Check for updates

Wiley
International Journal of Intelligent Systems
Volume 2025, Article ID 8884481, 19 pages
https://doi.org/10.1155/int/8884481

WILEY

Research Article

An Enhanced Approach for Predicting Breast Cancer Using Different Deep Learning Algorithms and Explainable AI Techniques in an IoT Environment

Belgacem Bouallegue,[1,2] **Yasser M. Abd El-Latif**,[3] **Hosam El-Sofany**,[1] **and Islam A. T. F. Taj-Eddin**[4]

[1]*College of Computer Science, King Khalid University, Abha, Saudi Arabia*
[2]*Electronics and Micro-Electronics Laboratory, Faculty of Sciences of Monastir, University of Monastir, Monastir, Tunisia*
[3]*Faculty of Science, Ain Shams University, Cairo, Egypt*
[4]*Faculty of Computers and Information, Assiut University, Assiut, Egypt*

Correspondence should be addressed to Hosam El-Sofany; helsofany@kku.edu.sa

Received 20 October 2024; Revised 9 May 2025; Accepted 16 June 2025

Academic Editor: Eugenio Vocaturo

Copyright © 2025 Belgacem Bouallegue et al. International Journal of Intelligent Systems published by John Wiley & Sons Ltd. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Breast cancer is the primary cause of death for women around the world, necessitating the development of highly accurate, interpreted, and technologically advanced predictive approaches to support early diagnosis and treatment. In this research, we introduce a deep learning (DL) model for predicting breast cancer using both public and private datasets. The model uses the internet of things (IoT) to improve data collection and real-time monitoring, and it also uses the SMOTE method to resolve issues of class imbalance. The proposed model combines an explainable AI approach with SHAP values to ensure model interpretability. To identify the best DL algorithm for this method, we assess and compare six different DL algorithms: temporal convolutional networks (TCNs), neural factorization machines (NFMs), long short–term memory (LSTM) networks, recurrent neural networks (RNNs), gated recurrent units (GRUs), and deep kernel learning (DKL). IoT devices allow for the continuous acquisition of patient data, which, when integrated with our predictive models, improve the capacity for early detection. Reliable cancer detection relies on our method's enhanced predictive accuracy and sensitivity. Furthermore, we offer crucial transparency in clinical settings by using SHAP to give detailed explanations of model decisions. By employing thorough statistical analysis and cross-validation, we guarantee that our model is resilient and can be applied to various patient populations. The results show that our proposed IoT integrated method has the potential to improve prediction performance and boost confidence in AI-powered medical diagnostics by making them more accessible and easier to use. From a performance perspective, the proposed approach, which uses the TCN algorithm and SMOTE, achieved the best accuracy for BC prediction. With the public dataset, the experimental results were 99.44%, 100.0%, 99.01%, 98.75%, 99.37%, and 99.89% for accuracy, sensitivity, specificity, precision, F1-score, and AUC, respectively. The experimental results for accuracy, sensitivity, specificity, precision, F1-score, and AUC using the private dataset were 97.33%, 93.33%, 100%, 100%, 96.55%, and 99.48%, respectively. On the other hand, with the combined datasets, the TCN algorithm achieved 100% for all performance metrics.

Keywords: AI applications; breast cancer; deep learning; DL-based apps; IoT; SHAP; SMOTE

1. Introduction

Breast cancer (BC), the leading cause of cancer deaths in women, is a global public health issue. Better patient outcomes and lower mortality rates require early and accurate detection. This is especially true considering that over 2.3 million new cases are identified every year [1]. The complex interplay of risk factors and the disease's heterogeneity makes accurate prediction of BC challenging, even with advances in molecular diagnostics and imaging [2, 3]. Recent

improvements in medical testing have been achieved thanks to deep learning's better performance in areas such as image analysis, disease classification, and predicting outcomes [4, 5]. The use of deep learning (DL) on structured data, such as clinical datasets, is still in its infancy, though. Incorporating patient demographics, tumor characteristics, and treatment histories into these datasets can improve BC predictive models. Since benign cases are more common than malignant ones, addressing class imbalance is one of the biggest challenges to creating trustworthy prediction apps. The development of models that benefit the majority class could result from this imbalance, which could cause important cancer cases to go unrecognized [6]. We have made extensive use of the synthetic minority oversampling technique (SMOTE) method to address this issue. This method involves the generation of synthetic samples for the minority class, which helps to achieve a more balanced dataset and enhances the sensitivity of the model [7, 8]. To fully benefit from combining SMOTE with DL frameworks for BC prediction, more research is required. While DL models offer high predictive accuracy, their "black-box" nature has raised concerns regarding interpretability and trustworthiness, especially in sensitive fields such as healthcare [9]. Clear explanations of prediction-making are essential for clinicians to trust and adopt these models in practice. Many researchers are interested in explainable AI (XAI) techniques such as SHapley Additive exPlanations (SHAP) because they can show how different features affect model predictions. Such an approach makes complex models easier to understand [10, 11]. SHAP provides a unified framework for interpreting any DL model's predictions, making it an ideal choice for explaining DL models used in BC prediction [12].

Using both public and private datasets, we present an integrated model that combines DL algorithms, SMOTE, and SHAP to create a model for predicting BC that is both accurate and easy to understand. The main objectives are to make the model more transparent through explanation and to improve its performance by fixing the class imbalance. Through extensive experiments, we compare the proposed method's performance to industry standards and examine how SMOTE and SHAP affect model accuracy and interpretability.

The main goal of this study is to develop a cutting-edge DL approach for BC dataset classification and then compare its performance with other existing approaches. The following are a few of the most important findings from the study:

- One of the primary contributions is the use of private and BC Wisconsin (diagnostic) datasets (WBCDs) for BC prediction. Other contributions include utilizing the internet of things (IoT) to improve real-time data collection and monitoring, as well as multiple DL classifications.
- This research makes use of the SMOTE method to raise the dataset's total number of balanced occurrences. This solves a common BC prediction problem in which the number of benign tumors is much higher than the number of life-threatening ones. Therefore, to improve the accuracy of BC prediction, we test the suggested method on a balanced dataset with SMOTE.
- The research also looked at six other DL classification algorithms to determine which one produced the best rate of BC prediction overall. We found that the temporal convolutional network (TCN) classifier is the most accurate at predicting BC cases.
- To understand the prediction process of the BC model, we applied an XAI method based on SHAP libraries. This approach may help us figure out which features are essential for accurate and reliable prediction.
- In general, this study significantly contributes to the expanding field of BC prediction systems that utilize IoT-enhanced deep learning. Positive healthcare outcomes and reduced risk of chronic medical complications related to BC may result from the study's conclusions.

2. Literature Review

The research community is genuinely interested in the topic of improving diagnostic accuracy in BC prediction, and many studies have explored various DL and machine learning methodologies to achieve this goal. This section reviews the recent literature on BC prediction, focusing on the use of DL, data preprocessing methods such as SMOTE, and understandable AI approaches such as SHAP. DL has been increasingly implemented for BC prediction; it is powered by its remarkable capability to autonomously acquire feature sets from datasets without human oversight. In analyzing medical images, convolutional neural networks (CNNs) have shown great success, especially in classifying mammograms, ultrasounds, and histopathology images. For instance, Spanhol et al. demonstrated that CNNs could achieve high accuracy in classifying images of BC histopathology, performing better than regular ML methods [13]. Similarly, Ha et al. employed a deep CNN to predict BC risk using mammograms, leading to cutting-edge outcomes [14].

Besides CNNs, researchers have explored other DL models such as long short–term memory (LSTM) and autoencoder networks for BC prediction using feature extraction and time-series data, respectively [15, 16]. For example, Kourou et al. utilized LSTMs for predicting BC recurrence, highlighting the potential of DL in temporal data analysis [17]. Moreover, researchers have proposed ensemble DL models, combining multiple neural networks, to further enhance prediction accuracy by reducing model variance [18]. Regardless of the encouraging results, the application of DL to structured data, such as CSV-based datasets containing clinical information, is still relatively underexplored. Structured data, which include patient demographics, tumor characteristics, and treatment histories, provide critical information that can complement imaging data to improve prediction models [4, 5]. Studies have begun to explore the integration of DL with structured data, showing that these models can significantly improve predictive performance when appropriately designed [19]. Class

ijis, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/int/8884481 by Assiut University Code 71515, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

imbalance is a pervasive issue in BC datasets, where malignant cases are often underrepresented. This imbalance can cause biased models that exhibit high accuracy but poor sensitivity, particularly in detecting minority-class instances such as malignant tumors [6]. The SMOTE method has been widely adopted to solve this issue by creating synthetic samples for the minority class, thereby balancing the dataset and improving model sensitivity [20, 21]. Various BC prediction studies have successfully implemented SMOTE. For instance, Batista et al. demonstrated significant gains in sensitivity by using SMOTE to enhance the performance of support vector machines and decision trees on datasets unbalanced for BC [22]. Similarly, Rahman et al. applied SMOTE in conjunction with DL models, achieving better classification performance in detecting BC [8]. Recent advances in SMOTE, such as adaptive synthetic sampling (ADASYN) and borderline-SMOTE, aim to improve the quality of synthetic samples by concentrating on instances that are more challenging to classify [23, 24]. These techniques have been shown to outperform traditional SMOTE in various applications, including BC prediction, by reducing the overlap between classes and generating more realistic synthetic samples [25]. While DL models have achieved remarkable accuracy in BC prediction, their black-box nature poses significant challenges in clinical adoption. To make reliable and practical predictions, clinicians need models that are simple to understand and use [9, 12]. Emerging as a solution, XAI tools provide ways to understand complicated models. The capacity of SHAP to deduce the relative importance of each feature in making predictions has made it stand out among them [10, 11]. SHAP provides a unified approach to model interpretation, applicable to any machine learning model, including DL networks. It assigns a SHAP value to each feature, indicating its contribution to the model's output. This has been particularly useful in BC prediction, where understanding the effect of clinical variables on the model's predictions is crucial for validating its use in practice [26]. For example, Lundberg et al. used SHAP to explain the BC risk model predictions, highlighting the most influential factors contributing to each prediction [16].

Recent studies have explored the integration of the SHAP method with DL models in BC prediction. Xie et al. used SHAP to understand a DL model that was trained on mammography data. This showed how the model can tell the difference between benign and malignant growths [27]. Similarly, Yang et al. used SHAP to decipher the results of a CNN-based model applied to BC histopathology images. This exemplifies how SHAP can facilitate the comprehension of DL models used in medical imaging [28]. Using SHAP to predict BC not only makes the model clearer but also helps to find any possible flaws in it. This makes sure that the predictions are based on features that are clinically relevant and not just random correlations [29]. This is critical for gaining clinicians' trust and ensuring reliable AI model deployment in healthcare settings [30]. Recent advancements in BC prediction have focused on integrating multimodal data, combining imaging, genomic, and clinical data to develop more comprehensive models [31, 32]. Studies have shown that multimodal DL models can outperform single-modality models by capturing complementary information from different data sources [33].

Zhao et al. developed a multimodal DL model that combines mammography and clinical data, achieving superior performance in predicting BC risk [34]. An additional developing pattern is the application of transfer learning to refine models that have been pretrained on bigger datasets using smaller datasets related to BC [35, 36]. Research has demonstrated that this approach enhances model generalization and lessens the reliance on large, labeled datasets, which are frequently in short supply in medical applications [37].

The authors in [38] used the best DL models for image classification to segment and mask a tumor mass image. Few datasets are public, such as BUSI. Segmentation and classification models can mask breast ultrasound images, according to experiments. The study obtains the best model with about 90% accuracy. This study proved that DL algorithms can classify and segment ultrasound breast images, making them suitable for clinical trials.

In conclusion, the literature highlights the significant advancements in DL, SMOTE, and XAI for BC prediction. However, challenges remain in developing apps that are not only accurate but also interpretable and generalizable across different populations. In Table 1, we can see a summary of the relevant literature, which includes the datasets used, methodologies, paper references, and considerations/weaknesses.

3. Integration of IoT Technology for Enhanced Data Collection and Real-Time Monitoring

IoT integration into healthcare systems is a giant step forward in medical data collection, monitoring, and analysis. Specifically, in BC prediction, IoT devices serve as vital tools to enhance the depth, breadth, and immediacy of data available for DL frameworks. This section explores how IoT technology contributes to improved data collection, real-time monitoring, and overall predictive accuracy in BC diagnosis.

3.1. Role of IoT in Data Collection and Monitoring. Many different types of networked devices constitute the IoT. Some examples include smart medical equipment, home monitoring systems, mobile health applications, and wearable sensors. These devices are designed to continuously collect and transmit data related to a patient's health status, offering an unprecedented level of detail and continuity in medical data acquisition. For instance,

- Wearable sensors: Wearable fitness monitors and smartwatches can record vitals, including heart rate, temperature, and activity levels. More advanced sensors can track specific biomarkers related to BC risk, such as hormonal fluctuations or changes in breast tissue composition [39, 40].
- Smart medical devices: Connected diagnostic tools, such as smart mammography devices or portable

ijis, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/int/8884481 by Assiut University Code 71515, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1: The related works' summary.

Ref.	Dataset used	Methodology	Considerations/weaknesses
Spanhol et al. [13]	Breast cancer histopathology images (public dataset)	CNN for image classification	Focused only on image data; limited interpretability
Ha et al. [14]	Mammography images	Deep CNN for BC risk prediction	Lack of model interpretability and dataset variability
Kourou et al. [17]	Time-series data (clinical follow-up)	LSTM for BC recurrence prediction	LSTM is effective in temporal data but lacks integration with structured/static data
[18] Ensemble DL models	Various datasets	Combined neural networks to reduce variance	Higher computational cost; interpretability concerns
[4, 5, 19] Various (CSV-based)	Structured clinical datasets (e.g., demographics and tumor info)	DL models on tabular data	Underexplored domain; requires robust feature engineering
Batista et al. [22]	Imbalanced BC datasets	SMOTE + SVM/decision trees	May introduce synthetic noise; limited to traditional ML
Rahman et al. [8]	Clinical and feature-extracted data	SMOTE with DL models	Lacks interpretability; black-box nature remains
[24, 25] ADASYN and borderline-SMOTE studies	Imbalanced clinical datasets	Enhanced SMOTE variants	Complex to tune; risk of overfitting on borderline samples
Lundberg et al. [10, 11, 16, 26]	Clinical prediction models	SHAP for DL model interpretation	Improves model transparency and clinical acceptance
Lundberg et al. [27]	Mammography data	SHAP with CNN	Focus limited to imaging; not multimodal
Yang and Zhu [28]	Histopathology images	SHAP with CNN	Imaging only; lacks structured data integration
Zhao et al. [31–34]	Multimodal (clinical + imaging)	Multimodal DL model	Data integration complexity; interpretability trade-offs
[35–37] Transfer learning studies	Pretrained models on large sets + small BC datasets	Transfer learning techniques	May not generalize well across different demographics
Gagliardi et al. [38]	BUSI dataset	DL segmentation/classification	Limited public datasets; moderate accuracy
The proposed approach	Public, private, and combined datasets (CSV + IoT data)	6 DL classifiers, IoT-enhanced data collection + TCN with SMOTE + SHAP for explainability.	Using TCN, the accuracy was 99.89% for WBCD datasets, 99.48% for private datasets, and 100% for combined datasets

ijis, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/int/8884481 by Assiut University Code 71515, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ultrasound machines, can collect high-resolution images and other diagnostic data that are directly transmitted to healthcare providers or cloud-based systems for real-time analysis [41].

- Home monitoring systems: These systems enable patients to monitor their health from home, decreasing the frequency of hospitalizations. Data collected from home can include weight, glucose levels, blood pressure, and more, which are crucial for monitoring conditions that may increase BC risk [42].
- Mobile health applications: These apps can gather subjective data from patients, such as self-reported symptoms, lifestyle factors, and medication adherence. These data can complement the objective information obtained by other IoT devices to give a more comprehensive picture of the patient's health [43].

Since IoT devices collect data continuously and automatically, there is less chance of missing important changes in health that could mean BC is starting or spreading. This collection of real-time data is necessary to keep track of changing physiological and environmental factors that may affect how cancer grows and spreads.

3.2. Data Transmission and Storage. Once data are gathered from IoT devices, they are sent safely to a central data store. Wireless communication technologies such as Bluetooth, Wi-Fi, or 5G can help with this transmission, depending on the device's capabilities and the speed at which the data need to be sent. During transmission, encryption protocols must be used to protect the privacy and security of data. The huge amount of data that have been gathered will be kept in a central repository, which is most likely a cloud platform. This platform should be able to grow to handle the huge amounts of data that IoT devices produce. It should also have advanced data management tools to set up and prepare the data for analysis [44, 45].

3.3. Data Preprocessing and Integration. IoT data are often noisy and unstructured, requiring significant preprocessing before it can be used in the predictive model. The key steps in data preprocessing include [46, 47] the following:

- Data cleaning involves removing errors, inconsistencies, and duplication from the dataset to ensure its reliability and accuracy.
- Data normalization involves standardizing the data to ensure consistency in format and scale, a crucial step when integrating data from multiple IoT devices with varying output formats.
- Feature extraction: We train the DL models using features extracted from the raw IoT data, such as trends in the patient's vital signs or abnormalities from the imaging data.
- Data integration: To produce a comprehensive dataset, we integrate IoT data using standard clinical data, including genetic and medical history. We then train and validate the DL model using this combined dataset.

3.4. Model Training and Real-Time Analysis. The integrated and preprocessed dataset is sent to the DL framework. It uses different types of classifiers, including TCNs, neural factorization machines (NFMs), LSTM networks, recurrent neural networks (RNNs), gated recurrent units (GRUs), and deep kernel learning (DKL). The continuous, real-time data streams provided by IoT data can greatly improve model training, as they mirror the dynamic nature of a patient's health [48].

- Training with IoT data: The model can be trained on a hybrid dataset comprising both historical and real-time IoT data. The continuous influx of new data helps the model adapt and improve over time, leading to better predictive performance and reduced overfitting.
- Real-time prediction and monitoring: Real-time tracking and prediction of IoT data can be achieved with the trained model. For example, if the model detects patterns indicative of potential BC development, it can alert healthcare providers for further investigation or intervention.

3.5. Integrate the Frameworks of DL. Training DL models to predict BC depends heavily on the quantity and quality of data. The IoT makes these models much better by providing a steady stream of real-time data that can identify correlations and trends that might not be seen in older, more static datasets.

- Enhanced data diversity and volume: The variety of data collected through IoT devices, ranging from imaging data to biochemical markers and behavioral data, allows DL models to analyze a more comprehensive and diverse set of features. This diversity is crucial for creating robust models that can generalize across different patient populations [49].
- Real-time data integration: DL models can be continuously updated with new data from IoT devices, enabling them to adapt to changes in a patient's health status. This real-time integration allows for dynamic risk assessment and early intervention, as the models can detect emerging patterns that may indicate the onset of BC or changes in its progression [50].
- Personalized predictions: The detailed and continuous data from IoT devices support the development of personalized predictive models. By accounting for individual variability in lifestyle, genetics, and environmental factors, these models can offer tailored predictions that are more relevant to each patient's unique risk profile [44].
- Improved sensitivity and specificity: The real-time, high-frequency data provided by IoT devices enhances the sensitivity and specificity of DL models. This improvement is critical in BC prediction, where early detection can significantly impact treatment outcomes and patient survival rates [51].

ijis, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/int/8884481 by Assiut University Code 71515, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.6. Benefits of Using IoT in Clinical Settings. Clinical settings can greatly benefit from the real-time monitoring and data collection capabilities of IoT devices. This is particularly true when it comes to early detection and management of BC.

- Proactive patient management: With IoT-enabled continuous monitoring, healthcare providers can detect changes in a patient's condition as they happen, allowing for timely interventions. For instance, if an IoT device detects an anomaly in breast tissue composition, a clinician can be alerted immediately to conduct further tests or adjust the patient's treatment plan [52].
- Reduced hospital visits: Patients, especially those residing in underserved or rural regions, may discover that they require fewer visits to the hospital as a result of remote patient monitoring and IoT. In addition to making patients more comfortable and increasing their compliance, this also aids in the efficient management of healthcare resources [53].
- Integration with XAI: The use of XAI techniques, such as SHAP, alongside IoT data, provides transparency in model predictions. Clinicians can better understand the factors contributing to a model's decision, which is critical for validating the model's use in a clinical setting and for building trust with patients [53].

4. The Proposed BC Prediction Methodology

In this work, we accurately predicted BC using six different DL classification techniques. These included TCN, NFMs, LSTM Networks, RNNs, GRUs, and DKL. Hyperparameter values and other implementation-specific choices may be required, depending on the dataset and goals. There are certain limitations to each of the mentioned diagnostic tools when it comes to early detection and prediction. These approaches include genetic analysis, mammography, MRI, CT, blood tests, and positron emission tomography. DL has the potential to enhance BC prediction in many ways, including improving accuracy, automating feature extraction, and integrating diverse data sources. One major benefit is that it can automatically identify complicated patterns in medical images such as MRIs, ultrasounds, and mammograms, all without manual feature extraction. As a result, DL models can recognize subtle and intricate details that traditional techniques might overlook. Researchers have found that CNNs are better at finding BC than other machine learning methods [54]. They do this by finding patterns in imaging data that are linked to cancer at the microscopic level [55, 56]. In addition, DL can handle multimodal data, integrating imaging, genetic, clinical, and histopathological data into a unified model. More accurate risk assessments and customized treatment programs are the result of this comprehensive strategy, which improves prediction performance by including all aspects of the disease [57]. Also, DL models can adapt to new information and get better at making predictions over time since they can learn from huge, constantly expanding datasets. Real-world applications have already demonstrated the scalability of DL in processing large-scale data, outperforming human experts in certain diagnostic tasks [58]. By combining imaging with genomic and proteomic data, DL has the potential to advance precision medicine in BC, providing clinicians with better tools for early detection, risk stratification, and treatment decision-making [59]. Recent research, on the other hand, has been focused on how AI models can improve diagnosis. The studies demonstrated how AI can be used in breast medical ultrasound imaging to find lesions, group them, classify them, measure breast density, and estimate BC risk. They also discussed new AI ideas to predict BC risk and the pros and cons of AI in breast medical imaging [60].

However, the integration of IoT technology into BC prediction frameworks significantly enhances the data collection process, providing real-time, continuous monitoring that DL models can leverage for improved predictive accuracy. This combination not only bolsters the early detection capabilities of predictive models but also supports personalized healthcare by enabling dynamic, patient-specific monitoring, and intervention. Novel opportunities for enhancing patient outcomes in BC and other diseases may arise as a result of the increasing role that IoT plays in advancing AI-driven medical diagnostics.

We present here the methodology and DL algorithms used to develop the proposed DL technique for BC prediction. Figure 1 illustrates the proposed BC prediction and detection approach. Gathering and preprocessing datasets were the first steps in removing the relevant inconsistencies.

IoT devices, as shown in Figure 1, represent wearable devices and sensors that continuously collect patient data. To resolve the problem of having more benign cases than malignant ones in BC datasets, we used the SMOTE technique when preparing the data. SMOTE is a popular method that creates new examples for the smaller group by blending existing examples with their closest neighbors. This approach enhances the diversity of the minority class while preventing overfitting that can result from simple duplication. In our approach, we used SMOTE only on the training data after splitting it into 75% for training and 25% for testing, which helps the model learn from a more balanced dataset while keeping the test set's original class distribution intact. As a result, the training dataset size increased, improving the model's ability to detect minority class instances and boosting sensitivity. However, since the test dataset remained relatively small (approximately 200 samples), the change may have impacted the robustness of the reported performance metrics [20]. Future iterations may benefit from expanding the test set or adopting cross-validation strategies to ensure more generalizable and statistically reliable results. SMOTE and feature extraction are two examples of data preprocessing that address class imbalance. The six DL algorithms applied are called DL classifiers. They are TCNs, NFMs, LSTM networks, RNNs, GRUs, and DKL. The TCNs algorithm achieved the best performance results.

SHAP explainability: This component explains the model's decisions using feature importance and interpretable output. Prediction outcomes and metrics such as accuracy, sensitivity, precision, F1-score, and area under the curve (AUC) are produced by the model. The term "real-time feedback loop" refers to IoT devices that provide continuous data for real-time monitoring and possible

ijis, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/int/8884481 by Assiut University Code 71515, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

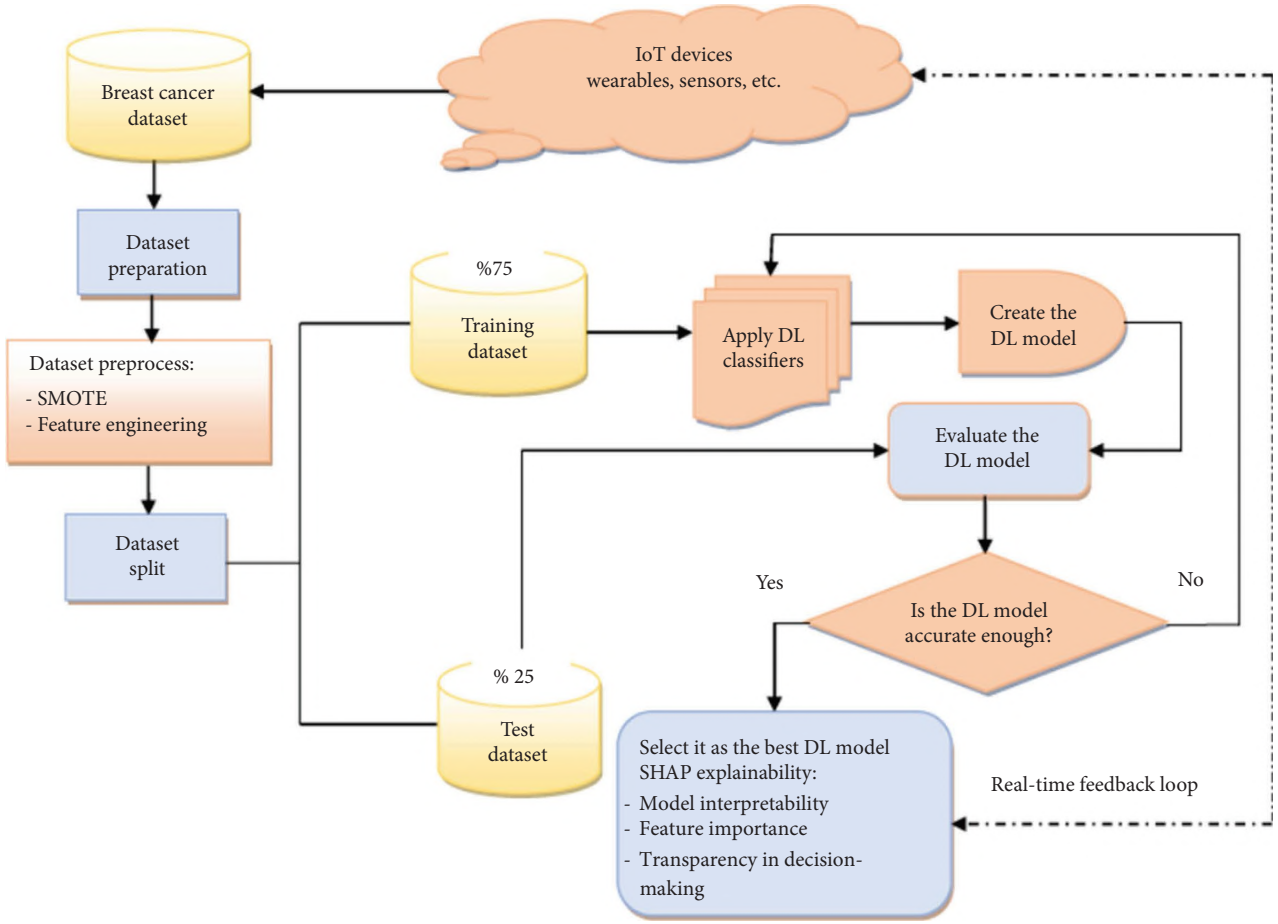

Figure 1: The proposed DL-based breast cancer prediction model.

model retraining. The dotted arrow represents the continuous feedback loop to further improve model performance in real-time.

A few of our enhancements include a more even distribution of class sizes and the substitution of average values for null instances. We created two datasets, one for testing and one for training. We next used several techniques to determine which classification algorithm was most suitable for this dataset. The designed application incorporates the best performance prediction model. The information gain method selects the highly significant features that are closer to the intended result. Equation (1) is the primary formula for determining the gain on the i^{th} feature f_i as compared to the class labels. Equation (2) determines the necessary information for classifying a tuple in dataset D, while equation (3) determines the expected and necessary information.

$$\text{Gain}(f) = \text{Info}(D) - \text{Info}_f(D), \tag{1}$$

$$\text{Info}(D) = -\sum_{i=1}^{n} P_i \log_2(P_i), \tag{2}$$

$$\text{Info}_f(D) = \sum_{i=1}^{n} \frac{|Dj|}{|D|} \times \text{Info}(Dj), \tag{3}$$

where $\text{info}_f(D)$ denotes the expected data needed for each feature f, info (D) is the expected data needed to categorize a record in D, and P_i denotes the probability that any record in D belongs to a BC class.

4.1. Datasets and Dataset Features. In this work, to evaluate the proposed approach for BC prediction, we used three datasets: the public WBCD dataset, a private dataset, and a combined dataset (which includes both public and private records). The Wisconsin Hospitals University in Madison created the Wisconsin dataset for BC classification [61]. We have used 719 samples from the WBCD dataset; the private dataset had 300 records, while the combined dataset had 1019 records. The private and WBCD datasets contain features generated by digitized breast mass fine needle aspirates (FNAs). These features enable the classification of breast lumps as benign or cancerous. The dataset contains the following features (attributes):

1. ID: Each sample's unique identification number.
2. Diagnosis: The class label indicates whether the mass is benign (B) or malignant (M).
3. Ten features: These real-number features, such as the average, standard error, and the worst values of various characteristics, are taken from the FNA

ijis, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/int/8884481 by Assiut University Code 71515, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

samples, which include perimeter, radius, smoothness, texture, area, concavity, compactness, fractal dimension, concave points, and symmetry.

In Figure 2, we can see the breakdown of the percentage of people with BC in both datasets. Figure 2(a) shows that 43.9% of the 719 records in the WBCD dataset have a diagnosis of BC, a malignant tumor, while the remaining 56.1% have benign tumors, where "B" refers to "benign" and "M" to "malignant." On the other hand, Figure 2(b) shows that 40.3% of the 300 records in the private dataset have a diagnosis of BC, while the remaining 59.7% have benign tumors.

The 300 patient records used for the private dataset are collected from a partner healthcare facility. These data were derived from digitized FNA tests, similar to those used in the public WBCD dataset, and include features such as texture_mean, radius_mean, perimeter_mean, smoothness_mean, and area_mean and their respective worst and standard error values. Each record was labeled as malignant (M) or benign (B) based on confirmed pathological reports. The class distribution in the private dataset is 40.3% malignant and 59.7% benign, as shown in Figure 2(b). Data were collected in compliance with ethical standards and anonymized to protect patient privacy.

In addition, IoT devices were used in the data acquisition phase to collect real-time patient vitals and tumor characteristics, enhancing temporal resolution and monitoring. The private dataset underwent preprocessing, including duplicate removal, null value handling via mean imputation, and standardization using the StandardScaler. We applied SMOTE to address class imbalances before feeding the data to our DL classifiers. This preprocessing ensured that the private dataset was comparable in structure and quality to the public WBCD dataset.

4.2. Preparing and Processing the Dataset. This study used both the open-source WBCD and some private datasets. You can use the holdout validation method by splitting the datasets into parts. There are 75% of the data in the training dataset and 25% in the test dataset. This study used the mutual information (MI) method to find out how the variables were related to each other. Many numbers mean that there is a lot of information and a lot of dependence. Looking at the importance of each feature in a dataset helps you figure out how useful and accurate they are. As we can see in Figure 3, this method gives the concave points_mean feature the most weight (29.53%), while the fractal_dimension_mean feature gets the least weight (0.01%). The minimum, maximum, and average values for the top 10 attributes from the WBCD and private datasets are shown in Table 2.

Preprocessing data includes removing duplicate values. We use a data balancer to perform data-balancing tasks. Thereafter, we use our approach to extract its features. We pass the data in its entirety to the data split process. In this case, we use a concrete method to partition the dataset into two parts: the training set and the testing set. We evaluated how accurate our BC predictions were by using DL classifiers such as NFMs, LSTM networks, RNNs, GRUs, and DKL. The StandardScaler module in this study resizes features and processes raw BC data. Many DL developers see data standardization as an absolute must. It is a usual practice to use feature selection as a preprocessing step before learning.

4.3. Advantages of Using DL Classifiers. As explained previously, this research employed six distinct DL algorithms for BC prediction. We used the GridSearchCV framework to determine the optimal values for each DL classifier's hyperparameters to prevent overfitting. DL classifiers have several advantages in BC prediction and detection, including the following:

- High accuracy and performance: DL models, especially CNNs, can extract complex features and patterns from medical images (such as mammograms, ultrasounds, or MRI scans) that may not be easily discernible to the human eye. These models often achieve higher accuracy compared to traditional ML methods, as they can learn directly from raw data without requiring handcrafted features.
- Automation and scalability: DL classifiers enable automated analysis of medical images and pathology slides, reducing the need for manual interpretation by radiologists and pathologists. This helps scale up the screening process, especially in regions where trained professionals may be scarce.
- Feature learning from raw data: Different from more conventional approaches, DL models may automatically learn hierarchical features from unprocessed data such as clinical records, gene expression data, or raw pictures. This enables more robust and comprehensive data analysis without extensive manual intervention.
- Improved sensitivity and specificity: DL classifiers typically show better accuracy (regarding both false negative [FN] and true positive [TP] rates) when it comes to detecting cancer. This reduces both FNs (missed cancer cases) and false positives (FPs) (incorrect cancer diagnoses), improving patient outcomes by facilitating earlier and more accurate diagnoses.
- Handling of large and complex datasets: BC diagnosis involves analyzing complex datasets, such as multimodal data (imaging, genetics, and clinical data), and DL excels at handling such large and varied data inputs. Training DL models on large datasets does not compromise their ability to generalize to novel, unseen data.
- Support for multiple imaging modalities: DL classifiers can analyze data from various imaging modalities, including mammography, MRI, ultrasound, and histopathological images. Integrating information from multiple sources helps provide a more comprehensive and accurate diagnosis.
- Personalized treatment insights: DL models can help predict treatment outcomes by analyzing gene expression patterns and other biological markers related to BC. As a result, treatment regimens can be fine-tuned to each patient's cancer subtype.

ijis, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/int/8884481 by Assiut University Code 71515, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

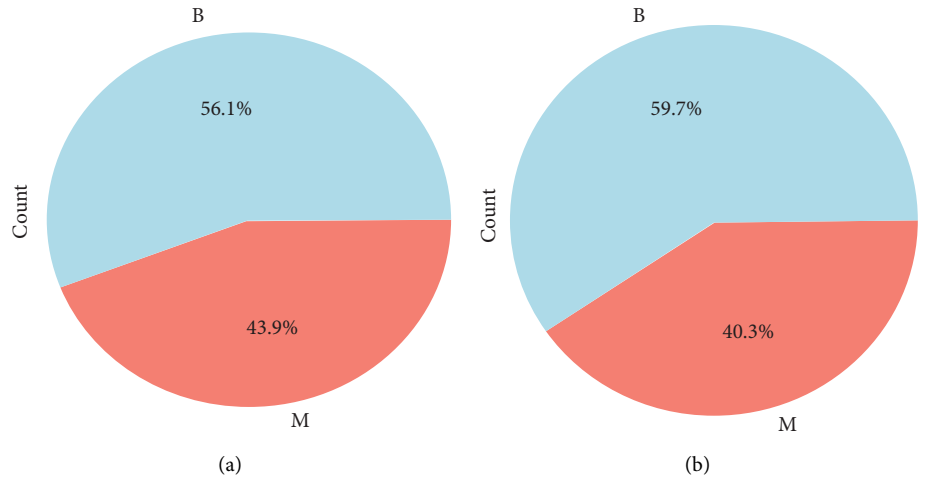

Figure 2: The percentage distribution of breast cancer among participants. (a) WBCD dataset. (b) Private dataset.

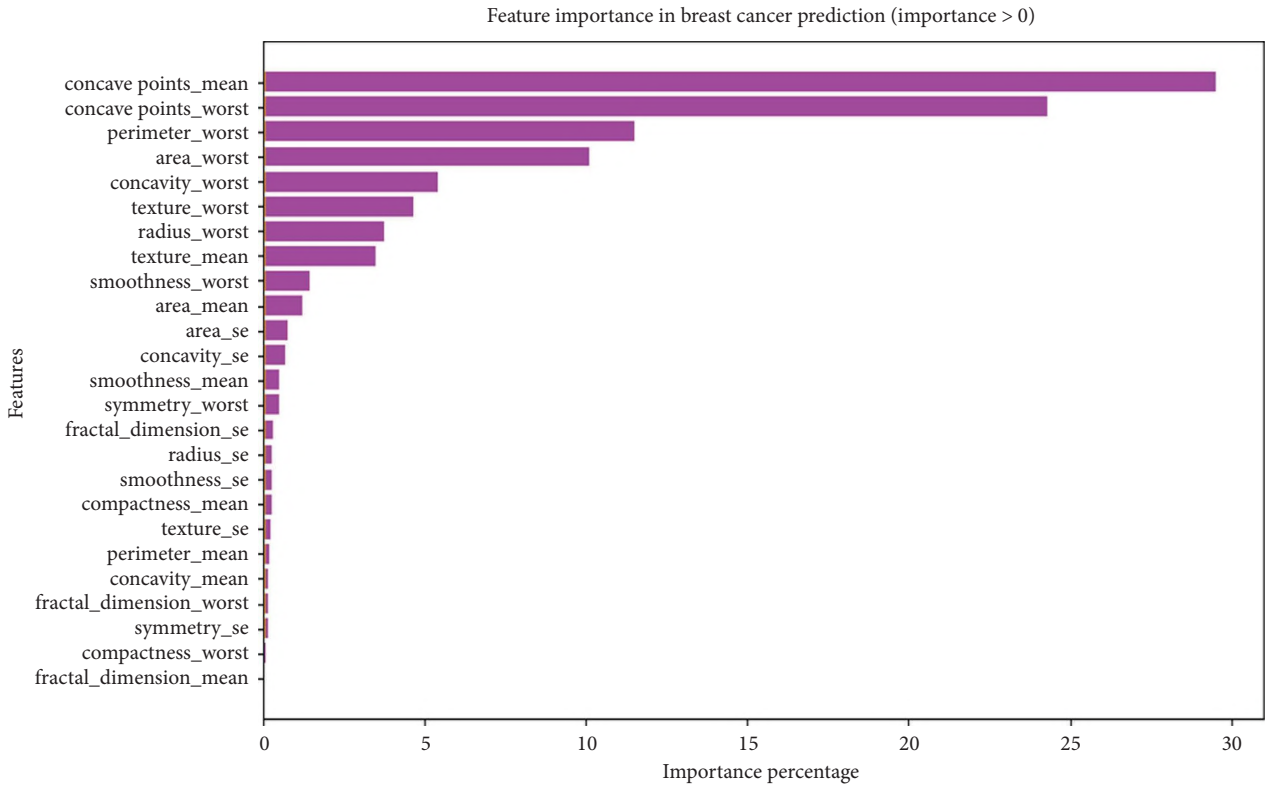

Figure 3: The evaluation of feature importance for the breast cancer dataset.

- Reduction in diagnostic time: DL classifiers can process and analyze medical images in seconds, vastly reducing the time required for diagnosis and allowing clinicians to focus on treatment decisions. Faster detection can lead to earlier intervention and better survival rates.
- Continuous improvement: DL models can continue to improve over time through transfer learning and fine-tuning. As we collect more data and update the models, their performance consistently improves, enhancing their reliability over time.
- Assistance in difficult or ambiguous cases: In cases where human experts may find it challenging to differentiate between benign and malignant lesions, DL classifiers can provide an additional layer of confidence by identifying subtle patterns that may go unnoticed in traditional manual review.
- Reduction in interobserver variability: In BC detection, variable diagnoses between different radiologists or pathologists are common. DL models provide a consistent and reproducible outcome, reducing human-induced bias or inconsistency in diagnosis.

ijis, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/int/8884481 by Assiut University Code 71515, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2: The top 10 WBCD and private dataset features' min, max, and average values.

1st ten important features	WBCD dataset			Private dataset		
	Max	Min	Average	Max	Min	Average
Concave points_mean	0.20	0.0	0.05	0.1913	0.0	0.05
Concave points_worst	0.29	0.0	0.12	0.29	0.0	0.12
Perimeter_worst	251.20	50.41	108.39	251.2	50.41	109.99
Area_worst	4254.0	185.2	890.18	4254.0	185.2	915.37
Concavity_worst	1.252	0.0	0.29	1.25	0.0	0.29
Texture_worst	42.79	12.49	25.56	41.61	12.49	25.55
Radius_worst	36.04	7.93	16.41	36.04	7.93	16.64
Texture_mean	30.72	10.38	19.10	30.72	10.38	19.08
Smoothness_worst	0.22	0.084	0.13	0.22	0.09	0.13
Area_mean	2501.0	143.5	655.98	2501.0	143.5	671.02

5. Findings and Discussion

This section introduces the findings of the proposed BC prediction system and describes the current implementation. We assessed the accuracy of six different DL classifiers in the first step of the process. By comparing several DL models, we were able to determine how accurate the predictions were. We used several metrics, such as the F1-score, AUC, accuracy, sensitivity, and precision, so that DL methods can be tested and assessed. The proposed system's accuracy was assessed using these numbers, which constitute performance metrics. True negative (TN), FN, TP, and FP are all defined by these types of metrics, respectively. T_P signifies that the system correctly predicts and receives a positive value. F_P suggests that the model predicts positively but ends negatively. T_N signifies that the system predicts incorrectly, and the outcome verifies this. F_N signifies that the model predicts a negative result but gets a positive one. We used a holdout validation method with a 75:25 train–test split to validate all DL classifiers.

Using hyperparameters, we further refined the DL classifiers utilized in this study. Every DL classifier has had its hyperparameters fine-tuned. The proposed approach obtained an accuracy rate of 99.44%, 97.33%, and 100% for using the WBCD, private, and combined datasets, respectively, with optimized hyperparameters.

We can see the expected numbers of TP, FP, TN, and FN in Figure 4 when the TCNs classifier is trained with the SMOTE method. For each entry in this confusion matrix, we can see how often a certain set of labels appears in the actual and predicted classes. There are a total of 100 different tumor types (TN) in the WBCD dataset (see Figure 4(a)). The experimental results reveal that 79 of these tumors are malignant, 1 is benign, and 0 are tumors that were initially thought to be benign but are malignant. Out of the 45 TN instances, 28 were found to be malignant, 0 to be benign, and 2 to be malignant that were mistakenly classified as benign, according to the results using the private dataset (see Figure 4(b)). "M" denotes malignant, and "B" indicates benign.

Figure 5 illustrates the relationship among the top 10 characteristics. The y-axis values for the following concave points are texture_worst, area_worst, texture_mean, radius_worst, smoothness_worst, perimeter_worst, and concavity_worst. Correlation coefficients that are either positive or negative indicate a strong connection between the two variables, while values of −1 or 1 imply no correlation at all. It is important to note that correlation only reveals linear relationships between variables. Each of these characteristics demonstrates a significant influence on the model's decision-making process, with at least a 96% correlation between the patient's prediction and each feature.

Figures 6 and 7 display the density and scatter plots for the combined dataset's first four important features. These diagrams help clarify the distributions and correlations of the variables within the BC dataset. By examining concave points_worst, concave points_mean, perimeter_worst, and area_worst, we can identify the potential patterns, outliers, correlations, and concentrations. In this context, a value of 1 denotes malignant (M), while 0 indicates benign (B).

We evaluate the system's performance by analyzing the results of different types of experiments. For the study's overall results to be accurate and to support the experimental findings, we need to test and evaluate each stage using the chosen classifiers. It is also essential to identify whether the classifier is capable of distinguishing between the various categories of features. To ascertain the correct functioning of the system, we evaluate the suggested model's performance using several metrics, such as accuracy, sensitivity, precision, specificity, AUC, and F1-score. Such metrics are based on the T_P, F_P, T_N, and F_N. We used a confusion matrix to figure out the system's performance.

- The following formula establishes the accuracy of the proposed approach by determining the percentage of correctly classified samples (equation (1)) as:

$$\text{accuracy} = \frac{T_P + T_N}{T_N + T_P + F_N + F_P}. \tag{4}$$

- Sensitivity is a metric that determines how complete the classifier's findings are. The formula that follows defines sensitivity as "the percentage of successfully identified positive samples" (equation (2)):

$$\text{sensitivity} = \frac{T_P}{T_P + F_N}. \tag{5}$$

- Specificity is a metric that measures the ability of a classifier to correctly identify negative samples. The metric provides an answer to the question, "What is

ijis, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/int/8884481 by Assiut University Code 71515, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

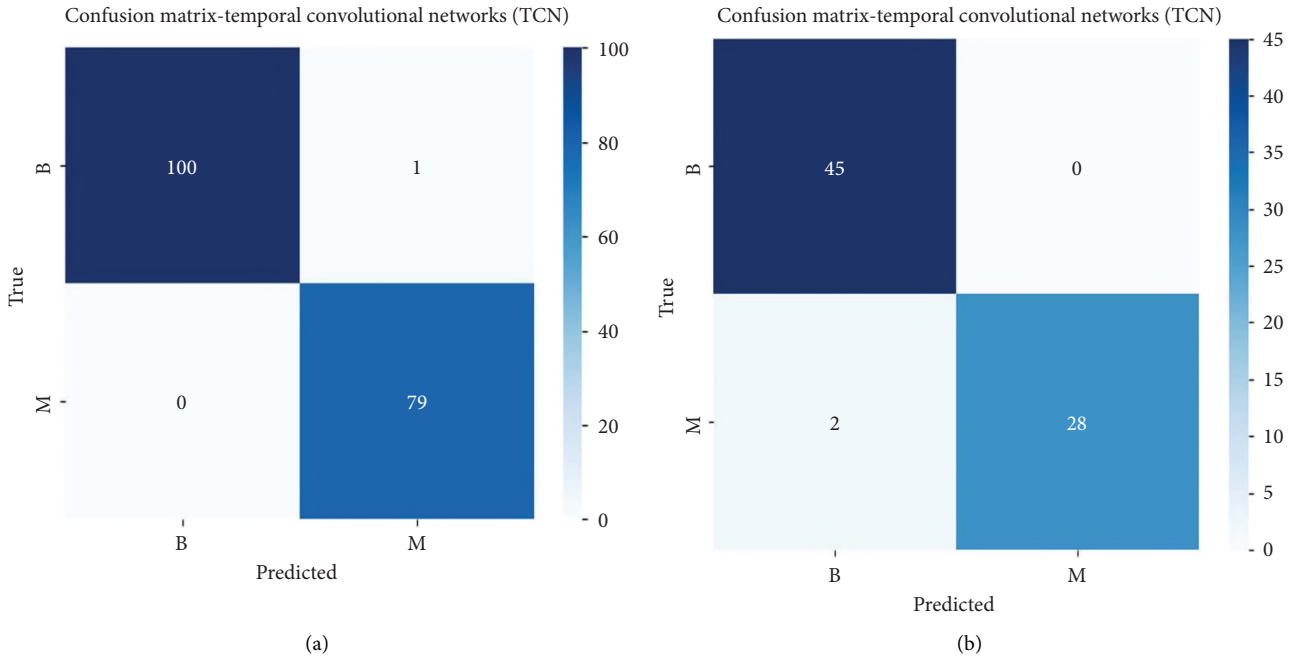

FIGURE 4: A breast cancer confusion matrix employing temporal convolutional networks and SMOTE. (a) Using the WBCD dataset. (b) Using a private dataset.

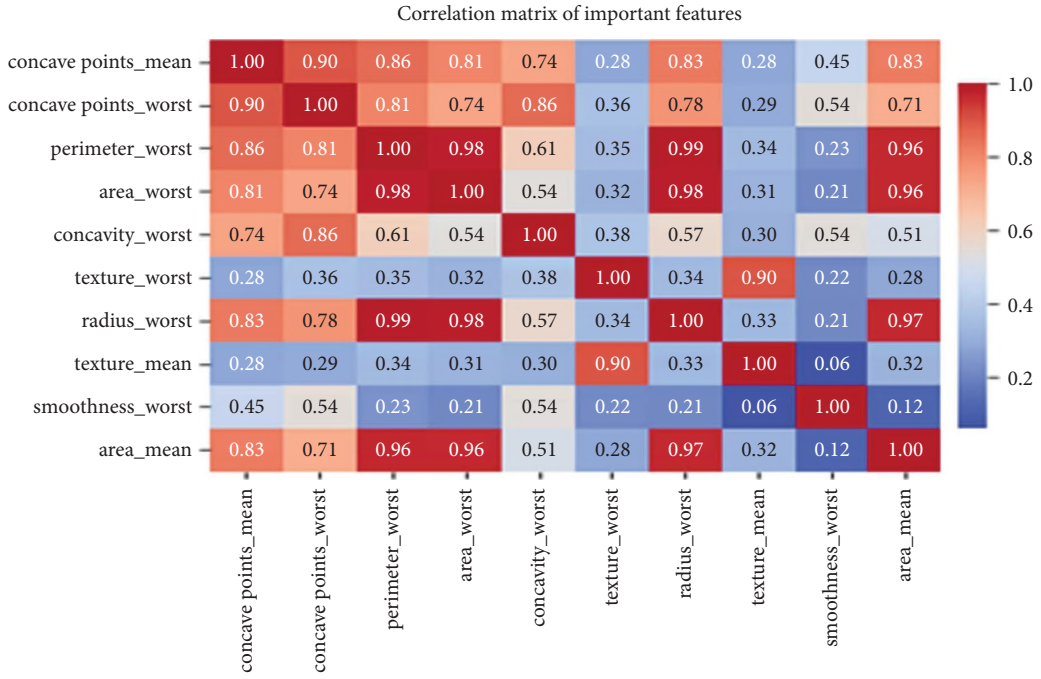

FIGURE 5: The correlation matrix between the first ten important features.

the correct classification percentage of actual negative samples?" It is defined mathematically by the following formula (equation (3)):

$$\text{specificity} = \frac{T_N}{T_N + F_P}. \tag{6}$$

- We use precision to determine the accuracy of the classifier result, and we can derive its value based on the provided information. By comparing the actual T_P to the expected T_P in equation (4), we demonstrate how the precision measure establishes the reliability of the proposed method's behavior.

ijis, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/int/8884481 by Assiut University Code 71515, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

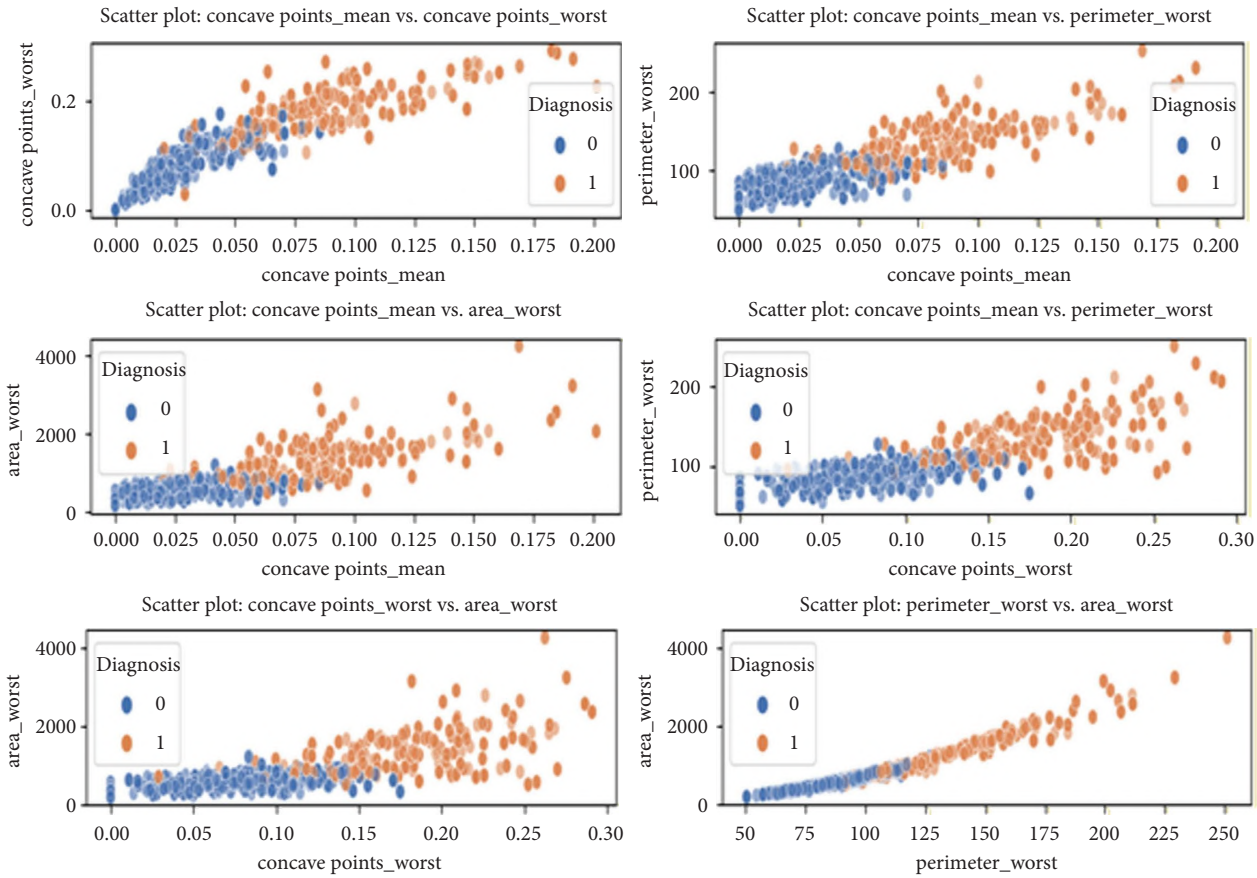

Figure 6: The scatter graph for the first 4 features.

$$\text{Precision} = \frac{T_P}{T_P + F_P}. \tag{7}$$

- The F1-score is a metric that balances sensitivity and precision in a single value, providing a harmonic mean of the two. It is particularly useful for evaluating classifiers in scenarios with imbalanced datasets, where one class significantly outweighs the other (see equation (5)).

$$\text{F1} - \text{score} = 2 * \frac{(\text{precision} * \text{sensitivity})}{\text{precision} + \text{sensitivity}}. \tag{8}$$

Table 3 compares the performance metrics of six different DL classifiers using a public dataset (WBCD) with SMOTE. The results demonstrated that the TCN algorithm's prediction obtained the highest overall performance, showing a rate of accuracy of 99.44%. The NFM classifier obtained the second-highest accuracy rate of 98.89%. On the other hand, the DKL algorithm achieved a minimum accuracy of 94.44%, as Figure 8 illustrates.

The findings in Table 4 show that the method using the TCN classifier performed very well on the private dataset that was improved with SMOTE. The accuracy, sensitivity, specificity, precision, F1-score, and AUC scores were 97.33%, 93.33%, 100%, 96.55%, and 99.48%, respectively.

Figure 9 illustrates the ROC curve for the TCN classifier. With an AUC of 99.89%, this curve displays the model's performance across all classification criteria using the WBCD dataset, while it achieved an AUC of 99.48% using the private dataset.

Afterward, we trained the DL-based model and evaluated it on both datasets using a domain adaptation strategy. First, we used a public dataset (WBCD) with 719 cases to train the BC prediction method; thereafter, we used a private dataset with 300 cases, and finally, we tested the model with a merged dataset consisting of 1019 cases from both WBCD and private datasets.

To understand and figure out how the model's findings make sense, we used an XAI approach that makes use of SHAP libraries. Figure 10 illustrates the crucial role of the SHAP library and how the SMOTE-based TCN classifier uses it.

The goals of this research are to use DL classifiers to make a model that can accurately recognize BC, fix class imbalance with the SMOTE method, improve data collection and real-time monitoring with IoT technology, and make the model's explanation easier with an AI approach based on SHAP values. According to the results of the experiment, the classifier that used the TCN algorithm had the highest accuracy rate in predicting the occurrence of BC. The following features are considered essential to BC

ijis, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/int/8884481 by Assiut University Code 71515, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

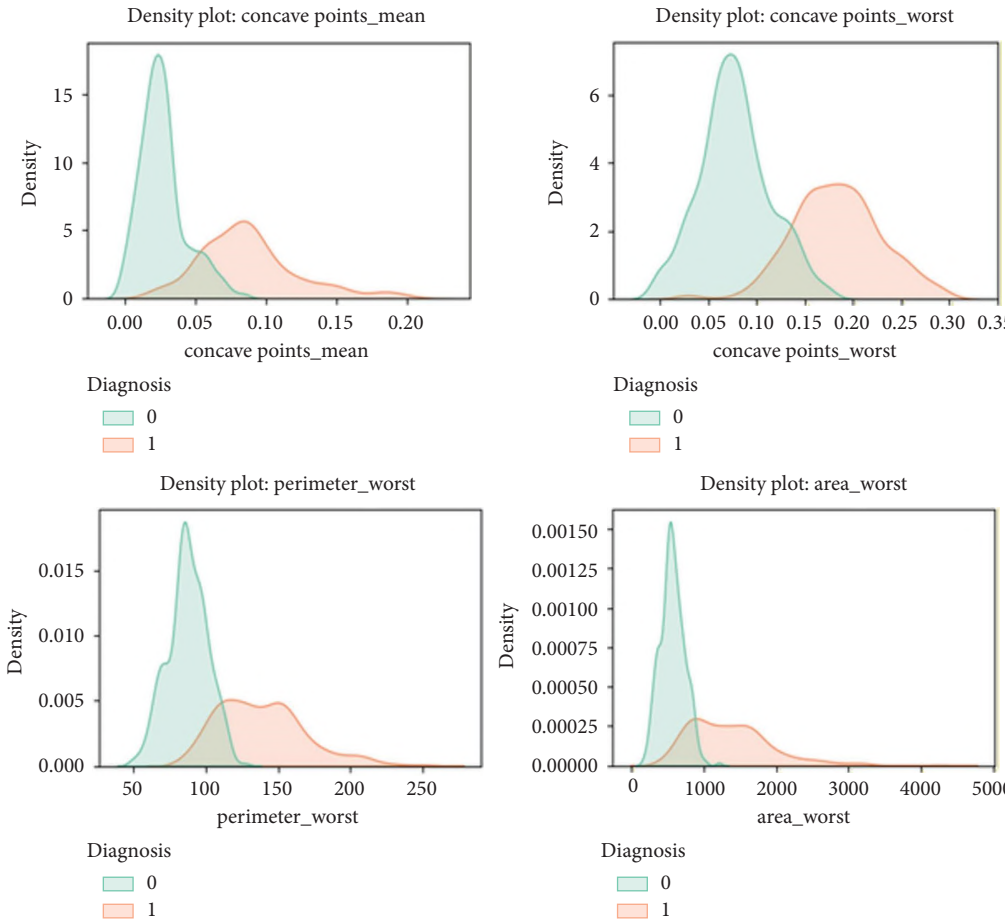

Figure 7: The density graph for the first 4 features.

Table 3: Performance outcomes for six DL classifiers using the public dataset (WBCD) SMOTE.

Ser.	Classifier	Accuracy	Sensitivity	Specificity	Precision	F1-score	AUC
1	Temporal convolutional networks	99.44	100	99.01	98.75	99.37	99.89
2	Neural factorization machines	98.89	98.73	99.01	98.73	98.73	99.95
3	Recurrent neural networks	97.22	96.20	9802	97.44	96.82	99.87
4	Long short–term memory networks	97.22	0.9620	0.9802	0.9744	0.9682	0.9987
5	Gated recurrent units	97.22	96.20	98.02	97.44	96.82	99.89
6	Deep kernel learning	94.44	90.41	97.20	95.65	92.96	98.32

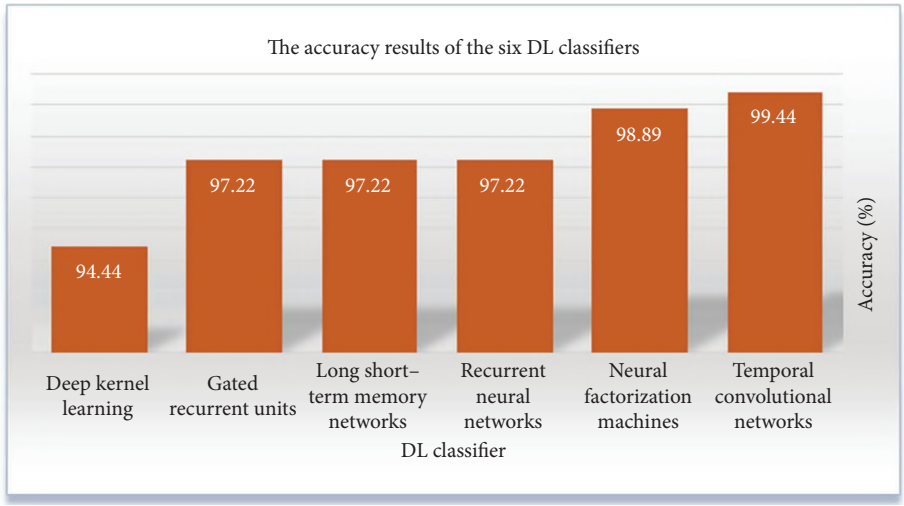

Figure 8: The accuracy results of the six DL classifiers.

ijis, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/int/8884481 by Assiut University Code 71515, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 4: SMOTE-based private dataset performance evaluation of the proposed approach.

Classifier	Accuracy	Sensitivity	Specificity	Precision	F1-score	AUC
Temporal convolutional networks	97.33	93.33	100	100	96.55	99.48

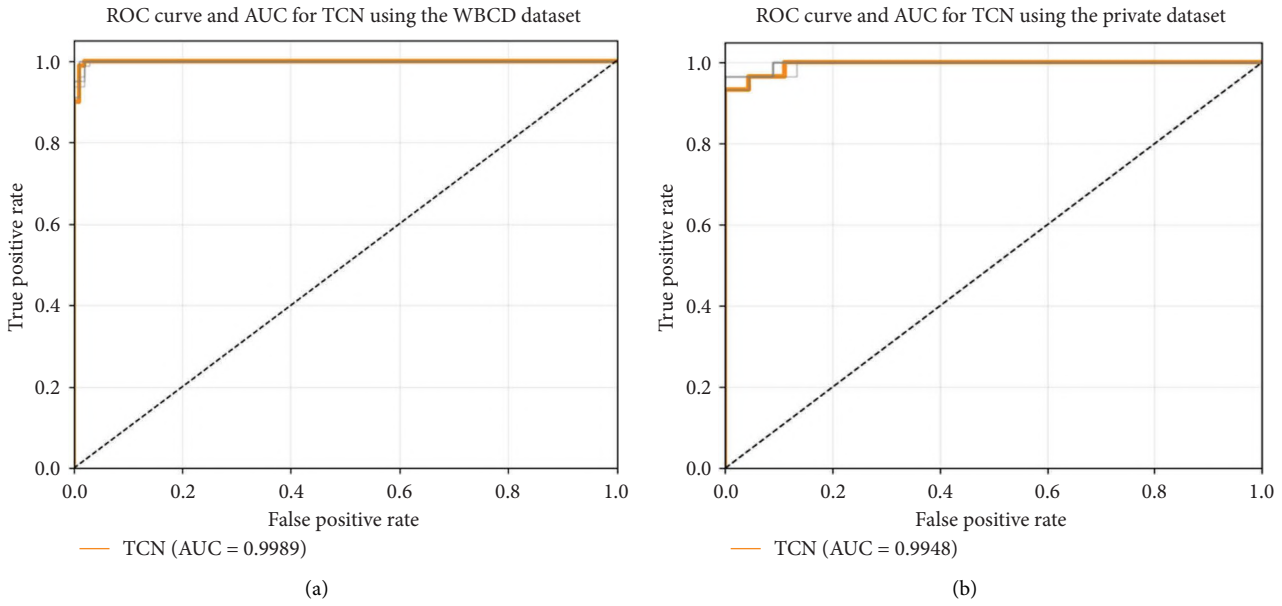

Figure 9: AUC and ROC curves for the temporal convolutional networks classifier.

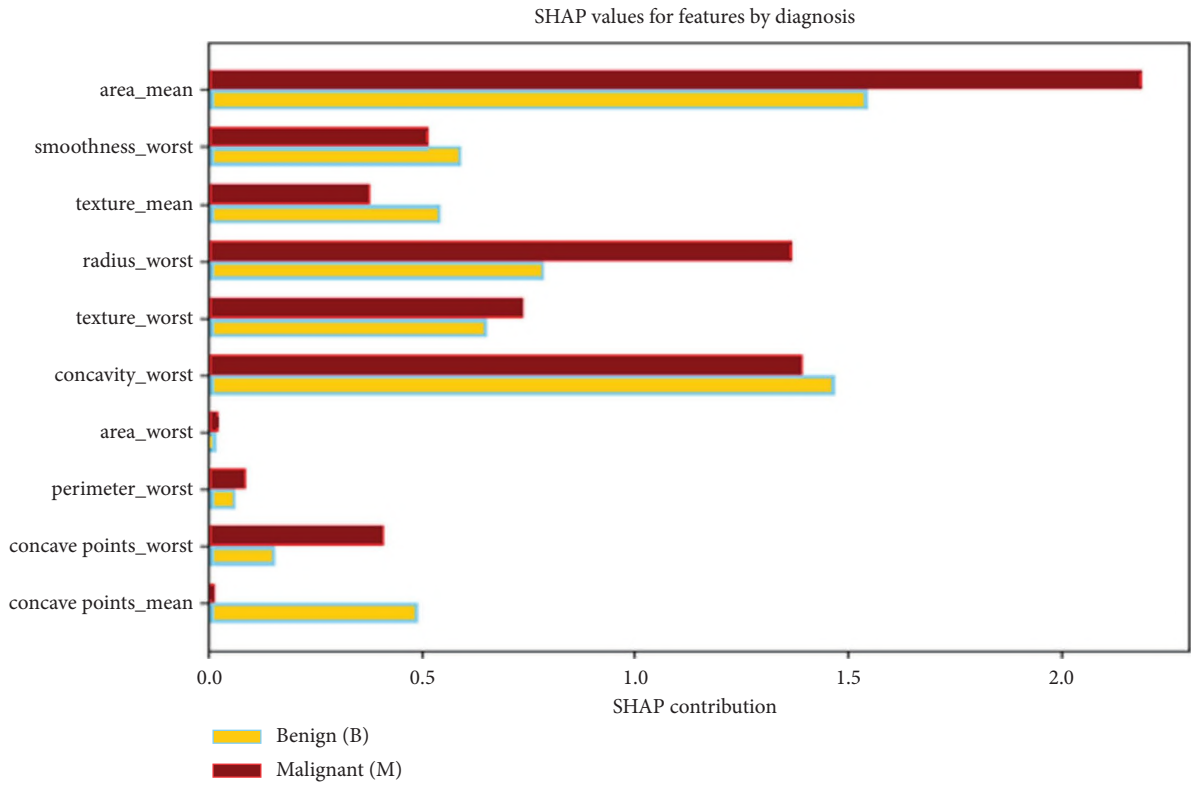

Figure 10: Feature importance (SHAP values), using the temporal convolutional network classifier.

ijis, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/int/8884481 by Assiut University Code 71515, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 5: Comparing performance on the WBCD dataset.

Study	Method	Accuracy	Sensitivity	Specificity	AUC	Description
Spanhol et al. [13]	CNN (histopathology)	85%–90%	Not reported	Not reported	N/A	Image-based, not structured data
Batista et al. [22]	SMOTE + SVM/DT	94%	92%	94%	95%	Limited to traditional ML
Rahman et al. [8]	SMOTE + deep learning	95%	93%	95%	96%	No explainability component
The proposed approach	TCN + SMOTE + SHAP	99.44%	100%	99.01%	99.89%	IoT-enabled; interpretable via SHAP

prediction according to the MI-based feature selection method: concave points_worst, perimeter_worst, concave points_mean, area_worst, concavity_worst, radius_worst, texture_worst, texture_mean, smoothness_worst, and area_mean. Using the collected data, we optimized hyperparameters, and using the SMOTE method, we optimized oversampling. The TCN classifier and the SMOTE method achieved the best performance for BC prediction. The evaluations performed with the public dataset gave 99.44% accuracy, 100.0% sensitivity, 99.01% specificity, 98.75% F1-score, 99.37% precision, and 99.89% AUC. Using the private dataset, the evaluation showed that the accuracy, sensitivity, specificity, precision, F1-score, and AUC were 97.33%, 93.33%, 100%, 100%, 96.55%, and 99.48%, in that order. However, when combining the datasets, the TCN algorithm achieved 100% for all performance metrics.

To contextualize the performance of our proposed model, we compared the results obtained using the WBCD public dataset with those reported in the existing literature. Our TCN model, integrated with SMOTE and XAI using SHAP, achieved an accuracy of 99.44%, sensitivity of 100%, specificity of 99.01%, precision of 98.75%, F1-score of 99.37%, and an AUC of 99.89%. These results significantly outperform previous studies using the same dataset. For instance, Batista et al. [22] applied SMOTE with support vector machines and decision trees, getting about 94% accuracy, while Rahman et al. [8] used SMOTE with a DL model, achieving around 95% accuracy but with less clarity on the results. Similarly, Spanhol et al. [13] used CNN models on histopathology images and found accuracies between 85% and 90%, but their research concentrated on image data instead of organized clinical information (see Table 5). These comparisons highlight how our approach is not only better at making predictions but also easier to understand and can be used in real time with IoT integration.

To ensure the reproducibility of our proposed method, we present the detailed hyperparameter configurations used for training each DL classifier. The TCN model, which achieved the highest performance, was trained using four convolutional layers with filter sizes of 32, 64, 128, and 256; a kernel size of 3; a dropout rate of 0.3 to prevent overfitting; and the ReLU activation function. The Adam optimizer was employed with a learning rate of 0.001 and a batch size of 32, and the model was trained for 100 epochs. For the NFMs, the embedding size was set to 10, with two hidden layers of 64 and 32 neurons, respectively, and a dropout rate of 0.2. The RNN, LSTM, and GRU models were all configured with two layers of 64 hidden units each, a dropout rate of 0.3, and were optimized using Adam with a learning rate of 0.001 over 100 epochs. The DKL model used a radial basis function (RBF) kernel, with two hidden layers (128 and 64 units) and was trained using the same optimizer and learning rate schedule. All models were implemented using scikit-learn and TensorFlow 2.0 in Python 3.10, and training was conducted on a workstation equipped with an NVIDIA RTX 3080 GPU and 32 GB of RAM. These consistent and transparent settings ensure that our results can be reliably replicated in future studies.

Finally, while our study compares the performance of six DL classifiers, we also benchmark our proposed method against the state-of-the-art (SOTA) models from recent literature using the same public dataset (WBCD) or similar BC diagnostic datasets. For example, Rahman et al. [8] integrated SMOTE with DL and achieved 95% accuracy and an AUC of 96%, but the model lacked interpretability. Batista et al. [22] used SMOTE with SVM and decision trees, reaching 94% accuracy, while Spanhol et al. [13] applied CNNs to histopathological images, achieving 90% accuracy. In comparison, our TCN + SMOTE + SHAP model reached an accuracy of 99.44%, a sensitivity of 100%, and an AUC of 99.89%, while also being easy to understand thanks to SHAP (see Table 6). These findings demonstrate superior performance and highlight the novelty of our method, which integrates IoT-based real-time data acquisition and XAI that are typically absent in prior SOTA methods. This positions our work as an advancement over current SOTA solutions in terms of both performance and practical clinical usability.

6. Limitations

The proposed IoT-enhanced DL method for predicting BC faces several challenges, including the following:

- Quality and availability of datasets: The success of the DL model is contingent on the amount and quality of the training and testing datasets. Limitations in availability and representativeness, as seen with the WBCD and private datasets, may restrict wider application.
- Class imbalance: Tackling class imbalance is essential, especially when the disease is infrequent. To address this issue, SMOTE creates synthetic samples, but we can also explore other strategies.
  - Class weights: They are easy to implement, but they depend on having accurate misclassification cost knowledge.
  - Ensemble methods: They are effective for data that are imbalanced and resistant to noise, but they require greater computational resources.

ijis, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/int/8884481 by Assiut University Cade 71515, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 6: Some comparisons with state-of-the-art methods.

Study	Method	Accuracy	AUC	Interpretability	IoT integration	Dataset type
Spanhol et al. [13]	CNN (image-based)	90%	N/A	×	×	Histology
Batista et al. [22]	SMOTE + SVM/DT	94%	95%	×	×	WBCD
Rahman et al. [8]	SMOTE + DL	95%	96%	×	×	WBCD
The proposed study	TCN + SMOTE + SHAP + IoT	99.44%	99.89%	✓	✓	WBCD

  - Cost-sensitive learning: It allows for fine-tuning based on misclassification costs, but it also adds complexity to the model selection process.
  - Anomaly detection: It is useful for identifying rare occurrences, but it may misidentify subtle patterns.
- Choice of algorithms: The selection of DL algorithms can significantly impact results. Future studies should assess a wider variety of algorithms to enhance predictive accuracy.
- Missing data: The study does not clarify whether the datasets have missing values, which are common in healthcare and can negatively influence predictions. It is important to address missing data using imputation or robust algorithms to guarantee the clinical prediction models' reliability and dependability.

7. Conclusions and Future Work

Using merged datasets of public and private datasets, this research introduced a DL model for BC prediction. By integrating IoT, the model enhances both data collection and real-time monitoring. To address the class imbalance, it employed the SMOTE. Our approach used an XAI component that incorporates SHAP to ensure clarity in model interpretations. We assessed six different DL algorithms to identify the most effective one for this task: TCN, NFM, RNNs, LSTM networks, GRUs, and DKL. IoT devices enable continuous data collection, which significantly enhances our predictive models' early detection capabilities. The findings reveal notable enhancements in both predictive accuracy and sensitivity, which are crucial for dependable cancer detection. In addition, SHAP provides insightful information on the model's decision-making procedure, promoting transparency essential for clinical environments. Rigorous cross-validation and statistical analyses validate the models, ensuring their robustness and applicability across various patient demographics. Our IoT-enhanced approach shows promising potential for improving predictive performance and fostering confidence in AI-assisted medical diagnostics through greater interpretability. The combination of the TCN algorithm and SMOTE produced the highest accuracy in predicting BC. The proposed approach outperformed the competition in predicting BC using the TCN algorithm in conjunction with SMOTE, according to performance metrics. Using the public dataset, the accuracy, sensitivity, specificity, precision, F1-score, and AUC were, respectively, 99.44%, 100.0%, 99.01%, 98.75%, and 99.37%. Using the private dataset, we achieved experimental results of 97.33% for accuracy, 93.33% for sensitivity, 100% for specificity, 96.55% for precision, and 99.48% for AUC. However, when the combined datasets were considered, the TCN algorithm reached a perfect score of 100% across all performance metrics.

Future research directions include the following:

- Expansion to multimodal data sources: Along with physiological data acquired through the IoT, future research may combine various data types, including genomic information, radiological images, and pathology slides. The accuracy and reliability of the model may be improved when applied to various diagnostic modalities using this multimodal strategy.
- Deployment in real-world clinical settings: Deploying the model in hospital settings could further validate its effectiveness. Its usefulness, scalability, and compatibility with electronic health record systems could be evaluated through real-time testing with real patients.
- Personalized treatment recommendations: The incorporation of patient-specific factors, such as lifestyle and genetic markers, has the potential to transform future models from simple diagnostic tools into systems that can also suggest personalized treatment plans. The result would be a significant advancement in precision medicine.
- Federated learning for privacy-preserving AI: Future iterations could explore federated learning to address data security and privacy concerns, especially when dealing with personal medical information. This approach would allow the model to be trained across decentralized data sources without requiring raw data sharing.
- Cross-population generalizability and bias mitigation: Further validation among diverse population groups is needed to ensure fairness and applicability across demographics. Investigation and mitigation of data or model biases should be a priority.
- Integration with wearable health technology: Closer collaboration with companies that make wearable tech could make it easier to add the predictive model to consumer health tech, which would make early screening easier for the whole population.

The main goals of future research should be to improve model understanding, explore advanced data augmentation techniques, and combine multimodal data to make sure that AI-driven BC prediction models can be used reliably in clinical practice [15, 62].

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ijis, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/int/8884481 by Assiut University Code 71515, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Conflicts of Interest

The authors declare no conflicts of interest.

Author Contributions

Hosam El-Sofany: created the original concept for the research and developed the methodology and contributed to the design, implementation, writing, reviewing, and editing. Proofreading and checking against plagiarism were performed using the iThenticate program provided by the King Khalid University. Belgacem Bouallegue: methodology, design, writing, reviewing, and editing. Yasser M. Abd El-Latif: methodology, design, writing, reviewing, and editing. Islam A. T. F. Taj-Eddin: methodology, design, writing, reviewing, and editing.

Funding

This study was funded by the Deanship of Scientific Research, King Khalid University, RGP1/15/46.

Acknowledgments

The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through a small-group research under grant number: RGP1/15/46.

References

[1] H. Sung, J. Ferlay, R. L. Siegel, et al., "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA: A Cancer Journal for Clinicians* 71, no. 3 (2021): 209–249, https://doi.org/10.3322/caac.21660.

[2] O. Yersal and S. Barutca, "Biological Subtypes of Breast Cancer: Prognostic and Therapeutic Implications," *World Journal of Clinical Oncology* 5, no. 3 (2014): 412, https://doi.org/10.5306/wjco.v5.i3.412.

[3] A. Aftab, M. Usman, A. Mehmood, and T. Saba, "Breast Cancer Prediction Using Machine Learning Models: a Systematic Review," *Expert Systems with Applications* 186 (2022).

[4] A. Esteva, A. Robicquet, B. Ramsundar, et al., "A Guide to Deep Learning in Healthcare," *Nature Medicine* 25, no. 1 (2019): 24–29, https://doi.org/10.1038/s41591-018-0316-z.

[5] Y. Liu, G. Sun, F. Wang, J. Li, D. Feng, and Y. Xiao, "Deep Learning in Breast Cancer Prognosis and Prediction," *Genes* 13, no. 3 (2022): 368.

[6] J. M. Johnson and T. M. Khoshgoftaar, "Survey on Deep Learning with Class Imbalance," *Journal of Big Data* 6, no. 1 (2019): 27–54, https://doi.org/10.1186/s40537-019-0192-5.

[7] G. Douzas, F. Bacao, and F. Last, "Improving Imbalanced Learning Through a Heuristic Oversampling Method Based on k-means and SMOTE," *Information Sciences* 465 (2018): 1–20, https://doi.org/10.1016/j.ins.2018.06.056.

[8] M. M. Rahman, D. N. Davis, and A. M. Aldahmash, "Addressing Data Imbalance in Cancer Prediction Using SMOTE," *Computers in Biology and Medicine* 133 (2021).

[9] C. Rudin, "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead," *Nature Machine Intelligence* 1, no. 5 (2019): 206–215, https://doi.org/10.1038/s42256-019-0048-x.

[10] S. M. Lundberg and S. I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems* (2017), 4765–4774.

[11] C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable* (Lulu.com, 2020).

[12] E. Tjoa and C. Guan, "A Survey on Explainable Artificial Intelligence (XAI): toward Medical XAI," *IEEE Transactions on Neural Networks and Learning Systems* 32, no. 11 (2021): 4793–4813, https://doi.org/10.1109/tnnls.2020.3027314.

[13] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A Dataset for Breast Cancer Histopathological Image Classification," *IEEE Transactions on Biomedical Engineering* 63, no. 7 (2016): 1455–1462, https://doi.org/10.1109/tbme.2015.2496264.

[14] R. Ha, P. Chang, S. Mutasa, et al., "Convolutional Neural Network Using a Breast MRI Tumor Dataset Can Predict Oncotype DX Recurrence Score," *Journal of Magnetic Resonance Imaging* 49, no. 2 (2019): 518–524, https://doi.org/10.1002/jmri.26244.

[15] G. Litjens, T. Kooi, B. E. Bejnordi, et al., "A Survey on Deep Learning in Medical Image Analysis," *Medical Image Analysis* 42 (2017): 60–88, https://doi.org/10.1016/j.media.2017.07.005.

[16] Y. Xie, J. Zhang, Y. Xia, M. Fulham, and Y. Zhang, "Fusing Texture, Shape and Deep model-learned Information at Decision Level for Automated Classification of Lung Nodules on Chest CT," *Information Fusion* 42 (2018): 102–110, https://doi.org/10.1016/j.inffus.2017.10.005.

[17] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine Learning Applications in Cancer Prognosis and Prediction," *Computational and Structural Biotechnology Journal* 13 (2015): 8–17, https://doi.org/10.1016/j.csbj.2014.11.005.

[18] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a Few Examples: a Survey on few-shot Learning," *ACM Computing Surveys* 53, no. 3 (2020): 1–34, https://doi.org/10.1145/3386252.

[19] A. Yala, C. Lehman, T. Schuster, T. Portnoi, and R. Barzilay, "A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction," *Radiology* 292, no. 1 (2019): 60–66, https://doi.org/10.1148/radiol.2019182716.

[20] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Oversampling Technique," *Journal of Artificial Intelligence Research* 16, no. 16 (2002): 321–357, https://doi.org/10.1613/jair.953.

[21] H. He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering* 21, no. 9 (2009): 1263–1284, https://doi.org/10.1109/tkde.2008.239.

[22] G. E. Batista, R. C. Prati, and M. C. Monard, "A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data," *ACM SIGKDD Explorations Newsletter* 6, no. 1 (2004): 20–29, https://doi.org/10.1145/1007730.1007735.

[23] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced Data: Review of Methods and Applications," *Expert Systems with Applications* 73 (2017): 220–239, https://doi.org/10.1016/j.eswa.2016.12.035.

[24] M. Buda, A. Maki, and M. A. Mazurowski, "A Systematic Study of the Class Imbalance Problem in Convolutional Neural Networks," *Neural Networks* 106 (2018): 249–259, https://doi.org/10.1016/j.neunet.2018.07.011.

[25] G. Douzas and F. Bacao, "Effective Data Generation for Imbalanced Learning Using Conditional Generative

ijis, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/int/8884481 by Assiut University Code 71515, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Adversarial Networks," *Expert Systems with Applications* 91 (2018): 464–471, https://doi.org/10.1016/j.eswa.2017.09.030.

[26] S. M. Lundberg, G. Erion, and S. I. Lee, "Consistent Individualized Feature Attribution for Tree Ensembles," *arXiv preprint arXiv:1802.03888* (2018).

[27] X. Yang, L. Wang, W. Zhang, H. Yu, and L. Xie, "Explainable Deep Learning for Medical Imaging: a Survey" (2019).

[28] Q. Zhang and S. C. Zhu, "Visual Interpretability for Deep Learning: a Survey," *Frontiers of Information Technology & Electronic Engineering* 19, no. 1 (2018): 27–39, https://doi.org/10.1631/fitee.1700808.

[29] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G. Z. Yang, "XAI—Explainable Artificial Intelligence," *Science Robotics* 4, no. 37 (2019): https://doi.org/10.1126/scirobotics.aay7120.

[30] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal, "Explaining Explanations: an Overview of Interpretability of Machine Learning," in *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)* (May 2018), 80–89, https://doi.org/10.1109/dsaa.2018.00018.

[31] W. Zhao, B. Lei, and Y. Yang, "Integrating Deep Learning with multi-modal Data for Precision Medicine in Cancer," *Nature Reviews Cancer* 21, no. 10 (2021): 611–630.

[32] S. Tonekaboni, S. Joshi, M. D. McCradden, and A. Goldenberg, "What Clinicians Want: Contextualizing Explainable Machine Learning for Clinical End Use" (2019).

[33] X. Li, Z. Hu, Y. Ma, and W. Zhao, "Multi-Modal Deep Learning: Methods and Applications," *Neural Networks* 134 (2021): 57–73.

[34] Z. Zhao, X. Zeng, L. Wang, Y. Wang, and L. Zhou, "A Novel CNN-based Method for Predicting Lung Cancer Growth on Serial Cts," *Computers in Biology and Medicine* 104 (2019): 97–106.

[35] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Transactions on Knowledge and Data Engineering* 22, no. 10 (2010): 1345–1359, https://doi.org/10.1109/tkde.2009.191.

[36] H. C. Shin, H. R. Roth, M. Gao, et al., "Deep Convolutional Neural Networks for computer-aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," *IEEE Transactions on Medical Imaging* 35, no. 5 (2016): 1285–1298, https://doi.org/10.1109/tmi.2016.2528162.

[37] F. Zhuang, Z. Qi, K. Duan, et al., "A Comprehensive Survey on Transfer Learning," *Proceedings of the IEEE* 109, no. 1 (2021): 43–76, https://doi.org/10.1109/jproc.2020.3004555.

[38] M. Gagliardi, T. Ruga, E. Vocaturo, and E. Zumpano, "Predictive Analysis for Early Detection of Breast Cancer Through Artificial Intelligence Algorithms," in *Innovations in Computational Intelligence and Computer Vision. ICICV 2024. Lecture Notes in Networks and Systems*, ed. S. Roy, D. Sinwar, N. Dey, and J. M. Tavares (2024).

[39] S. M. Riazul Islam, D. Kwak, M. Humaun Kabir, M. Hossain, and K. S. Kwak, "The Internet of Things for Health Care: a Comprehensive Survey," *IEEE Access* 3 (2015): 678–708, https://doi.org/10.1109/ACCESS.2015.2437951.

[40] M. Patel and M. Patel, "Internet of Things-IoT: Definition, Characteristics, Architecture, Enabling Technologies, Application & Future Challenges," *International Journal of Engineering Science and Computing* 6, no. 5 (2016): 6122–6131.

[41] A. Kundu, P. Saini, and R. Sinha, "Iot-based real-time Breast Cancer Detection Using Naive Bayes Classifier," *Biomedical Signal Processing and Control* 68 (2021): https://doi.org/10.1016/j.bspc.2021.102785.

[42] M. Chen, J. Yang, J. Zhou, Y. Hao, and J. Zhang, "5G-Smart Diabetes: toward Personalized Diabetes Diagnosis with Healthcare Big Data Clouds," *IEEE Communications Magazine* 56, no. 4 (2018): 16–23, https://doi.org/10.1109/MCOM.2018.1700494.

[43] L. Zhou, J. Bao, and I. M. Setiawan, "Personalized Health Monitoring Systems for Elderly Healthcare Based on Cloud Computing and Internet of Things," *IEEE Access* 8 (2020): 26241–26251, https://doi.org/10.1109/ACCESS.2020.2971393.

[44] V. Chang and M. Ramachandran, "Towards Achieving Data Security with the Cloud Computing Adoption Framework," *IEEE Transactions on Services Computing* 9, no. 1 (2016): 138–151, https://doi.org/10.1109/TSC.2015.2491281.

[45] N. Bui and M. Zorzi, "Health Care Applications: a Solution Based on the Internet of Things," in *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies* (ACM, August 2011), 1–5.

[46] S. Li, L. D. Xu, and S. Zhao, "The Internet of Things: a Survey," *Information Systems Frontiers* 20, no. 2 (2018): 243–259.

[47] Y. Zhang, M. Qiu, C. W. Tsai, M. M. Hassan, and A. Alamri, "Health-CPS: Healthcare Cyber-Physical System Assisted by Cloud and Big Data," *IEEE Systems Journal* 11, no. 1 (2017): 88–95, https://doi.org/10.1109/jsyst.2015.2460747.

[48] X. Zhang and L. Wang, "A Survey on Internet of Things (Iot) and its Applications," *Computer Networks* 174 (2021): https://doi.org/10.1016/j.comnet.2020.107181.

[49] Y. Zhang, H. Zhang, J. Cai, and X. Liu, "Real-Time Human Health Monitoring Using the Internet of Things Framework," *IEEE Internet of Things Journal* 4, no. 5 (2017): 1057–1065, https://doi.org/10.1109/JIOT.2017.2724037.

[50] A. Kaushik and D. Gupta, "Internet of Things and Machine Learning in Healthcare," *Journal of King Saud University-Computer and Information Sciences* 34, no. 1 (2020): 295–306, https://doi.org/10.1016/j.jksuci.2020.05.011.

[51] M. S. Hossain and G. Muhammad, "Cloud-Assisted Industrial Internet of Things (IIoT)–enabled Framework for Health Monitoring," *Computer Networks* 101 (2016): 192–202, https://doi.org/10.1016/j.comnet.2016.01.009.

[52] M. Kaushal, M. Singh, and S. Singh, "Iot-based Framework for Predictive Analytics of Breast Cancer," *Journal of Ambient Intelligence and Humanized Computing* 11, no. 9 (2020): 3573–3582, https://doi.org/10.1007/s12652-020-01838-7.

[53] I. Azimi, A. M. Rahmani, P. Liljeberg, and H. Tenhunen, "Internet of Things for Remote Elderly Monitoring: a Study from User-Centered Perspective," *Journal of Ambient Intelligence and Humanized Computing* 8, no. 2 (2017): 273–289, https://doi.org/10.1007/s12652-016-0387-y.

[54] H. El-Sofany, S. A. El-Seoud, O. H. Karam, Y. M. Abd El-Latif, and I. A. T. F. Taj-Eddin, "A Proposed Technique Using Machine Learning for the Prediction of Diabetes Disease Through a Mobile App," *International Journal of Intelligent Systems* 2024 (2024): 1–13, https://doi.org/10.1155/2024/6688934.

[55] G. Litjens, T. Kooi, B. E. Bejnordi, et al., "A Survey on Deep Learning in Medical Image Analysis," *Medical Image Analysis* 42 (2017): 60–88, https://doi.org/10.1016/j.media.2017.07.005.

[56] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep Learning to Improve Breast Cancer Detection on Screening Mammography," *Scientific Reports* 9, no. 1 (2019): https://doi.org/10.1038/s41598-019-48995-4.

[57] A. Esteva, A. Robicquet, B. Ramsundar, et al., "A Guide to Deep Learning in Healthcare," *Nature Medicine* 25, no. 1 (2019): 24–29, https://doi.org/10.1038/s41591-018-0316-z.

ijis, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/int/8884481 by Assiut University Code 71515, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[58] S. M. McKinney, M. Sieniek, V. Godbole, et al., "International Evaluation of an AI System for Breast Cancer Screening," *Nature* 577, no. 7788 (2020): 89–94, https://doi.org/10.1038/s41586-019-1799-6.
[59] X. Zhang, et al., "Multi-Modal Deep Learning Model for Survival Prediction of Breast Cancer," *Scientific Reports* 10 (2020): 8147.
[60] E. Vocaturo and E. Zumpano, "Artificial Intelligence Approaches on Ultrasound for Breast Cancer Diagnosis," in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (Houston, TX, USA, June 2021), 3116–3121, https://doi.org/10.1109/BIBM52615.2021.9669690.
[61] "Wisconsin (Diagnostic) Dataset" (2024).
[62] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks," *Communications of the ACM* 60, no. 6 (2017): 84–90, https://doi.org/10.1145/3065386.

ijis, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/int/8884481 by Assiut University Code 71515, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License