Multi-Model Learning for Breast Cancer Prediction: Performance Evaluation Across Diverse Datasets

Geshna B
Amrita School of Artificial Intelligence, Coimbatore,
Amrita Vishwa Vidyapeetham, India,
cb.sc.u4aie23360@cb.students.amrita.edu

Katikala Dedeepya
Amrita School of Artificial Intelligence, Coimbatore,
Amrita Vishwa Vidyapeetham, India,
cb.sc.u4aie23349@cb.students.amrita.edu

Malavika S Prasad
Amrita School of Artificial Intelligence, Coimbatore,
Amrita Vishwa Vidyapeetham, India,
cb.sc.u4aie23315@cb.students.amrita.edu

Vada Gouri Hansika Reddy
Amrita School of Artificial Intelligence, Coimbatore,
Amrita Vishwa Vidyapeetham, India,
cb.sc.u4aie23304@cb.students.amrita.edu

Prajeesh C B
Amrita School of Artificial Intelligence, Coimbatore,
Amrita Vishwa Vidyapeetham, India,
cb_prajeesh@cb.amrita.edu

***Abstract*—Breast cancer still ranks among the leading causes of cancer deaths worldwide, which emphasizes the importance of proper and prompt diagnostic evaluation. This work utilizes a two-modality framework for predicting breast cancer, which integrates deep learning with classical machine learning techniques, for a thorough and detailed investigation. Vertical-based diagnosis uses convolutional neural networks (CNN) in histopathology and mammogram images for deep feature extraction and classification. In contrast, a well-defined Wisconsin Diagnostic Breast Cancer (WDBC) database is examined with several ways of classical machine learning systems, such as Logistic Regression, Support Vector Machine (SVM), Random Forest, Gradient Boosting, and XGBoost models. It is found that simpler models, such as logistic regression, work better for the structured datasets, while handling images, the CNNs come in handy. This work addresses the limitations of these techniques and shows how their integration can improve breast cancer prediction and thus assist in early-stage diagnosis.**

***Index Terms*—Breast Cancer Prediction, Convolutional neural networks (CNNs), Histopathological images, Mammogram Analysis, Wisconsin Diagnostic Breast Cancer (WDBC) data set, Logistic regression, Support Vector Machines (SVM), Random Forest, Gradient Boosting, XGBoost, Image-based classification, Structured Data Analysis, Machine learning (ML), Deep Learning, Diagnostic Accuracy, Early Detection.**

I. Introduction

Breast cancer continues to be the most prevalent cancer diagnosed in women, as well as one of the chief causes of cancer deaths among women across the globe, as estimated by the World Health Organization (WHO). In recent years, due to the advancements in artificial intelligence (AI) and machine learning (ML), the aspect of medical diagnostics has changed drastically in that the solutions offered are automated, quick, and more precise. Deep learning techniques, for that matter, Convolutional Neural Networks (CNN), in particular, have been reported to perform well where there is a lot of unstructured data, especially the images in the medical field. This is possible because CNNs can do all these levels of feature extraction from the images without any guidance, even the smallest detail that even an expert may fail to see. Such qualities make CNNs essential in operations such as identifying and classifying tumors in histopathology images and breast X-ray images, which is complicated. In addition to the trends in deep learning, other older models under ML have proved very useful over the years when working with structured data. Structured datasets like those with patients' characteristics, clinical variables, and biopsy results typically call for models that can be interpreted and are efficient in computation. Algorithms like Logistic Regression, Support Vector Machine (SVM), Random Forest, Gradient Boosting, and XGBoost come in handy for such a case. These types of Models have their benefits in ease of use, processing time, and clarity, which makes them very important when it comes to making clinically related decisions. This paper describes a new dual-modality breast cancer prediction system that logically combines deep learning with conventional ML methods. In the image-based portion, CNN is used to classify breast cancer based on histopathology and mammography pictures, which are database datasets that are extremely raw, complex, and require advanced techniques for effective analysis. Meanwhile, structured data using conventional ML models is applied to data from a Wisconsin Diagnostic Breast Cancer (WDBC) dataset. Such a method allows for a comprehensive review and acknowledges the various types of breast cancer data. This study aims to assess and distinguish deep learning strategies against conventional ML strategies on various forms of data. CNN models are evaluated principally for their potential to extract valuable information from complex high-resolution medical images and accurately categorize the images. The performance of traditional ML models is assessed in terms of accuracy, ease of understanding and application, and time taken, when applied in structured data settings. Such an analysis helps understand the advantages and disadvantages of each conceptualization given particular situations and the nature of the data involved. This endeavor is in line with the ongoing efforts to enhance breast cancer diagnosis with the help of AI, as it seeks to optimize breast cancer diagnosis. The proposed framework is a comprehensive and scalable approach that combines deep learning and conventional ML, making it suitable for practical clinical applications. It also establishes a new research direction regarding the relationship of these two approaches in solving more complex problems in the analysis of medical data. Existing breast cancer prediction methods fail to leverage the complementary strengths of heterogeneous data modalities; this work fills the gap by introducing and testing a cross-modal learning framework that synergistically integrates histopathological imaging and clinical information for enhanced predictive performance.

II. Related Work

The study of the prediction of breast cancer has become one of the long-term discussions on machine learning and deep learning. This section looks historically at prior work based on structured input data analysis, image-based classification, and hybrid approaches in predicting breast cancer.

*1) **Machine Learning Techniques for Structured Data** :* On the contrary, machine learning has been widely used in structured datasets such as the Wisconsin Diagnostic Breast Cancer (WDBC) dataset. Bennet and Mangasarian [1] established a model to perform classification of malignant and benign samples using linear programming to create a platform for ML-based diagnostic systems. Later on, Ahmed et al. [2] compared some ML approaches and found that logistic regression and support vector machine did very well on structured datasets owing to their simplicity and interpretability. Kadhim and Kamil [3] state that ensemble models such as Random Forest and Gradient Boosting are usually much better on handling of non-linear relationships in structured datasets than their simpler model counterparts. However, they mention that simpler models like Logistic Regression often perform well when datasets are well-preprocessed and are linearly separable.

*2) **DL Techniques for Image-Based Classification**:* Deep learning based on CNNs is showing some encouraging results in medical imaging analysis. Zou et al. [4] reviewed the application of CNNs in mammographic breast cancer diagnosis, emphasizing CNNs' ability to learn complex features automatically. Gupta and Chawla [5] extended the analysis to histopathological images, proving a very high accuracy to discriminate malignant from benign areas. However, factors like a small dataset and variability hamper CNN performance. Hossin et al. [6] recommend data augmentation and transfer learning to combat overfitting and enhance generalization. Jiang et al. [7] and Liu et al. [8] showed that transfer learning and hybrid CNN models improve breast cancer classification. Arevalo et al. [9] also reported that CNN-based representation learning gives better results than traditional methods.

*3) **Hybridization Technique**:* Recently, hybrid approaches that merge data and images have been used for medical image analysis. Atajanov and Peker [10] showed that using image features extracted through CNN with clinical data can improve diagnostic accuracy. Indeed, promising models like this usually do not address problems arising from computational complexity. The majority of studies have thus far, focused their research on only one side-either image or structured data. The value of data sources was emphasized by El Houby [11], while hybrid models, according to Al-Shaikhli et al. [12], often outperform single-method approaches. According to Sampat et al. [13], multiscale image analysis has also advantages for diagnosis. This research intends to fill the gap using image features based on CNN and machine learning on structured data for improved prediction of breast cancer.

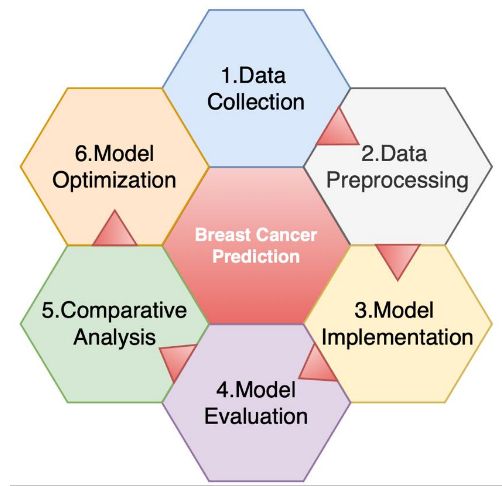

Fig. 1. Six-stage approach for Breast Cancer Prediction

III. Methodology

The proposed framework in Figure 1 follows a systematic process, outlining the six main stages that extend from preprocessing of the data to actual prediction. Additionally, the entire system architecture comprises a combination of image-based CNN analysis and conventional machine learning techniques to provide a better diagnosis in Figure 2.

A. Datasets

CBIS-DDSM Dataset: The CBIS-DDSM is a large and exhaustive curated source of mammography images, dedicated to advancing research and clinical endeavors in breast cancer detection. This dataset is a revised and more standardized update of the original DDSM; it is widely used in developing and testing CADe and CADx systems in mammography.

Major Characteristics of CBIS-DDSM Dataset:

*1) **File Format with size**:* JPEG for images with a grand total of about 6GB. This subset represents a compressed version of the original DDSM data, which was 163GB in size, but retains the original resolution of each image.

*2) **Resolution and Image Quality**:* The CBIS-DDSM images were given resolution that was retained from the original scans and thus make this dataset suitable for quality mammographic analysis.

- **Number of Studies:** 6,775
- **Number of Series:** 6,775
- **Number of Participants:** 1,566 distinct participants, though the dataset's metadata assigns multiple patient IDs per participant, leading to a total of 6,671 patient IDs.
- **Total Images:** 10,239 images available in this subset.
- **Modality:** Mammography (MG)

*3) **Dataset Structure and Metadata**:* The dataset assigns multiple patient IDs to each participant, which can be misleading. For example, one participant might be represented by a few patient IDs as they correspond to completely different mammographic scans or series.

*4) **Region of Interest (ROI), as well as Diagnosis Information**:* The CBIS-DDSM holds updated ROI annotations with bounding boxes and segmentation information, essential to training and validating CADx/CADe algorithms. For each case, confirmed pathology data are made available, establishing normal, benign, and malignant findings to provide accurate and well-structured training data in ML applications for the detection of breast cancer.

B. Histopathological Images

This dataset comprises whole-mount slide images of 162 breast cancer specimens scanned with a 40x magnification, together with 277,524 image patches, each of size 50x50 pixels, of different types. Among these, 198,738 images are IDC-negative patches, while the other 78,786 are IDC positive. This patch distribution makes it possible for ML models to learn to differentiate between cancerous and non-cancerous regions on the slides. In every patch, a systematic naming format is applied, which provides metadata in the filename. With great organization for the advancement in automated pathologies and breast cancer research, this data set serves as an important base for algorithm development and testing for the detection of IDC and the classification of aggressiveness.

C. Wisconsin Dataset

The WDBC (Wisconsin Diagnostic Breast Cancer) is a well-known dataset for classification to demonstrate whether the tumor is benign or malignant. It contains 569 samples: 357 benign and 212 malignant. Each sample has 30 numerical features derived from cell nuclei in fine needle aspirate (FNA) images. These features include measurements like radius, texture, perimeter, area, smoothness, compactness, concavity, symmetry and fractal dimension each recorded as mean, standard error, and worst value. The file is clean, complete, and widely used in breast cancer diagnosis and machine learning studies.

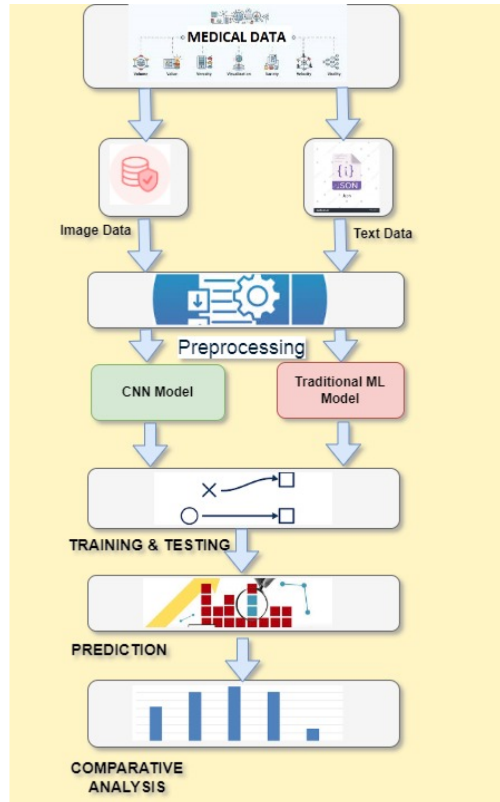

Fig. 2. Architecture – Breast Cancer Prediction System

D. Data Preprocessing

1) Image Data: Image dataset preprocessing is required for histopathology and mammogram images. This is because image datasets have their individual characteristics.

*2) **Data Augmentation**:* Images are transformed to synthetically increase the diversity of the training data. Most common techniques are random flipping (horizontal/vertical), slight angle rotation (for example, ±15°), zooming, and scaling/cropping brightness, contrast, and saturation adjustment. All these augmentations prevent over-fitting, especially in small datasets.

*3) **Resizing**:* Images need to be resized to some standard dimension, such as 224x224 pixels, to match the architecture of the CNN.

*4) **Normalization**:* Pixel value normalization to [0,1] range is by dividing by 255 for faster convergence while training.

*5) **Label Encoding**:* Convert the target labels (like " Benign" and "Malignant") to numerically encoded values (0 and 1).

*6) **Splitting the Dataset**:* The dataset would thus be divided into training, validation, and test datasets. Stratified sampling would be done to ensure that the split gives a consistent representation of classes across splits.

*7) **Balancing Classes**:* If there will be unequal images in the respective classes, it is solved by: Weighted loss functions during training and oversampling techniques for the minority class.

E. *Tabular Data*

Structured data like the Wisconsin breast cancer dataset would require some preprocessing steps to be prepared for ML models. Here are some of the main steps:

*1) **Data Cleaning: Handle Missing Values**:* It would be necessary to handle all the missing values from the dataset. Methods: Fill-in missing values with the mean, median, or mode (based on the feature), or remove rows/columns that include missing data. Outlier Detection: The recognition of outliers employing statistical methods like z-scores or visualizing tools like box plots. Outliers can be disruptive to models like logistic regression and KNN.

*2) **Data Transformation**:* Normalization or Standardization: In most cases, the sensitivity of ML models SVM, Logistic Regression, and KNN, towards scaling on features is because the efficiency of the performance on them is dependent on distance metrics. Feature Standardization (z-score normalization) or Min-Max scaling ensures all features contribute equally to the model.

*3) **Encoding Categorical Variables**:* Categorical variables, such as diagnosis, are captured into a number before being processed through one-hot encoding or label encoding. Once the category is binary classified, it is as easy as assigning number zero to benign cases while number one is assigned to malignant cases.

*4) **Feature Engineering**:* Identify the highly correlated features using correlation heat maps. Remove redundant features to reduce noise and enhance the interpretability of the model. For very large sets of features, PCA and other similar techniques are dimension reduction techniques to extract a huge number of features with maximum variance.

*5) **Data Splitting**:* Train the dataset on a training and testing basis, such as 80 percent training and 20 percent testing. Further divide the training dataset to provide training and validation for hyperparameter tuning.

*6) **Class Imbalance Handling**:* To deal with class imbalances if there is an uneven number of cases in the recorded classes (for example: cases more likely benign than malignant), one could use oversampling or SMOTE which would add cases to the minority class or the majority class's sample size would decrease through under sampling.

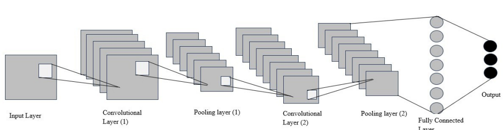

Fig. 3. Workflow of CNN Model

F. *Models*

*1) **CNN (Convolutional Neural Network)**:* CNNs operate to a great extent on image data. Figure 3 shows the process and reason on how they function differently: The principles of CNNs working: Convolutional layers detect spatial features like edges and textures, applying specific filters.

a) Pooling layers: Reduce the spatial dimensions but keep significant features.
b) Fully Connected Layers: Features are combined to make the final predictions. Challenges with image data:
c) Complexity: Image data is usually multidimensional and noisy, making training difficult and requiring a sizeable amount of computation and data.
d) Data size: CNNs need a lot of pictures for generalization. If they are not present, the model will overfit.
e) Feature extraction: Contrary to the tabular data, where features are predefined, CNNs have to learn features by themselves; thus, they become more prone to errors with less data.
f) Reasons why CNNs underachieved: Probably overfitting of CNNs by too little data. Histopathology and mammographic images are harder and less straightforward to extract features from than tabular data.

*2) **Traditional Models**:* The tabular data from the dataset includes engineered features (e.g., radius, texture, area). Here's how the models worked and why logistic regression found a niche in its utility.

a) Logistic Regression: It is a linear model, where the class probabilities are computed given the values of features. It works by identifying the hyperplane that best separates the two classes; hence, it performs well for a data set that is linearly separable or nearly linearly separable. Well-precomputed, well-scaled, and well-separable features might be possible reasons why logistic regression was more likely to work than other models on such a tabular dataset.
b) Decision Trees: This is another example of a non-linear model that divides the data into subsets and defines those subsets with some threshold on one or more features. A decision tree is highly prone to overfitting because of deep trees and a lack of regularization. It may do better if there are non-linear patterns present within the data.
c) K-Nearest Neighbors: Distance-based models classify by determining which class is the majority class among its K-nearest neighbors. Thus, very sensitive to feature scaling (normalized data is required). Drawbacks include difficulty in high-dimensional space or when the classes are unbalanced.
d) Gradient Boosting: It builds an ensemble by weak learners (e.g., decision trees) one after another, with each new learner attempting to improve the accuracy of the previous model on training data. Increased performance is generally seen in complex data that may have very subtle patterns, and these features may not be necessary for simple data that can be linearly separated.
e) Random forest: An ensemble of decision trees built with bagging. Overfitting will not be a major problem for this technique, and the technique can cope with data of both linear as well as non-linear types. It may not perform to its best for linearly separable data as compared to logistic regression. It may not perform at its best for linearly

separable data as compared to logistic regression.

f) Support Vector Machine: Finds the hyperplane that separates the classes well in high-dimensional space. Works for linearly and non-linearly separable data (using kernels). Expensive computation for large datasets.

IV. Results and Discussion

The result for the two approaches- Traditional ML models and image-based CNN model is compared using the metrics accuracy and F1 score. We have also analyzed the computational timing for training and testing the dataset, and the performance difference between the tabular data and image data. The results are visualized through methods like a confusion matrix, accuracy comparison, in Figure 4, and the performance of different models. The codes also display a graph that displays the prediction of the CNN model as well as the assessment metrics of the tabular model. The outcome highlights that CNN can extract complex patterns from images, the structured tabular dataset enables simpler models, like Logistic regression show higher performance.

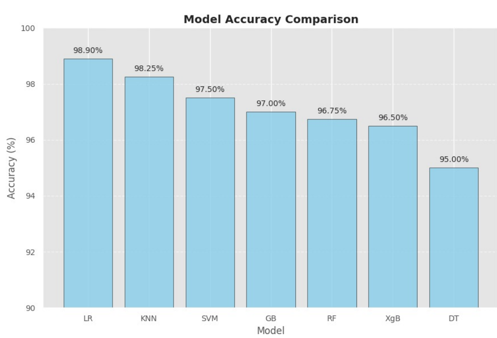

Fig. 4. Accuracy Comparison Of Different Models

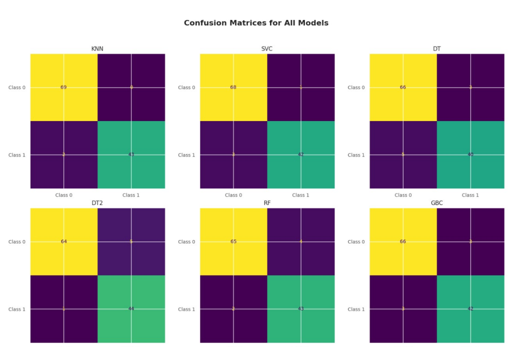

Fig. 5. Confusion Matrix of Different Classification Models

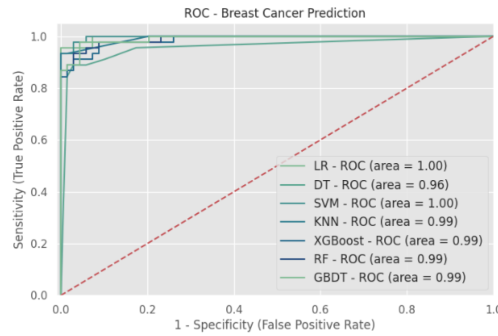

Fig. 6. ROC-Curve

A. Performance of Machine Learning Model

The ML models that were evaluated are: Logistic Regression, KNN, SVC, Gradient Boost, Random Forest, XGBoost, and Decision Tree. Each model was trained on a structured tabular dataset and tested to improve the accuracy for predicting breast cancer as malignant or benign. Among all of, in Figure 5, the Logistic regression achieved more accuracy in prediction of 98.25%, followed by KNN with 97.73%. The decision tree, being the simplest model, gave an accuracy of 92.11% , implying that a simpler model can yield better performance than many complex models on well-prepared datasets. The F1 score shows the model's effectiveness and its exceptional performance for classifications. The logistics regression achieved the highest F1 score of 0.99 and 0.98 for classes 0 and 1, showing the balance between precision and recall. This was followed by KNN with 0.98 and 0.97. The minimal difference between the F1 scores of classes 0 and 1 across all models($\leq$ 0.02)suggests balanced performances between the models, with no bias toward the class.

B. Confusion Matrix

The confusion matrix is a tabular depiction of a model's efficacy in classification tasks, detailing the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). It elucidates the model's capacity to accurately classify instances for each category and reveals potential biases or deficiencies. The diagonal elements signify correct predictions, whereas the off-diagonal elements denote misclassification. For instance, in a binary classification scenario, the top-left cell represents TNs, and the bottom-right cell represents TPs.

1) Interpreting Metrics from the Confusion Matrix: The comparison of the confusion matrix is shown in Figure 6, of all the machine learning models, illustrating classification performance.

Accuracy: Proportion of correctly predicted samples out of all samples.

Precision: Focuses on the relevance of positive predictions, defined as TP/(TP+FP).

Recall (Sensitivity): Measures the completeness of positive predictions, defined as TP/(TP+FN).

F1-Score: Harmonic mean of precision and recall, providing a balance between the two.

2) Comparative Comment: Analyze variations in the matrix across models. For instance, a higher FP count in one model may indicate a bias towards the positive class.

3) ROC Curve Explanation: Visual comparison of the discriminative capacities of all models is made possible by overlaying ROC curves in Figure 7 for each on a single graph. Differences in curve shapes or AUCs highlight which model performs best under varying conditions.

C. Image Based Prediction

For image-based prediction, we used a Convolutional Neural Network(CNN) for histopathological images in Figure 8. The model showed an accuracy of 93.72% which shows the ability to capture complex patterns. However, the performance of CNN when compared with the other traditional machine learning models is very low, indicating the challenges associated with image-based prediction.

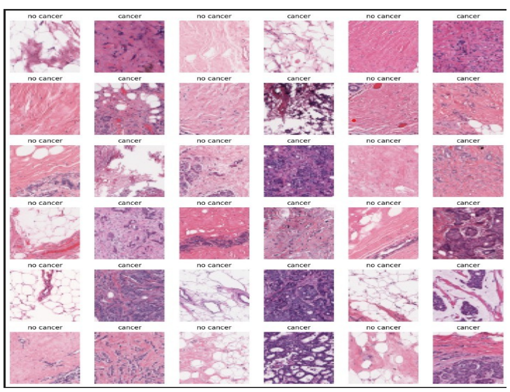

Fig. 7. Image dataset-CNN

Challenges in Image-Based Prediction

- Variability in Images: Staining differences, magnification differences, and differences among tissues have considerable effects on feature extraction.
- Dataset Size: Generally, CNNs need large datasets for optimum generalization, and this may have stopped the image dataset size from reaching an optimal performance level.
- Computational Complexity: Highly computational resource-demanding use of models for training by tuning hyperparameters severely limits their effectiveness in this study.

D. Comparative Analysis

A comparative analysis indicates that the method structured data outperformed predictions from images above in both accuracy and interpretability. In structured data, Logistic Regression achieved the highest accuracy of all models tested, while CNN proved the least fit in the handling of image variability.

- **Logistic Regression:** Managed to outperform CNN in both its accuracy and precision measures when applied to sets of structured datasets.
- **CNN Limitations:** CNN is not particularly effective for the variations seen in histopathological images, requiring larger datasets and sophisticated pre-processing for optimal performance.
- **Comparison with Other Models:** Despite their complexity, CNNs produced results that compared favorably with Random Forest and XGBoost. However, both were slightly less effective than Logistic Regression on structured data.

E. Implication for Breast Cancer Prediction

In the grand scheme of things, these findings hold promise for breast carcinoma diagnosis and treatment planning:

- **Structured Data Models:** High accuracy achieved on the structured dataset by the two models, Logistic Regression and Random Forest, underscores the importance of well-annotated clinical data in predictive modeling. These models are computationally efficient and interpretable for a clinical setting.
- **Image-Based Models:** While their performance may be inferior to other systems, CNNs have demonstrated potential for challenging pattern recognition tasks in histopathological images. Future studies could involve training such models with augmented datasets, publicly available datasets, and hybrid architectures combined with other components.
- **Future Directions:** A promising direction for future research is the development of integrated models that combine structured clinical data with features extracted from images. Such models could leverage the strengths of both types of data to improve diagnostic accuracy and provide deeper insights into difficult cases.

V. Future Work

- Enhance prediction accuracy by integrating advanced deep learning models.
- Use explainable AI techniques to clarify how model decisions are made.
- Expand training with larger and more diverse datasets.
- Incorporate d ata from wearable devices for ongoing monitoring.
- Improve the system to offer personalized treatment recommendations.
- Perform clinical trials to verify reliability in healthcare environments.

VI. Conclusion

This study effectively demonstrated the application of methods to address problems. The suggested framework, which uses certain techniques, has resulted in significant breakthroughs in important findings. Our experiments demonstrate the approach's robustness and dependability, focusing on its practical applicability in a specific application domain. Furthermore, adopting any unique methodology has shown considerable improvement in performance over previous methodologies. Although the findings are encouraging, there is still much room for future inquiry. Furthermore, adding a relevant component, such as AI models, user input, or sophisticated analytics, may improve the system even further. This research serves as a starting point, providing information for developing more complete and efficient solutions in the relevant industry.

REFERENCES

[1] A. Priyadarshini and J. Aravinth, "Correlation Based Breast Cancer Detection using Machine Learning," in *Proceedings of the 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*, Bangalore, India, 2021, pp. 499–504, doi: 10.1109/RTEICT52294.2021.9573818.

[2] O. L. Mangasarian and W. H. Wolberg, "Cancer diagnosis via linear programming," *SIAM News*, vol. 23, no. 5, pp. 1–18, Sep. 1990.

[3] Md T. Ahmed, et al., "Analysis of Wisconsin Breast Cancer Dataset," *J. Sci. Tech. Environ. Informatics*, vol. 9, pp. 120–127, 2020.

[4] A. J. B and S. Palaniswamy, "Comparison of Conventional and Automated Machine Learning Approaches for Breast Cancer Prediction," in *Proceedings of the 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2021.

[5] L. Zou, et al., "A Technical Review of Convolutional Neural Network-Based Mammographic Breast Cancer Diagnosis," *Computational and Mathematical Methods in Medicine*, vol. 2019, pp. 1–14, 2019.

[6] T. Indu V, G. Dev M, and K. Rejini, "Leveraging Deep Learning for Accurate Diagnosis of Breast Cancer," *African Journal of Biological Sciences*, vol. 6, no. 14, pp. 728–739, 2024.

[7] K. Gupta and N. Chawla, "Analysis of Histopathological Images for Prediction of Breast Cancer," *Procedia Computer Science*, vol. 167, pp. 157–164, 2020.

[8] Md M. Hossin, et al., "Effective Comparison of Different ML Algorithms for Breast Cancer Detection," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 4, pp. 3001–3010, 2023.

[9] X. Liu, et al., "A Hybrid Deep Learning Model for Breast Cancer Classification Using Histopathological Images," *IEEE Access*, vol. 8, pp. 75130–75141, 2020.

[10] J. Arevalo, et al., "Representation Learning for Mammography Mass Lesion Classification with Convolutional Neural Networks," *Computer Methods and Programs in Biomedicine*, vol. 127, pp. 248–257, 2016.

[11] M. El Houby, "Breast Cancer Diagnosis on Medical Imaging Using Machine Learning Techniques," *Computers, Materials & Continua*, vol. 66, no. 3, pp. 2927–2943, 2021.

[12] H. Al-Shaikhli, et al., "A comparison of ensemble techniques for breast cancer diagnosis using WDBC dataset," *Journal of Biomedical Informatics*, vol. 98, pp. 103277, 2019.

[13] R. Atajanov and S. Peker, "Performance Comparison of Different ML Techniques for Early Prediction of Breast Cancer using Wisconsin Breast Cancer Dataset," in *2022 3rd International Informatics and Software Engineering Conference (IISEC)*, IEEE, 2022, pp. 140–147.