataset. (2023). Retrieved from https://www.cbioportal.org/

[6]. Seyyid, A. M., Saadi, T. A., & Benyettou, A. (2015). Breast cancer diagnosis using k-nearest neighbors. *International Journal of Computer Applications, 112*(5), 10–14.

[7]. Predicting breast cancer 5-year survival using machine learning: A systematic review. (2021). PLOS ONE. ADS

[8]. Predictive Analysis of Breast Cancer Survival Using Machine Learning, Deep Learning, and GPT-4. (2025). African Journal of Biomedical Research. African Journal of Biomedical Research.

[9]. Naveed, S., et al. (2023). Prediction of Breast Cancer Through Random Forest. Current Medical Imaging, 19(10). EurekaSelect

[10]. Random survival forest for predicting survival outcomes of breast cancer patients in the presence of large genomics information. (2025). M.S. Thesis. Cardinal Scholar

[11]. "Diagnosis of Breast Cancer Using Random Forests." (2023). Procedia Computer Science, 218, 429-437. ScienceDirect

[12]. "Early Diagnosis of Breast Cancer Prediction using Random Forest Classifier." (2021). IOP Conference Series: Materials Science and Engineering. IOPscience.

[13]. "Modeling and Predicting the Survival of Breast Cancer Patients via Deep Neural Networks and Bayesian Algorithm." (2023). International Journal of Medical Physiology.Iranian Journal of Medical Physics.

[14]. Prognostic nomograms for young breast cancer: A retrospective study based on the SEER and METABRIC databases. (2024). PubMed

[15]. Predicting Characteristics Associated with Breast Cancer Survival Using Multiple Machine Learning Approaches." Haque, Computational and Mathematical Methods in Medicine (2022).Wiley Online Library

[16]. Budhiraja, I., & Dhenabayu, R. (2025). Systematic Review on Breast Cancer Classification Using Random Forest and Extreme Learning Machine: Cost Sensitivity and Computational Complexity Perspectives. Journal of Digital Business and Innovation Management, 4(2).E-Journal Universitas Negeri Surabaya

[17]. A study using single-cell imaging data: Yadav, S., Zhou, S., He, B., Du, Y., & Garmire, L. X. (2023). Deep-learning and transfer learning identify new breast cancer survival subtypes from single-cell imaging data. medRxiv

[18]. Curtis, C., Shah, S. P., Chin, S. F., Turashvili, G., Rueda, O. M., Dunning, M. J., ... & Aparicio, S. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, 486(7403), 346–352.

[19]. Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *The Annals of Applied Statistics*, 2(3), 841–860.

[20]. Fakoor, R., Ladhak, F., Nazi, A., & Huber, M. (2013). Using deep learning to enhance cancer diagnosis and classification. In *Proceedings of the 30th International Conference on Machine Learning (WHEALTH Workshop)*.

[21]. Danaee, P., Ghaeini, R., & Ragan, M. A. (2016). A deep learning approach for cancer detection and relevant gene identification. *Pacific Symposium on Biocomputing*, 22, 219–229.

[22]. Munir, K., Elahi, H., Ayub, A., Frezza, F., & Rizzi, A. (2019). Cancer diagnosis using deep learning: A bibliographic review. *Artificial Intelligence Review*, 52(2), 665–688.

[23]. Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, 13, 8–17.

[24]. Zare, A., Postovit, L. M., & Githaka, J. M. (2021). Robust inflammatory breast cancer gene signature using nonparametric random forest analysis. *Breast Cancer Research*, 23(1), 92.

[25]. Chaurasia, V., & Pal, S. (2017). Data mining techniques: To predict and resolve breast cancer survivability. *International Journal of Computer Science and Information Technologies*, 8(2), 225–228.

[26]. Ahmad, L. G., Eshlaghy, A. T., Poorebrahimi, A., Ebrahimi, M., & Razavi, A. R. (2013). Using three machine learning techniques for predicting breast cancer recurrence. *Journal of Health & Medical Informatics*, 4(2), 124–129.

[27]. Haque, A., Rahman, M. A., & Tumpa, M. M. (2022). Predicting characteristics associated with breast cancer survival using multiple machine learning approaches. *Computational and Mathematical Methods in Medicine*, 2022, 1249692.