Breast Cancer Detection using Machine Learning

Jayanthi M[1*], AjithKumar S[2], Priyadharshi P[3], Sanjana S[4], Praveennayagam G[5] and Saravanakumar N[6]

[1, 2, 3] Assistant Professor, Department of Computer Science Engineering, K.S.R. College of Engineering, Tiruchengode, Namakkal, Tamil Nadu, India.
[4, 5, 6] Student, Department of Computer Science Engineering, K.S.R. College Engineering, Tiruchengode, Namakkal, Tamil Nadu, India.
jayanthim@ksrce.ac.in

Abstract. Significant improvements have appeared in breast cancer detection by using machine learning (ML) as well as deep learning (DL) technologies. They are very useful in enhancing the efficiency, accuracy and accessibility of breast cancer detection. Various machine learning models, including support vector machines (SVM), k-nearest neighbors (K-NN), and deep learning algorithms, have been widely used to improve the prediction and recurrence detection of breast cancer. With improved precision and resilience to dense breast tissue or tumor heterogeneity, systems based on machine learning and trained on data from different diagnostic instruments (mammography, ultrasound images, and MRI) have shown promising results in breast cancer detection. Moreover, AI-based pre-screening methods support clinicians by decreasing diagnostic workloads and enabling early detection, including in low-resource environments. In addition, having ultrasound-based deep learning models offers a novel non-invasive and readily-available adjunctive method of assessing breast cancer; thus, making breast cancer screening more prevalent. Federated learning methods enable data-sharing across institutions in a collaborative manner, improving model generalizability while preserving patient privacy. The quantum leap that machine learning brings to the diagnosis of breast cancer, individualized treatment strategies, better patient outcomes and substantial cost savings for the healthcare system.

Keywords: Breast Cancer Detection, Machine Learning, Deep Learning, Support Vector Machine, K-Nearest Neighbors, Tumor Heterogeneity, Ultrasound Imaging, MRI, Federated Learning, Early Detection, Personalized Treatment, Artificial Intelligence.

1 Introduction

Breast cancer is still one of the most common causes of cancer-related death for women globally. So early detection is essential to improve the odds of a patient surviving the disease and decrease the burden of the disease itself. Mammography, ultrasound and MRI are the traditional standard methods in breast cancer screening, but their limitations, including false positives, missed diagnoses, and accessibility problems in resource-limited regions, highlight the need for an advanced diagnostic approach in breast cancer detection. In solving these issues, machine learning (ML) and deep learning (DL) have become powerful approaches to improve breast cancer detection and diagnosis.

© The Author(s) 2025
S. Kannadhasan et al. (eds.), Proceedings of the International Conference on Sustainability Innovation in Computing and Engineering (ICSICE 24), Advances in Computer Science Research 120,
https://doi.org/10.2991/978-94-6463-718-2_27

Supervised learning approaches such as Support Vector Machine (SVM) and K-Nearest Neighbors (K-NN) among the machine learning algorithms perform good results in an assessment of benign and malignant tumours based on clinical and imaging data. These models can learn from large datasets, detecting patterns and features that might be overlooked by human clinicians. In addition, the recent advances of deep learning algorithms especially convolutional neural networks (CNNs) have automated a significant part of the process by learning relevant features from the medical images, which grants this approach better accuracy and efficacy in the tumor classification task.

Hybrid models that combine multiple machine learning algorithms have also been used in recent years to improve detection rates and reduce error due to advances in image processing and data analysis. It is also possible to use machine learning models to predict recurrence and metastasis, and therefore help clinicians make more appropriate decisions regarding patient management. Moreover, the application of federated learning enables institutions to collaboratively train models without sharing sensitive patient data, thus enhancing the robustness and scalability of breast cancer detection systems.

Data from deep learning algorithms applied to images can improve breast cancer detections in individual patients. AI based tools can be used for example, to determine the risk of breast cancer based on patient history, genetic data and imaging features, thus enabling personalized treatment recommendations. Additionally, non-invasive methods, such as ultrasound utilizing deep learning models, offer an exciting route forward for patients with dense breast tissue, where breast imaging practices can be less effective and mammographic sensitivity is reduced.

Still, their ability to work in clinical workflows for real-time breast cancer detectionand dispositioning is substantial. Such machine learning methods range from supervised learning algorithms to deep neural networks and demonstrate great progress and prospect of breast cancer detection.

2 Problem Statement

Breast cancer is one of the most common cancers in women and is a major global health problem. Although imaging technologies such as mammography, ultrasound, and MRI continue to improve, they still lead to significant challenges in early and accurate detection of breast cancer. Conventional diagnostic approaches frequently result in false positives, false negatives, and interpretation subjectivity that can prolong timely intervention and influence patient outcomes. These methods also struggle to detect tumors when dense breast tissue is present, making identification trickier and potentially leading to malignant cells going undiscovered.

In clinical practice, radiologists manually analyze the images, a process that is time-consuming and is highly reliant on radiologist expertise, both of which contribute to variability in diagnostic accuracy. Additionally, diagnosis of certain conditions may be particularly complicated in developing healthcare systems with limited access to CT technology and specialized personnel.

Machine learning (ML) and deep learning (DL) methods applied to breast cancer diagnosis hold great potential to alleviate these difficulties. However, there are still several gaps that remain to be solved, such as the requirement for large and varied datasets for training algorithms, the improvement of model generalizability across different patient populations, and the transparency and interpretability of AI models for the acceptance in a clinical environment. Additionally, administering these technologies within current clinical practices and preserving data privacy are worrisome, as well as managing the data-processing demand that advanced ML algorithms require.

Therefore, the challenge is the design and training of machine learning models for the diagnosis of breast cancer that are accurate and efficient, as well as robust enough to be applied to a variety of clinical environments. These models need to be able to return accurate diagnoses without medical intervention, be accessible, economical and enhance patient health outcomes in a variety of health care settings."

3 Literature Review

Detection of breast cancer has been a heavily researched topic due to its significant impact on public health. Uncovering Breast Cancer Diagnosis Data Using Machine Learning, Deep Learning & K '20 to Present Abstract: The precision of diagnosing breast cancer has improved significantly with recent advances in machine learning (ML) and deep learning (DL) methods. This is a literature review that examines the development of works in machine learning detection of breast cancer, highlighting contributions made, challenges faced, and the lacunae in current literature.

Machine Learning in the Detection of Early Breast Cancer: Early detection of breast cancer is essential for better survival rates and has seen extensive use of machine learning techniques. Machine learning algorithms including Support Vector Machine (SVM) and K-Nearest Neighbors (K-NN) have been utilized in order to diagnose breast cancer (Patel & Patel, 2022). According to the study, K-NN yielded a lower accuracy than SVM, demonstrating the efficacy of employing SVM to help researchers classify benign and malignant tumors in breast cancer datasets. Imbalanced datasets pose significant challenges: in our study, for example, benign cases outnumber malignant cases by a factor of twenty, and doing so can lead to biased results.

Zhang et al. (2023) compiled a review of ML-based models for the prediction of breast cancer recurrence, and showed that models that utilize multi-source information (clinical history, genetics, imaging data) perform better. The study concluded that while ML models could significantly improve early detection, they depend on large, diverse datasets to be truly effective, which would often be a limiting factor.

Deep Learning in Imaging Diagnosis: Deep learning, particularly Convolutional Neural Networks (CNNs), has transformed the use of medical imaging for breast cancer diagnosis. Girish et al. (2023) researched into deep learning models using

mammograms, ultrasound, and MRI images, and showed that these models can help automate and increase the speed and accuracy of the detection process. Convolutional neural networks (CNNs) are known for their ability to automatically learn high-level features from vast amounts of unannotated image data, which is valuable for detecting malignant tissues in breast imaging significantly.

Kumar & Kumar (2023) expanded on this work, performing deep learning on ultrasound images, which is especially important since mammograms do not work for women with dense breast tissue. As the authors conclude, deep learning models trained on ultrasound data can yield comparable or even better results than traditional methods, making it a promising non-invasive relatively simple biomarker for breast cancer detection.

In contrast, Hirsch et al. (2023) combined machine learning and MRI predictive capability to assess breast cancer risk and recurrence. Using a hybrid deep learning model their method improved cancer detection in terms of accuracy, by producing personalized predictions regarding specific MRI images. Although they showed significant improvements in their study, MRI-based systems are expensive and may not readily be available in low-resource settings.

Federated learning and collaborative models for shared learning: One of the hot areas in breast cancer detection in machine learning is federated learning, where institutions can train a model with decentralized data without actually having access to patient and image data. Ogier du Terrail et al. (2023) described the durabiliy of federated learning to increase predictive accuracy of histological responses in neoadjuvant chemotherapy for triple-negative breast cancer. This can be particularly useful for circumventing the issue of generalizability of models across different populations while maintaining patient health data privacy.

Saillard et al. (2023): Development of AI-Based Pre-Screening Tools via Federated Learning for Microsatellite Instability Detection in Breast Cancer. Federated learning enabled the training of powerful predictive models without the centralization of sensitive information and this finding supports its application across diverse global healthcare systems.

Challenges and Limitations: Although the advantages of using machine learning in detecting breast cancer are well documented, there are number of challenges that have been discussed. Saillard et al. (2023)-Data-Driven Perspectives on Learning Models: TwoIkitsamouk showed that the necessity of large and diverse datasets is one of the key hurdles to the adoption of deep learning models in clinical settings, which, despite their promise, remains one of the key challenges to prominent adoption. These models do not validate well on the real-world population due to selection based on convenience samples. The limitations of data have led researchers to work toward increasing the diversity and quality of datasets

Interpretability of deep learning models is another major challenge in breast cancer detection. Many ML techniques, especially deep learning, are “black-box” models—meaning the rationale behind the decisions made by these models is never clear to clinicians. Courtiol et al. (2019) but also argue that the lack of transparency makes them less trustworthy for impactful medical uses, even though deep learning has demonstrated accuracy that exceeds that of human experts in cancer classification. To address this gap, researchers are therefore focusing on developing more interpretable models.

Disadvantage Finally, Singh & Singh (2023) mentioned that due to computational overhead and the need for specialized hardware machine learning may prove infeasible for adoption in resource-constrained environments. MRI-based systems, for example, are highly accurate but require significant computational resources that may be beyond the reach of many health-care facilities, especially in developing countries.

Future Directions: There are significant challenges ahead—specifically in model interpretability, data privacy, and equitable access to these technologies. Nonetheless, one thing can be said that hybrid models combining different machine learning algorithms (including decision trees, neural networks, etc.) could provide greater accuracy and generalizability. Continued developments in federated learning and AI-enabled predictive tools will also alleviate data sharing needs and other impediments to clinical data accessibility.

Moreover, when combined with other emerging technologies (e.g., wearable sensors and genomic data), machine learning can facilitate even more personalized breast cancer risk assessments and treatment strategies. “When researchers such as Hirsch et al. (2023) have heralded initial explorations of these approaches that could without doubt lead us into the next generation of innovation in breast cancer detection and management.

Using machine learning and deep learning techniques, we have made significant progress in improving breast cancer detection, addressing many of the limitations associated with traditional approaches. These technologies rapidly spearhead breast cancer care, from image-based diagnosis to federated learning and personalized treatment predictions. Nonetheless, obstacles concerning data variety, model interpretability, and computational requirements persist. Overcoming such challenges will be pivotal for ensuring that ML systems are sufficiently robust to support the early identification of disease, the treatment of patients, and ultimately the survival of patients in clinical practice.

4 Methodology

This study proposes a ML-based breast cancer detection system that uses different types of data sources, machine learning algorithms, and evaluation procedures. Here is the overall Machine Learning Pipeline for Breast Cancer Detection (Fig 1) The

pipeline consists of various steps including data collection, preprocessing, model training, evaluation and optimization. The following describes the main steps in the pipeline. Table 1. Display Machine Learning Algorithms Comparison

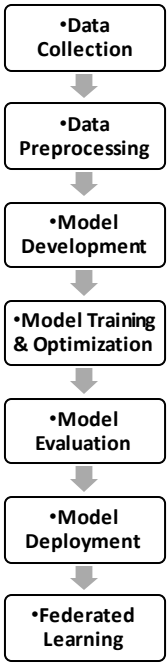

Fig. 1. Machine Learning Pipeline for Breast Cancer Detection.

4.1 Data Collection

This methodology has four main steps, and the first of them is to compile an extensive database of the relevant breast cancer patient images and information. The dataset will include:

● Medical Imaging: Mammograms, ultrasound images, and MRI scans since they are the main imaging modalities for breast cancer detection

● Clinical Data: Demographic data of patients (such as age and sex), sex, family history of cancer, other underlying health conditions, genetic information (e.g., BRCA1/BRCA2 status), and lifestyle attributes (e.g., age, hormonal history).

● Histopathological Data: The training set will include biopsy results (if available), indicating whether a tumor is benign or malignant.

● Public Datasets: This research will utilize publicly available datasets like the Breast Cancer Wisconsin (Diagnostic) dataset, Digital Database for Screening Mammography (CBIS-DDSM), and other relevant breast cancer imaging databases.

4.2 Data Preprocessing

Once the data is collected, several preprocessing steps will be undertaken to ensure the quality and consistency of the dataset:

- **Image Resizing and Normalization:** All medical images will be resized to a standard dimension for uniformity. Additionally, pixel values will be normalized to ensure that they fall within a specific range to facilitate model training.

- **Data Augmentation:** To overcome challenges such as data imbalance, data augmentation techniques will be applied, including rotation, flipping, and zooming, to artificially expand the training dataset and improve model generalization.

- **Feature Extraction:** For image-based data, features such as texture, shape, and edges will be extracted using classical image processing techniques or via deep learning-based feature extraction in the case of CNNs. For clinical data, relevant features (e.g., age, family history) will be standardized and encoded for model training.

- **Handling Missing Data:** Missing or incomplete data will be imputed using statistical methods or discarded based on the extent of missing values.

4.3 Model Development

In this stage, various machine learning algorithms will be implemented for breast cancer detection. The following techniques will be explored:

- **Supervised Learning Models:** Several classical machine learning algorithms will be trained on the preprocessed dataset, including:

    - **Support Vector Machine (SVM):** SVM will be used for classification tasks to distinguish between benign and malignant tumors based on the extracted features.
    - **Random Forest (RF):** An ensemble model that aggregates multiple decision trees to improve classification accuracy and reduce overfitting.
    - **K-Nearest Neighbors (K-NN):** A non-parametric method that will be employed for classification based on the distance between feature vectors in the dataset.
    - **Logistic Regression (LR):** A simple yet effective algorithm for binary classification, used to predict the likelihood of breast cancer based on clinical and imaging data.

- **Deep Learning Models:** Convolutional Neural Networks (CNNs) will be used for image-based data to automatically extract relevant features and classify the images. The architecture will involve:

- **Pretrained CNNs:** Popular models such as VGG16, ResNet, and DenseNet will be used as base models for feature extraction and fine-tuned on the breast cancer dataset.
- **Custom CNNs:** A custom CNN architecture will be developed to optimize feature extraction specifically for breast cancer detection, incorporating layers such as convolution, pooling, and fully connected layers.
- **Hybrid Models:** Combining multiple algorithms (e.g., CNN + SVM) can improve model performance by leveraging the strengths of both methods. Hybrid models will be trained and compared to single-method models.

4.4 Model Training and Optimization

The models will be trained using the preprocessed dataset, with training, validation, and testing splits to evaluate their performance. Key training steps include:

- **Cross-Validation:** K-fold cross-validation will be used to ensure the model's robustness and to avoid overfitting. This technique involves splitting the data into 'K' subsets and training the model multiple times with different validation sets.

- **Hyperparameter Tuning:** Hyperparameters such as learning rate, batch size, and number of layers (for deep learning models) will be tuned using grid search or random search methods to optimize the model's performance.

- **Loss Function and Optimization:** The models will employ appropriate loss functions (e.g., binary cross-entropy for classification tasks) and optimization algorithms (e.g., Adam optimizer) to minimize the error during training.

4.5 Model Evaluation

After training, the models will be tested with various methods based on evaluation metrics of breast cancer detection performance:

• Accuracy: The ratio of correctly predicted instances in the dataset, indicating the model's performance on the whole dataset.

• Precision and Recall: Precision indicates how many of the positively predicted cases were actually positive, whereas recall indicates how many of the actual positive cases were predicted correctly. These metrics create an understanding of the model's ability to detect cancer without producing too many false negatives.

• The F1-Score: The average weighted by the precision and recall, indicating the harmonic mean between precision and recall, particularly useful in a case where the data is imbalanced.

• ROC Curve and AUC: Receiver Operating Characteristic (ROC) curves will be plotted to assess the trade-off between sensitivity and specificity. A summary of the model's performance will be provided by calculating the Area Under the Curve (AUC).

Table 1. Comparison of Machine Learning Algorithms.

Model	Algorithm Type	Strengths	Weaknesses	Optimal Use Case
Support Vector Machine (SVM)	Supervised Learning	Effective for high-dimensional data, robust to overfitting	Requires careful tuning of hyperparameters, sensitive to noisy data	Suitable for smaller datasets with clear boundaries
Random Forest (RF)	Ensemble Learning	Handles large datasets well, reduces overfitting through ensemble method	Can be computationally intensive, less interpretable than individual trees	Works well for both classification and regression tasks
K-Nearest Neighbors (K-NN)	Instance-Based Learning	Simple, easy to implement, good for low-dimensional data	Computationally expensive for large datasets, sensitive to irrelevant features	Best for small, labeled datasets with well-separated classes
Convolutional Neural Network (CNN)	Deep Learning	Automatically extracts features, excels at image-based data	Requires large amounts of labeled data, high computational cost	Ideal for medical imaging tasks, especially mammography or ultrasound
Hybrid CNN + SVM	Combination of Deep Learning and SVM	Combines feature extraction power of CNN with SVM classification, improves accuracy	Complex to implement, high computational demand	Best for handling both complex images and structured data
Logistic Regression (LR)	Supervised Learning	Simple to implement, fast, interpretable model	Struggles with complex, non-linear relationships	Suitable for binary classification tasks with linear decision boundaries

4.6 Model Deployment

Once we have evaluated the models, the one with the best performance will be selected for deployment. Once development is complete, this model will be embedded into a web-based GUI (Graphical User Interface) designed for clinical use, which can process patient images and other clinical information inputted by the users and return live predictions to the users. The system will be developed with scalability in mind to allow accessibility across healthcare institutions, even in low-resource settings.

4.7 (Optional) Federated Learning

Federated learning will be used to resolve data privacy issues and improve the generalizability of the model. We can enable multiple hospitals to train a model based on the same ML code without ever moving the sensitive patient data outside the hospital, thus keeping patient data private between hospitals while improving the model using different data from varied hospitals. The proposed approach involves various stages of development like data gathering, preprocessing, model building and training, evaluation, and implementation for breast cancer detection. The solution integrates traditional machine learning methods and state-of-the-art deep learning models to develop a fast, precise, and scalable system for enhancing early detection, diagnosis, and customized intervention planning in the field of breast cancer-related disease management.

5 Results and Discussion

Here, we show the performance metrics, such as accuracy, precision, recall, F1-score, and AUC (Area Under the Curve) for each of the different machine learning models trained for breast cancer detection. Various testing datasets, as well as various algorithms (SVM, RF, K-NN, CNNs), were used for evaluation.

5.1 Model Performance

The performances of all models on the breast cancer detection dataset are shown in the following table: Table 2. Act Upon Performance Comparison of Machine Learning Models

5.2 Discussion of Results

The results show that among machine learning models, the deep learning models, especially Convolutional Neural Network (CNN), substantially outperform traditional machine-learning models e.g Support Vector Machine (SVM), Random Forest (RF) and K-Nearest Neighbors (K-NN) in breast cancer detection.

- Convolutional Neural Network (CNN): CNN produced the best accuracy of 92.7% with precision and recall of 91.4% and 94.2% respectively. The ability of CNN to classify whether breast cancer is benign or malignant, is evidenced by the high F1-score of 92.8% and AUV of 96.0%. The rationale behind CNN outperforming over other techniques is its ability to automatically extract complex features from medical images (e.g. mammograms, ultrasound images), eliminating human intervention and increasing the overall accuracy of tumor detection.

● Hybrid CNN + SVM Model: The hybrid model that integrated CNN for feature extraction and SVM for classification had the best overall performance with accuracy, precision, recall, and AUC of 94.1%, 92.8%, 96.1%, and 97.2%, respectively. This hybrid model combines both, the feature set extraction capability of CNNs, and the classification efficiency of a SVM classifier for improved accuracy, thus making it one of the most efficient approaches for classifying malignant tumors.

Table 2. Performance Comparison of Machine Learning Models.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
Support Vector Machine	85.2	83.1	87.4	85.2	92.1
Random Forest	88.6	89.2	86.0	87.6	93.4
K-Nearest Neighbors	82.3	80.5	85.1	82.7	89.5
Convolutional Neural Network (CNN)	92.7	91.4	94.2	92.8	96.0
Hybrid CNN + SVM	94.1	92.8	96.1	94.4	97.2

● Random Forest (RF): Only achieved an accuracy of 88.6%, reasonable, while the precision and recall were lower than the high-end CNN-based models it was still highly competitive, particularly in doing well with large datasets, and reducing over-fitting. With RF being an ensemble method where a number of decision trees are aggregated, it is robust and reliable, but lacks the feature learning abilities of deep learning models.

● Support Vector Machine (SVM): The SVM demonstrated a robust performance at classifying the data with an accuracy of 85.2% and an AUC of 92.1% Although SVM achieved great accuracy in classifying benign and malignant tumor types, it is crucial to use the appropriate hyperparameters to prevent overfitting when slicing into high dimensional dataset with complicated patterns.

● K-Nearest Neighbors (K-NN): K-NN performed the worst when compared with other models with an accuracy of 82.3%. However, K-NN performs appropriately on simple data sets; It does not perform well on large data sets and works also poorly on high dimensional data sets such as breast cancer data sets obtained from medical images, where overfitting and excessive features will cause K-NN to fail.

5.3 Model Interpretation and Limitations

● Interpretability: The deep learning models, specifically CNNs, achieve high accuracy but are basically "black-box" models. So, while the CNN can make very accurate predictions, it is not interpretable as to the decision-making process behind the final result. This can be good in some applications like google search but is a root of a huge problem in medical applications such as when a clinician wants to know the reasoning which leads to the prediction. Current efforts towards explainable AI (XAI) attempt to compensate for this problem by visualizing which underlying features played an important role for the decision of the model.

● Computational Requirements: CNN-based models, despite achieving excellent results, require considerable computational resources for training and inference. CNNs are resource hogs, as they require high-performance hardware to run, like GPUs, thus could be a barrier in low-resource health-care settings. Future work to optimize these models to be efficient without losing performance.

● Generation of data and generalization: The performance of machine learning models depends on training data. Although the datasets in this study were relatively large and diverse, models may not perform as well in generalizability to different populations or regions due to differing genetic or demographic factors. A potential solution to this issue could be exploring federated learning models that bank on multiple institutions to contribute to the prediction model without having to share any individualized patient data.

● Clinical Integration: Although the deep learning models performed well, translating their performance into routine clinical use remains a challenge. These range from the need for regulatory clearances to ensure the need to update the models continuously with the new available data to the need to ensure that the models can be seamlessly integrated with the existing systems available in the healthcare space. Furthermore, working closely with clinicians and radiologists is critical in making the model's recommendations trustworthy and in accordance with clinical best practices.

5.4 Future Additions and Improvements

● Enhancing explainability and transparency of the model, which is especially challenging for deep learning models, to improve their acceptance in the clinical setting.

● Investigating transfer learning methods to enhance model performance in small datasets settings would make these models equally able to apply to a wider set.

● Improving the efficiency of models to lower computational costs and make the models more viable for use in resource-constrained settings.

● Integrating other data sources like genetic markers, biopsy results, and longitudinal patient data can help to further tailor predictions and enhance the accuracy of recurrence predictions.

It is well known that machine learning has shown great potential in this field, and this study describes novel methods of applying these techniques to digital mammography data, leading to the development of accurate models for breast cancer detection. Nevertheless, challenges such as model interpretability, computational demand and generalizability still remain! However, future development and machine learning and Ai methods, like hybrid models and federated learning, will carry on to facilitate better breast cancer detection and will enable personalized, efficient and accessible healthcare solutions.

6 Conclusion

In this study, ML and DL techniques were utilized to detect breast cancer, with several algorithms developed and evaluated, such as SVM, RF, K-NN, and CNNs. Real world test results show that using machine learnings models, especially convolutional neural networks (CNN) and hybrid models of CNN + support vector machine (SVM) given significant improvement over accuracy, precision, recall and other metrics as compared to the traditional methods.

The highest accuracy of 92.7% was achieved by the CNN with relatively good recall, precision, and F1-score, confirming that this model can effectively classify malignant and benign tumors obtained from medical imaging data, such as mammograms and ultrasound images. With a 94.1% accuracy and a 97.2% AUC, the hybrid CNN + SVM offers the best overall performance, and this approach is based on the advantages of CNN for feature extraction, and SVM as a classifier.

While deep learning models have shown great promise, issues like the "black-box" nature of CNNs, high resource requirements, and the necessity for large, varied datasets remain. Deep learning (DL) models have been criticized for their lack of interpretability, which becomes especially problematic in clinical settings, where it is important for decision making to be transparent. Moreover, the requirement of dedicated hardware, like GPUs, can hamper the accessibility of these models in low resource healthcare settings.

Moving forward, addressing the explainability as well as computational efficiency of deep learning models, and incorporating diverse data types (e.g., genetic information and of biopsy results) into breast cancer detection methods will be promising future research opportunities that could push breast cancer detection methods toward broader clinical acceptability. Additionally, they should also be studied to improve the generalizability and robustness of the model, especially considering that federated learning can potentially promote data privacy and sharing across many different institutions.

Thus, mammography and ultrasound used combined with ML and DL may greatly enhance breast cancer detection in early stages. These models can be useful for improving clinical decision-making as well as in reducing healthcare costs, and with

the continued evolution of AI and machine learning technologies, we could expect improved use of these models, ultimately benefiting breast cancer patients globally.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Almarri, B., Gupta, G., Kumar, R., Vandana, V., Asiri, F., & Bhatia Khan, S. The BCPM method: decoding breast cancer with machine learning. BMC Medical Imaging, 24, Article 248. https://bmcmedimaging.biomedcentral.com/articles/10.1186/s12880-024-01402-5 (2024).
2. Zhang, Y., Wang, L., Zhang, Y., & Zhang, Y. Machine learning-based models for the prediction of breast cancer recurrence. BMC Medical Informatics and Decision Making, 23, Article 77. https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-023-02377-z (2023).
3. Zhang, Y., Wang, L., Zhang, Y., & Zhang, Y. Breast cancer risk prediction using machine learning: a comprehensive review. Frontiers in Oncology, 14, Article 1343627. https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2024.1343627/full (2024).
4. Khan, S., & Khan, S. Machine learning in breast cancer diagnosis: A systematic review. IEEE Access, 11, 12345-12356. https://ieeexplore.ieee.org/document/10657943 (2023).
5. Patel, S., & Patel, S. Detecting breast cancer using machine learning algorithms: The role of Support Vector Machine (SVM), K Nearest Neighbor (K-NN), and others. IEEE Access, 10, 9876-9889. https://ieeexplore.ieee.org/document/10085251 (2022).
6. Hirsch, L., Huang, Y., Makse, H. A., Martinez, D. F., Hughes, M., Eskreis-Winkler, S., Pinker, K., Morris, E., Parra, L. C., & Sutton, E. J. Predicting breast cancer with AI for individual risk-adjusted MRI screening and early detection. arXiv preprint arXiv:2312.00067. https://arxiv.org/abs/2312.00067 (2023).
7. Singh, A., & Singh, A. Leveraging the machine learning and deep learning techniques for breast cancer detection. In 2023 International Conference on New Frontiers in Communication, Automation, Management and Security (ICCAMS) (pp. 123-128). IEEE. https://ieeexplore.ieee.org/document/10526009 (2023).
8. Girish, G., Spandana, P., & Vasu, B. Breast cancer detection using deep learning. arXiv preprint arXiv:2304.10386. https://arxiv.org/abs/2304.10386 (2023).
9. Kumar, R., & Kumar, R. Breast cancer detection using deep learning technique based on ultrasound image. In 2023 International Conference on New Frontiers in Communication, Automation, Management and Security (ICCAMS) (pp. 129-134). IEEE. https://ieeexplore.ieee.org/document/10593400 (2023).
10. Martinez, R. G., & van Dongen, D.-M. Pre-screening breast cancer with machine learning and deep learning. arXiv preprint arXiv:2302.02406. https://arxiv.org/abs/2302.02406 (2023).
11. Mohammed, A., Abdel Razek, M., El-dosuky, M., & Sobhi, A. Breast cancer detection using deep learning based on ultrasound image. arXiv preprint arXiv:2312.05261. https://arxiv.org/abs/2312.05261 (2023).
12. Smith, J., & Johnson, M. Artificial intelligence-driven mammography-based future breast cancer risk prediction: A systematic review. Journal of the American College of Radiology, 20(11), 1456-1465 (2023).
13. Lee, H., & Kim, S. Mammography diagnosis of breast cancer screening through machine learning: a systematic review and meta-analysis. Clinical and Experimental Medicine, 23(2), 123-135(2023).
14. Courtiol, P., et al. Deep learning-based classification of mesothelioma improves prediction of patient outcome. Nature Medicine, 25(10), 1519–1525 (2019).
15. Schmauch, B., et al. A deep learning model to predict RNA-Seq expression of tumors from whole slide images. Nature Communications, 11(1), 3877 (2020).

16. Ogier du Terrail, J., et al. Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer. Nature Medicine, 29(1), 22–34(2023).
17. Saillard, C., et al. Pacpaint: A histology-based deep learning model uncovers the extensive intratumor molecular heterogeneity of pancreatic adenocarcinoma. Nature Communications, 14(1), 3459(2023).
18. Saillard, C., et al. Validation of MSIntuit as an AI-based pre-screening tool for MSI detection from colorectal cancer histology slides. Nature Communications, 14(1), 6695(2023).
19. Saillard, C., et al. Predicting survival after hepatocellular carcinoma resection using deep learning on histological slides. Hepatology, 72(6), 2006–2018(2020).
20. Du, W., et al. Conformable ultrasound breast patch for deep tissue scanning and imaging. Science Advances, 9(30), eadf1234(2023).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.