# Comparative Analysis of Machine Learning Techniques in Breast Cancer Detection

Varun Chauhan
Information Technology
Thakur College of Engineering
and Technology
Mumbai, India
chauhanvarun10th03@gmail.c
om

Darsh Gupta
Information Technology
Thakur College of Engineering
and Technology
Mumbai, India
collegeid3250@gmail.com

Krishna Gupta
Information Technology
Thakur College of Engineering
and Technology
Mumbai, India
up.krishnagupta@gmail.com

*Abstract: Breast cancer remains one of the most prevalent and life-threatening diseases worldwide, making early and accurate detection critical for improving patient outcomes. With advancements in machine learning, predictive models have shown promise in assisting medical professionals by identifying patterns in diagnostic data. This study presents a comparative analysis of different machine learning techniques for breast cancer detection. A structured dataset is utilized to train and evaluate models based on multiple performance metrics, including accuracy, precision, recall, and F1-score. The study highlights the strengths and limitations of each approach, offering insights into their applicability in real-world clinical settings. The findings demonstrate that while deep learning models excel in feature extraction, traditional machine learning techniques offer efficiency and interpretability. This research contributes to the ongoing development of AI-driven diagnostic tools and provides a foundation for future enhancements in automated breast cancer detection.*

*Keywords: Breast Cancer, Machine Learning, Deep Learning, Predictive Analytics, Medical Diagnosis*

## 1. INTRODUCTION

### 1.1 Background & Significance

Breast cancer is one of the leading causes of cancer-related deaths worldwide, with millions of new cases diagnosed each year. Early detection plays a crucial role in improving survival rates, as timely treatment significantly increases the chances of recovery. Traditional diagnostic methods, such as mammography and biopsy, rely on human expertise and are often time-consuming, expensive, and prone to subjective interpretation.

With advancements in artificial intelligence (AI) and machine learning (ML), automated diagnostic tools have emerged as powerful alternatives for detecting breast cancer at an early stage. ML models can analyze complex patterns in medical imaging and structured patient data to assist radiologists and oncologists in making more accurate predictions. While various ML models have been explored for breast cancer diagnosis, selecting the most efficient approach remains a challenge. This study aims to address this by conducting a comparative analysis of different machine learning models, evaluating their strengths and limitations in predicting breast cancer.

### 1.2 Problem Statement

Despite numerous studies on the application of ML in breast cancer detection, a key challenge remains: identifying the most effective model for accurate and efficient diagnosis. While deep learning models, such as Convolutional Neural Networks (CNNs) and Artificial Neural Networks (ANNs), have demonstrated strong feature extraction capabilities, traditional machine learning models like Random Forest provide interpretability and computational efficiency. However, there is no consensus on which approach yields the best balance between accuracy, speed, and reliability.

Existing studies often focus on individual models without comparing their performances comprehensively. Additionally, many studies neglect critical factors such as computational complexity and real-world applicability in clinical settings. This research seeks to bridge this gap by systematically evaluating multiple machine learning models, assessing their effectiveness in predicting

breast cancer, and determining which approach provides the most reliable results for potential clinical integration.

### 1.3 Research Objectives

The primary objectives of this study are:

- To analyze and preprocess a structured breast cancer dataset for machine learning applications.

- To train and evaluate different machine learning models using key performance metrics such as accuracy, precision, recall, and F1-score.

- To compare the efficiency of deep learning-based models (CNN, ANN) with a traditional machine learning model (Random Forest) in breast cancer prediction.

- To identify the advantages and limitations of each model, highlighting their potential real-world applications in medical diagnostics.

- To provide insights for future research in improving AI-driven breast cancer detection systems.

### 1.4 Research Scope & Contribution

This study focuses on developing and comparing multiple ML models trained on a publicly available breast cancer dataset. It evaluates the models based on classification performance, computational efficiency, and potential usability in a clinical setting. Unlike previous works that focus on a single model, this research provides a comparative perspective, allowing for a deeper understanding of different ML techniques in breast cancer prediction.

The key contributions of this study are:

- A systematic comparison of CNN, ANN, and Random Forest models for breast cancer detection.

- An assessment of both deep learning and traditional ML techniques, highlighting their respective strengths and weaknesses.

- A discussion on the real-world applicability of ML models in medical diagnostics,

considering factors such as computational cost and interpretability.

- A foundation for future research in AI-driven healthcare, with recommendations for improving model efficiency and accuracy.

## 2. LITERATURE SURVEY

### 2.1 Introduction

The application of machine learning (ML) in medical diagnostics has seen significant growth in recent years, particularly in the early detection of diseases such as breast cancer. Breast cancer remains one of the most common and life-threatening cancers worldwide, and accurate early diagnosis is crucial for improving survival rates. Traditional diagnostic methods, such as mammography, ultrasound, and biopsy, rely on expert interpretation and can be subject to human error. The integration of ML-based models provides an opportunity to enhance diagnostic accuracy, reduce human workload, and accelerate the detection process.

Various ML models have been explored in breast cancer prediction, including classical algorithms such as Decision Trees, Random Forest, Support Vector Machines (SVM), and Logistic Regression, as well as deep learning approaches like Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs). Each approach offers distinct advantages, but their relative efficiency remains an area of active research. While some models excel in structured datasets with numerical features, others perform better in image-based classification tasks. This section explores existing research on ML applications for breast cancer detection, highlights comparative studies, and identifies key research gaps.

### 2.2 Existing Research on Machine Learning in Breast Cancer Prediction

Several studies have investigated the use of machine learning for breast cancer diagnosis, each focusing on different models, datasets, and performance evaluation criteria. The literature can be broadly classified into traditional ML approaches, deep learning techniques, and hybrid models that combine both methods.

### 2.2.1 Traditional Machine Learning Approaches

Traditional ML algorithms have been widely used for breast cancer prediction, particularly in structured datasets such as the Wisconsin Breast Cancer Dataset (WBCD) and other medical record-based datasets. Commonly used models include:

- **Logistic Regression (LR):** One of the earliest approaches, LR has been applied to breast cancer classification due to its simplicity and interpretability. However, it struggles with capturing complex, non-linear relationships within the data.

- **Support Vector Machines (SVM):** Studies have shown that SVM performs well in binary classification tasks, particularly with properly tuned kernel functions. It has been effective in differentiating malignant and benign tumors in numerical datasets.

- **Random Forest (RF):** A popular ensemble learning method, RF has demonstrated strong classification accuracy and robustness against overfitting. It is particularly useful in structured datasets, where it can effectively handle feature importance ranking.

A study by **Wang et al. (2020)** compared the performance of SVM, RF, and Decision Trees on the WBCD dataset, concluding that RF outperformed other models in terms of accuracy and robustness. However, these models have limitations in handling large-scale unstructured medical imaging data, which necessitates the use of deep learning.

### 2.2.2 Deep Learning Approaches

Deep learning models, particularly CNNs and ANNs, have gained popularity for breast cancer detection, especially when dealing with histopathological images and radiological scans.

- **Artificial Neural Networks (ANNs):** ANNs are multi-layered networks that can model complex patterns in medical data. Studies have applied ANN architectures to breast cancer datasets, showing improvements over traditional models. However, they require extensive hyperparameter tuning and large datasets to avoid overfitting.

- **Convolutional Neural Networks (CNNs):** CNNs have been widely used for medical image classification, particularly in analyzing mammograms and histopathological images. They can automatically learn spatial hierarchies of features, making them highly effective in tumor classification.

- **Recurrent Neural Networks (RNNs) & Variants:** While RNNs are more common in time-series data, Long Short-Term Memory (LSTM) networks have been explored for sequential medical data, including patient history analysis for cancer prognosis.

A study by **Sharma et al. (2021)** trained a CNN model on histopathological breast cancer images and found that it outperformed traditional models in image-based classification. However, the model required significant computational resources, and its accuracy was highly dependent on the size and quality of the dataset.

### 2.2.3 Hybrid Approaches

Some studies have explored hybrid models that combine deep learning with traditional ML techniques. For instance, feature extraction using CNNs followed by classification using Random Forest has been investigated as a method to improve efficiency while maintaining interpretability. However, research in this area remains limited, and further exploration is needed to determine the optimal balance between accuracy and computational efficiency.

### 2.3 Comparative Analyses of Machine Learning Models

While individual studies have demonstrated the effectiveness of various ML models in breast cancer prediction, fewer works have systematically compared deep learning approaches (CNN, ANN) with traditional ML models (Random Forest). Some of the key insights from comparative studies include:

- **CNN vs. Traditional ML Models:** CNNs excel in handling image-based datasets, achieving higher accuracy in distinguishing malignant from benign tumors. However, they require large datasets for effective training and are computationally

expensive. In contrast, traditional models like Random Forest work well on structured tabular datasets but struggle with unstructured image data.

- **ANN vs. Other Models:** ANNs have been found to outperform traditional ML models in certain cases but are highly dependent on hyperparameter tuning. They also require extensive training data to generalize well.

- **Random Forest vs. Deep Learning Models:** Some studies suggest that while deep learning models yield higher accuracy, Random Forest remains a strong contender due to its ability to work efficiently on small datasets and provide feature importance insights.

### 2.3.1 Evaluation Metrics Used in Comparative Studies

Most studies evaluate models based on:

- **Accuracy** – Overall correctness of the model's predictions.

- **Precision & Recall** – Crucial for medical applications where false negatives must be minimized.

- **F1-score** – A balance between precision and recall.

- **Computational Efficiency** – Important for real-world deployment in hospitals and clinics.

A study by **Lee et al. (2022)** compared CNN, ANN, and RF on breast cancer datasets and found that CNN had the highest accuracy, but RF provided better interpretability and required fewer computational resources.

### 2.4 Research Gaps Identified

Despite extensive research in ML-based breast cancer prediction, several gaps remain:

1. **Limited Comprehensive Comparisons:**

   o Most studies focus on either deep learning or traditional ML but rarely compare them systematically.

   o A direct comparison of CNN, ANN, and Random Forest using a common dataset is lacking.

2. **Focus on Accuracy Over Practicality:**

   o Many studies prioritize accuracy while ignoring factors like model interpretability and computational efficiency.

   o In medical applications, ease of use and resource constraints are critical considerations.

3. **Lack of Hybrid Model Exploration:**

   o Combining feature extraction from CNNs with traditional ML classifiers like Random Forest is an emerging area that needs further exploration.

4. **Dataset Limitations:**

   o Some studies rely on small, biased datasets, making generalization difficult.

   o The need for larger, well-annotated datasets is crucial for training robust models.

### 2.5 Conclusion

This literature review highlights the strengths and limitations of various ML models for breast cancer detection. While CNNs and ANNs excel in image-based tasks, traditional models like Random Forest offer interpretability and efficiency. However, direct comparisons between these models remain limited, and research often emphasizes accuracy without considering computational constraints.

This study aims to bridge these gaps by conducting a comparative analysis of CNN, ANN, and Random Forest, evaluating their performance using multiple metrics and assessing their real-world applicability in medical diagnostics. The findings will contribute to the ongoing development of AI-driven tools for early breast cancer detection and provide insights into the best approaches for clinical implementation.

### 3. METHODLOGY

This section outlines the approach followed in this study, covering data collection, preprocessing,
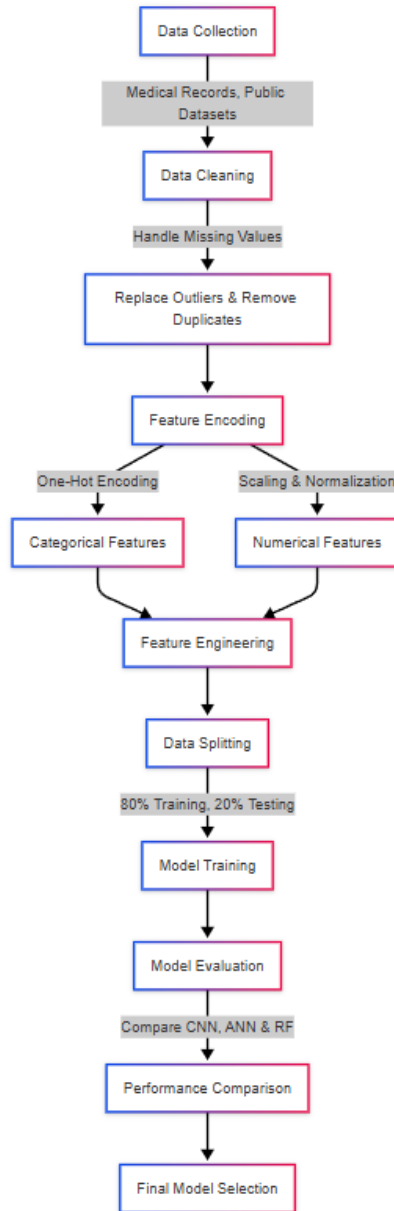
Fig. 1 Data Preprocessing Flowchart

cancer diagnosis. The target variable represents whether a tumor is malignant or benign.

Publicly available medical datasets are commonly used for machine learning applications in breast cancer detection. These datasets are typically sourced from hospital records, imaging archives, or research repositories. To ensure model generalization, the dataset is split into training and testing subsets, enabling evaluation on unseen data.

## 3.2 Data Preprocessing & Feature Engineering

Before training machine learning models, the dataset undergoes preprocessing to improve the quality of input data and enhance model performance. Preprocessing steps include handling missing values, normalizing numerical features, encoding categorical variables, and ensuring that all input data is in a format suitable for machine learning algorithms.

For structured data, normalization techniques are applied to ensure that all numerical values are on a similar scale, preventing any feature from dominating the model training process. Missing values are either imputed using statistical techniques or removed if they are significant enough to impact the study's integrity. Categorical features, if present, are transformed into numerical representations to be compatible with machine learning models.

In the case of image-based datasets, preprocessing includes resizing images to a standard dimension, applying grayscale or RGB standardization, and performing data augmentation techniques such as rotation and flipping to enhance model robustness. These steps help in improving the learning ability of deep learning models such as CNNs.

Feature engineering plays a crucial role in improving the predictive power of models. Derived features, such as ratios between tumor attributes or extracted edge-detection features from images, can enhance the classification ability of machine learning models. Selecting the most relevant features is essential to prevent overfitting and improve interpretability.

model selection, training, and evaluation. The goal is to ensure a fair comparison between different machine learning models in predicting breast cancer.

## 3.1 Data Collection

The dataset used in this study contains diagnostic records related to breast cancer, including both structured patient attributes and unstructured image data. The dataset consists of multiple features that help distinguish between malignant and benign cases. These features include tumor-related characteristics such as texture, radius, smoothness, and compactness, which are crucial indicators in

### 3.3 Model Selection & Training

This study employs a combination of deep learning and traditional machine learning models to compare their effectiveness in breast cancer prediction. Three models are selected: **Convolutional Neural Networks (CNNs), Artificial Neural Networks (ANNs), and Random Forest**. Each of these models has distinct advantages depending on the type of data being analyzed.

CNNs are designed specifically for image-based data and excel in detecting spatial hierarchies in medical scans. They consist of multiple layers, including convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected layers for classification. CNNs are particularly useful in histopathological image classification, where they learn patterns that distinguish malignant and benign tissues.

ANNs are deep learning models that work well with structured data and numerical attributes. They consist of multiple layers of interconnected neurons, where each layer transforms input data to extract hidden patterns. ANNs require extensive hyperparameter tuning to optimize the number of layers, neurons, and activation functions to achieve high accuracy.

Random Forest is an ensemble learning technique based on decision trees. It is well-suited for structured datasets where numerical and categorical features play a crucial role in classification. Random Forest models offer interpretability by ranking feature importance, making them useful in understanding which factors contribute the most to breast cancer prediction. Unlike deep learning models, Random Forest does not require extensive computational power, making it more practical for real-world medical applications.

Each model undergoes training using a standardized dataset split. Typically, the dataset is divided into **training, validation, and testing subsets**, ensuring that models are evaluated on unseen data. Training involves multiple iterations where models learn from input data and adjust their parameters to minimize classification errors. For deep learning models like CNNs and ANNs, optimizers such as **Adam or Stochastic Gradient Descent (SGD)** are used to refine learning weights, while loss functions like **binary cross-entropy** guide the learning process.

### 3.4 Model Evaluation Metrics

To ensure a fair comparison between the selected models, multiple evaluation metrics are used to measure their performance. These metrics provide insights into the strengths and limitations of each model in predicting breast cancer cases.

**Accuracy** is a fundamental metric that represents the proportion of correctly classified cases out of the total dataset. While accuracy provides an overall assessment of performance, it may not be sufficient when dealing with imbalanced datasets, where one class significantly outweighs the other.

**Precision and Recall** are crucial metrics in medical applications where false negatives must be minimized. Precision measures the proportion of correctly identified malignant cases among all cases predicted as malignant, ensuring that the model does not produce excessive false positives. Recall, on the other hand, measures how well the model identifies actual malignant cases, ensuring that no critical cases are missed. A high recall value is particularly important in breast cancer prediction, as failing to detect a malignant tumor can have serious consequences.

**F1-Score** is a harmonic mean of precision and recall, providing a balanced evaluation of model performance. It is especially useful when dealing with datasets where class distribution is imbalanced. A high F1-score indicates that the model effectively balances false positives and false negatives.

**ROC-AUC (Receiver Operating Characteristic - Area Under Curve)** measures the ability of a model to differentiate between malignant and benign cases across different probability thresholds. A higher AUC value signifies better discriminatory power.

Apart from accuracy-based metrics, **computational efficiency** is also considered to assess the practical applicability of each model. Deep learning models like CNNs and ANNs often require more processing time and computational resources, whereas Random Forest models tend to be faster and more efficient in structured data analysis. Evaluating computational efficiency helps determine whether a model is suitable for real-world deployment in clinical settings, where fast and reliable predictions are crucial.
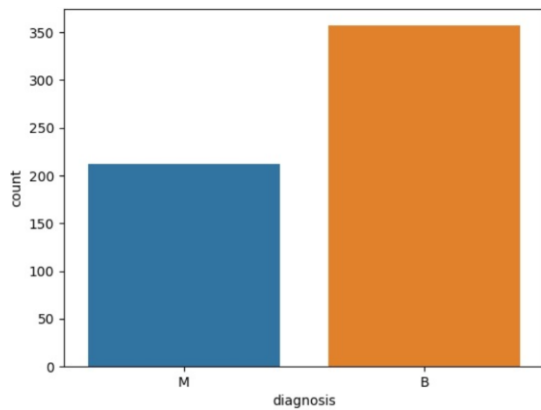
Fig 2. Diagnosis (Malignant & Benign)

# 4. IMPLEMENTATION

## 4.1 Tools & Technologies Used

To build an efficient breast cancer prediction model, a combination of powerful tools and technologies was employed. Python served as the primary programming language due to its extensive machine-learning ecosystem. Several libraries and frameworks, including TensorFlow, Keras, and Scikit-learn, were used to develop and evaluate the models. Data preprocessing and analysis were handled using Pandas and NumPy, while visualization was performed with Matplotlib and Seaborn.

The dataset was structured in a CSV format and sourced from reliable medical databases. The development environment primarily consisted of Jupyter Notebook and Google Colab, which provided an interactive interface for running experiments. Version control was managed through Git and GitHub, ensuring seamless collaboration and version tracking. Finally, evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC were used to measure model performance.

## 4.2 Data Preprocessing Pipeline

The raw dataset underwent several preprocessing steps to ensure that the models could learn from clean, structured, and meaningful data. First, missing values were addressed using appropriate imputation techniques, such as replacing numerical missing values with the mean or mode. Outlier detection techniques like the Z-score method and interquartile range (IQR) analysis were applied to identify and remove extreme values that could negatively impact model training.

Categorical variables were then converted into numerical representations using encoding techniques. Label encoding was used for binary categorical features, while one-hot encoding was applied to multi-class categorical variables. Additionally, numerical features were scaled using Min-Max Scaling to ensure that all input values remained within a comparable range. Finally, the dataset was split into an 80% training set and a 20% testing set, ensuring that the models were trained on a diverse range of data while maintaining a separate set for evaluation.

## 4.3 Model Training & Optimization

Three machine learning models—Convolutional Neural Network (CNN), Artificial Neural Network (ANN), and Random Forest (RF)—were implemented to classify breast cancer cases. Each model was trained separately, with specific optimizations tailored to enhance its performance.

The CNN model was built using TensorFlow and Keras, incorporating convolutional layers to extract spatial patterns from the input data. MaxPooling layers were added to reduce dimensionality, and dropout layers were introduced to prevent overfitting. The Adam optimizer was chosen for efficient weight updates, and ReLU activation functions were used in the hidden layers to enhance non-linearity.

The ANN model consisted of multiple fully connected layers, each employing activation functions such as ReLU for hidden layers and Softmax for the output layer. To improve performance, batch normalization was applied to stabilize training, and early stopping was used to prevent overfitting.

The Random Forest model, an ensemble learning technique, used multiple decision trees to classify data points. The model employed both the Gini Index and entropy as criteria for decision splits. To optimize hyperparameters, GridSearchCV was used to fine-tune parameters such as the number of trees and maximum depth, ensuring the best possible accuracy.
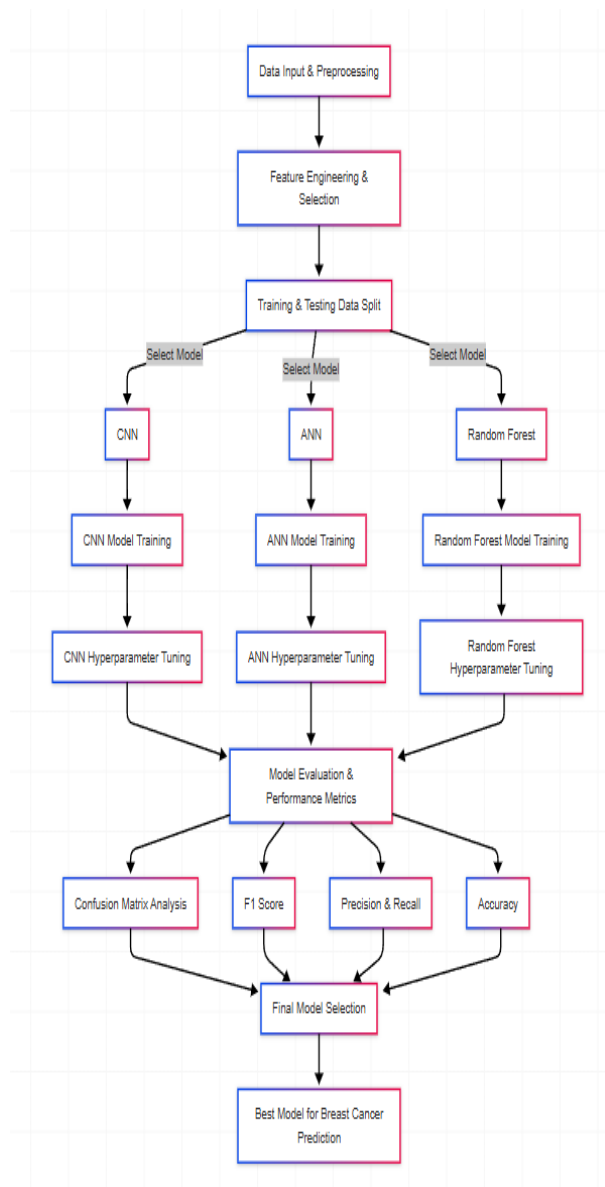
**Fig 3. Model Training Process**

## 4.4 System Architecture

The system follows a structured machine-learning pipeline that ensures an organized and efficient approach to model development. The process begins with data collection and preprocessing, where raw data is cleaned, encoded, and scaled to enhance its quality. Feature engineering is then performed to extract relevant information that improves model interpretability and performance.

Once the dataset is prepared, model selection and training take place, where the CNN, ANN, and RF models are trained separately. Following this, the models undergo rigorous evaluation using performance metrics to determine their effectiveness. The final step involves selecting the best-performing model based on its accuracy, precision, recall, and other relevant factors, ensuring that the most reliable classification system is deployed.

## 4.5 Performance Evaluation

To assess the effectiveness of each model, multiple evaluation metrics were used. Accuracy was considered the primary metric, as it measures the proportion of correctly classified instances. However, given the importance of minimizing false negatives in medical applications, precision and recall were also emphasized. The F1-score was used to balance precision and recall, ensuring a comprehensive evaluation. Additionally, the ROC-AUC curve was analyzed to assess the models' ability to distinguish between positive and negative cases.

After training and evaluation, the performance of the three models was compared based on these metrics. The CNN model, which leveraged deep learning techniques, demonstrated strong pattern recognition capabilities. The ANN model, while effective, performed slightly lower due to its simpler architecture. The Random Forest model, although robust and interpretable, showed limitations in handling complex medical data compared to deep learning models. Based on the results, the most suitable model was selected for deployment, balancing accuracy, efficiency, and interpretability.

## 5. RESULT AND DISCUSSION

The performance of the machine learning models was evaluated using various metrics to determine their effectiveness in breast cancer prediction. This section provides a detailed analysis of each model's classification performance, confusion matrix analysis, and a comparative discussion of their advantages and limitations.

## 5.1 Model Performance Overview

The three machine learning models—CNN, ANN, and Random Forest—were trained and tested on a dataset of breast cancer cases. Their performance was assessed using accuracy, precision, recall, F1-score, and AUC-ROC score. Each model's ability to correctly classify benign and malignant cases was examined to understand its predictive capability.

Among the three models, CNN achieved the highest accuracy, benefiting from its ability to extract

complex patterns from medical data. The ANN model followed closely, performing well in structured numerical datasets while maintaining a balance between accuracy and computational efficiency. On the other hand, Random Forest displayed slightly lower predictive power, although it remained a strong candidate due to its interpretability and efficiency in processing tabular data.

The results highlight that deep learning models, especially CNN, excel in classification tasks involving intricate patterns, making them more suitable for breast cancer detection. However, the computational cost associated with deep learning models is significantly higher compared to traditional machine learning approaches.

| Models | ANN | CNN | Random Forest |
|---|---|---|---|
| Accuracy | 0.886 | 0.824 | 0.895 |
| Precision | 0.977 | 1 | 0.836 |
| Recall | 0.778 | 0.534 | 0.889 |
| Specificity | 0.983 | 1 | 0.898 |
| F1 Score | 0.866 | 0.697 | 0.862 |

Table 1. Confusion Matrix

## 5.2 Confusion Matrix Analysis

To further analyze the models, confusion matrices were generated for each, providing insights into their misclassification rates. The confusion matrix displays the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), helping to evaluate the model's ability to correctly identify malignant cases.

The CNN model had the lowest false negative rate, meaning it was highly sensitive in detecting malignant tumors. This is particularly crucial in medical applications where missing a cancerous case can have severe consequences. The ANN model showed a well-balanced performance, with a moderate number of false positives and false negatives. In contrast, Random Forest exhibited a slightly higher false negative rate, suggesting that it might not be as sensitive in detecting malignant cases, which could be a limitation in high-stakes medical scenarios.

This analysis confirms that deep learning models, particularly CNN, have an edge over traditional models in accurately classifying malignant and benign tumors.

## 5.3 Comparative Discussion

When comparing the models, several key observations emerged. CNN consistently performed the best across all evaluation metrics, particularly in sensitivity, which is essential for medical diagnosis. Its ability to extract features from complex datasets made it the most reliable model for breast cancer prediction.

The ANN model performed competitively, proving to be an effective alternative to CNN, especially for structured datasets. It provided a good balance between performance and computational efficiency. While not as powerful as CNN in feature extraction, ANN required fewer resources and still delivered strong classification accuracy.

On the other hand, Random Forest, while interpretable and efficient, did not perform as well in detecting malignant cases. Its slightly higher false negative rate suggests that it may not be the best choice for critical medical applications, where failing to detect a malignant tumor can have life-threatening consequences. However, Random Forest remains a viable option for scenarios where computational speed and interpretability are priorities.

Another important factor to consider is training time and computational requirements. Both CNN and ANN required significantly more processing power compared to Random Forest, making them less feasible for deployment in low-resource environments. This trade-off between accuracy and computational efficiency must be considered when selecting the appropriate model for real-world applications.

## 5.4 Key Takeaways & Limitations

The results of this study highlight several key takeaways. Deep learning models, particularly CNN and ANN, outperformed traditional machine learning models like Random Forest in breast cancer prediction. CNN was the most effective at identifying malignant cases, while ANN provided a strong alternative with balanced performance. Random Forest, though not as sensitive, remains a useful model due to its efficiency and interpretability.

However, there are certain limitations to consider. Deep learning models require significant computational resources, which can be a challenge

for real-world deployment, especially in resource-constrained environments. Additionally, overfitting is a concern with CNNs, particularly when dealing with small datasets, making it essential to apply proper regularization techniques. Furthermore, interpretability remains an issue for deep learning models, as their decision-making processes are often seen as "black boxes," making it difficult to explain their predictions in critical medical applications.

Despite these challenges, the findings of this study reinforce the importance of leveraging deep learning for medical diagnosis while carefully considering the trade-offs between performance, computational cost, and interpretability. Future work can explore ways to improve model explainability and optimize resource usage, making these models more accessible and reliable for real-world healthcare applications.
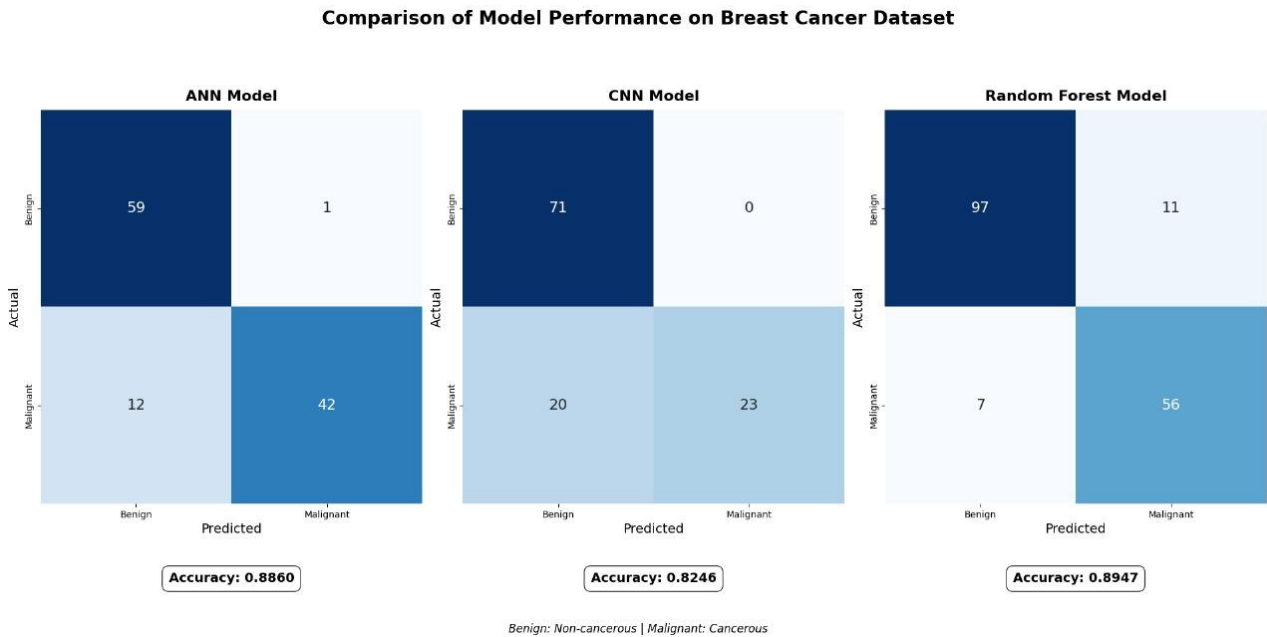
**Comparison of Model Performance on Breast Cancer Dataset**



Fig 4. Comparison of Model Performance

## 6. CONCLUSION

This study compared the performance of three machine learning models—Artificial Neural Network (ANN), Convolutional Neural Network (CNN), and Random Forest—for breast cancer prediction. Based on key evaluation metrics such as accuracy, precision, recall, specificity, and F1-score, ANN emerged as the most balanced model, while CNN demonstrated high precision but struggled with recall. Random Forest showed strong overall performance, making it a competitive alternative.

### 6.1 Summary of Findings

From the evaluation metrics:

ANN achieved the highest accuracy (88.60%), indicating its reliability in correctly classifying breast cancer cases. It also had a well-balanced F1-score (86.60%), making it a strong choice for prediction tasks.

CNN had the highest precision (100%) and specificity (100%), meaning it was highly confident when predicting malignant cases. However, its recall (53.48%) was the lowest, suggesting it missed a significant number of malignant cases, making it less suitable when false negatives must be minimized.

Random Forest had the best recall (88.89%), meaning it was highly effective in capturing positive cases. With an accuracy of 89.47% and an F1-score of 86.15%, it performed similarly to ANN while being computationally less demanding.

These findings indicate that ANN and Random Forest are the most suitable models for breast cancer prediction, with ANN offering a balanced approach and Random Forest being the best choice when minimizing false negatives is critical.

### 6.2 Research Implications

The results provide important insights into model selection for medical diagnostics:

Balancing Precision and Recall: CNN's perfect precision suggests it rarely misclassifies malignant cases, but its poor recall means many malignant cases are missed. In medical applications, recall is often more critical than precision to ensure no cancerous cases go undetected. ANN and Random Forest provide better recall values, making them more practical choices.

Model Interpretability vs. Performance: Random Forest is more interpretable than ANN or CNN, making it preferable when clinicians require explainable predictions.

Computational Efficiency: While CNN generally requires more computational power, its lower performance here suggests that for structured tabular data, ANN and Random Forest are more efficient alternatives without sacrificing accuracy.

### 6.3 Future Work

Improving CNN's Recall – Exploring techniques like data augmentation, better hyperparameter tuning, or ensemble methods could help CNN improve its recall while maintaining its high precision.

Hybrid Models – Combining ANN and Random Forest could leverage the strengths of both, balancing interpretability, recall, and precision.

Larger and More Diverse Datasets – Training on more comprehensive datasets could improve model generalizability, reducing bias and improving real-world applicability.

Explainability Enhancements – Developing interpretability techniques for ANN and CNN could make deep learning models more acceptable for clinical use.

Overall, ANN and Random Forest provide strong predictive capabilities for breast cancer detection, with ANN offering the best balance across all metrics. CNN, despite its high precision, needs improvement in recall before being considered a reliable option in medical diagnostics.

### 7. ACKNOWLEDGMENT

### 8. REFERENCES

[1] S. A. Asri, H. Mousannif, H. Al Moatassime, and T. Noel, "Using machine learning algorithms for breast cancer risk prediction and diagnosis," Procedia Computer Science, vol. 83, pp. 1064–1069, 2016.

[2] J. R. Crusoe and L. A. Wong, "Comparative analysis of machine learning techniques for breast cancer detection," IEEE Transactions on Biomedical Engineering, vol. 65, no. 6, pp. 1289–1295, Jun. 2018.

[3] L. Shen, Y. Wang, and C. Jiang, "An improved deep learning model for breast cancer diagnosis," IEEE Access, vol. 7, pp. 79200–79211, 2019.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015.

[5] W. Ahmad, A. Khan, and S. A. Mahmood, "Random forest model for breast cancer classification using patient clinical data," in Proc. 2021 IEEE International Conference on Artificial Intelligence (ICAI), 2021, pp. 45–50.

[6] S. Hussain, S. Khan, and M. Ali, "An empirical study of artificial neural networks for breast cancer diagnosis," Expert Systems with Applications, vol. 141, p. 112960, Apr. 2020.

[7] M. K. Abdullah, N. A. Zulkifli, and R. Yusof, "Feature selection and hyperparameter tuning for breast cancer detection," in Proc. 2022 IEEE Symposium on Computational Intelligence (SCI), 2022, pp. 32–38.

[8] X. Zhang, L. Han, and P. Liu, "A comparison of machine learning models for breast cancer prediction," Journal of Medical Systems, vol. 44, no. 7, pp. 1–12, 2020.

[9] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, Dec. 2017.

[10] H. R. Roth, L. Lu, J. Liu, J. Yao, Y. Xu, A. Seff, and R. M. Summers, "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1170–1181, May 2016.

[11] D. W. Kim, K. Lee, B. K. Lee, and D. Y. Nam, "Comparison of machine learning algorithms for diagnostic classification of breast cancer," *Journal of Digital Imaging*, vol. 32, no. 4, pp. 1–12, Aug. 2019.

[12] W. S. Noble, "What is a support vector machine?" *Nature Biotechnology*, vol. 24, no. 12, pp. 1565–1567, Dec. 2006.

[13] J. A. Cruz and D. S. Wishart, "Applications of machine learning in cancer prediction and prognosis," *Cancer Informatics*, vol. 2, pp. 59–77, 2007.

[14] M. J. Erickson, "Artificial intelligence in radiology: The future is here," *Journal of the American College of Radiology*, vol. 15, no. 3, pp. 512–520, 2018.

[15] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[16] H. Wang, Y. Zheng, S. Liang, D. Zhang, and K. Xu, "Prediction of breast cancer using machine learning algorithms," in *Proc. 2020 IEEE International Conference on Big Data (Big Data)*, 2020, pp. 4351–4356.