



Article

# A Hybrid Deep Learning Framework with Decision-Level Fusion for Breast Cancer Survival Prediction

Nermin Abdelhakim Othman <sup>1,2,\*</sup>, Manal A. Abdel-Fattah <sup>1</sup>  and Ahlam Talaat Ali <sup>1,3,\*</sup>

<sup>1</sup> Faculty of Computers and Artificial Intelligence, Helwan University, Cairo 11795, Egypt

<sup>2</sup> Faculty of Informatics and Computer Science, British University in Egypt, Cairo 11837, Egypt

<sup>3</sup> Faculty of Computer Science, Nahda University, Beni Suef 62521, Egypt

\* Correspondence: drnermin@fci.helwan.edu.eg (N.A.O.); ahlam.ali@nub.edu.eg (A.T.A.)

**Abstract:** Because of technological advancements and their use in the medical area, many new methods and strategies have been developed to address complex real-life challenges. Breast cancer, a particular kind of tumor that arises in breast cells, is one of the most prevalent types of cancer in women and is. Early breast cancer detection and classification are crucial. Early detection considerably increases the likelihood of survival, which motivates us to contribute to different detection techniques from a technical standpoint. Additionally, manual detection requires a lot of time and effort and carries the risk of pathologist error and inaccurate classification. To address these problems, in this study, a hybrid deep learning model that enables decision making based on data from multiple data sources is proposed and used with two different classifiers. By incorporating multi-omics data (clinical data, gene expression data, and copy number alteration data) from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) dataset, the accuracy of patient survival predictions is expected to be improved relative to prediction utilizing only one modality of data. A convolutional neural network (CNN) architecture is used for feature extraction. LSTM and GRU are used as classifiers. The accuracy achieved by LSTM is 97.0%, and that achieved by GRU is 97.5, while using decision fusion (LSTM and GRU) achieves the best accuracy of 98.0%. The prediction performance assessed using various performance indicators demonstrates that our model outperforms currently used methodologies.

**Keywords:** breast cancer; multimodal data; deep learning; LSTM; GRU; decision-level fusion; voting classifier



**Citation:** Othman, N.A.;

Abdel-Fattah, M.A.; Ali, A.T. A

Hybrid Deep Learning Framework with Decision-Level Fusion for Breast Cancer Survival Prediction. *Big Data Cogn. Comput.* **2023**, *7*, 50. <https://doi.org/10.3390/bdcc7010050>

Academic Editors: Guarino Alfonso, Rocco Zaccagnino and Emiliano Del Gobbo

Received: 29 January 2023

Revised: 9 March 2023

Accepted: 13 March 2023

Published: 16 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The medical field faces a fundamental problem in offering reliable and easily accessible diagnoses [1]. Because some physicians lack the necessary experience, diagnostic mistakes can happen, such as a patient's diagnosis being completely missed, improperly delayed, or incorrect. As a result, researchers have attempted to build computer-aided methods to assist physicians in making decisions. Among the different cancers that affect women, breast cancer has a higher incidence rate than lung cancer [2]. Medical decision-making systems are used to diagnose patients using data from laboratory examinations in the form of text, numbers, and images. The greatest threat to women's health is breast cancer, the most prevalent type of cancer among women in the world [3]. Accurate diagnostic results are critical in this context for precision medicine. Accurate survival prediction is critical for patients with breast cancer because it allows physicians to make informed decisions and further direct appropriate therapies [4].

Due to the implications of gene modification, researchers are now interested in copy number variation, gene expression, and clinical information when studying breast cancer. Even medical specialists with decades of expertise have trouble predicting and treating breast cancer because it takes a lot of cognitive effort for a human to understand the

pertinent information from various sources [5]. Two categories of cancer patient life expectancies are long-term survival (survival of over five years) and short-term survival (fewer than five years of survival). Once sufferers are anticipated to survive for a short term, clinicians can utilize predictive models to help them suggest targeted cancer treatments, sparing them from needless adjuvant therapy and the pain brought on by its harmful side effects [6]. Combining clinical and genetic data may improve prognosis and diagnosis prediction models.

Machine learning and deep learning are frequently employed in the prediction of breast cancer. However, the majority of researchers use just one deep learning model, for example, CNN, RNN, LSTM, or GRU. As a result, it was determined that the performance of these models was inadequate. Hybrid DL models can be used to effectively enhance classification performance [7–11]. Deep learning and hybrid deep learning algorithms have made significant progress in recent years in a variety of areas, including computer vision and natural language processing. Curating appropriate training datasets and selecting an appropriate evaluation measure are the two most important phases in developing an effective deep learning approach [12].

The accuracy of breast cancer classification using only one modality still fails to meet therapeutic needs. Because natural factors are so complicated, it is difficult for a single modality to provide complete knowledge for analysis. Therefore, multimodal data provide more advantages for complex analyses [13]. Multimodal data fusion is a core method of multimodal data mining that attempts to combine data from various distributions, sources, and types into a global space that can uniformly represent both intermodality and cross modality [14]. Using multimodal medical data fusion algorithms has led to improvements in clinical accuracy. There are three different types of fusion techniques: data-level, feature-level, and decision-level fusion [15]. In order to improve survival classification accuracy, we propose a hybrid deep learning model based on multimodal data to predict the survival of breast cancer patients in order to assist algorithms in responding to changes. In this research, we investigate a model that predicts the survival of patients with breast cancer by integrating clinical information and genetic traits. The main contributions of this paper are summarized as follows:

- In this study, we present a novel hybrid DL (CNN-LSTM+ CNN-GRU) model that automatically extracts features from the METABRIC dataset and classifies patients as long-term survivors and short-term survivors to minimize pathologist errors;
- It is suggested that the hybrid DL model (CNN-LSTM + CNN-GRU) be used to effectively classify breast cancer survival prediction in medical research;
- An ensemble model is presented that provides highly accurate breast cancer prediction. The final prediction is made via a voting mechanism;
- The suggested CNN-LSTM, CNN-GRU, and hard voting (LSTM-GRU) models were evaluated, and their major performance measures were compared to current prediction models using the same dataset (METABRIC). In comparison to other models, we found that the suggested hybrid deep learning model achieves outstanding classification results;

The remainder of this study is organized as follows: In Section 2, we briefly present some related work that motivated our research. In Section 3, we describe our proposed method; in Section 4, we evaluate the results; in Section 5, experimental results are presented; in Section 6, we present a discussion; and finally, we conclude this paper by mentioning our intentions for our future work.

## 2. Related Work

In this section, we briefly review and discuss the state-of-the-art algorithms for breast cancer prediction and related fields. Pathologists can benefit from deep learning algorithms, as they can improve diagnostic accuracy while reducing computing time. Several algorithms have been utilized for machine learning and deep learning.

In recent decades, several approaches have been adopted for computer-aided diagnostics using only source data for breast cancer prediction, such as pathological images. EMR. R. Sanyal et al. [16] proposed a novel attention method for breast histology image classification. CNN was used for feature extraction, and BLSTM was used as an encoder network. The MLP decoder predicts the image class, achieving an accuracy of 85.50% for patch classification and 96.25% for image classification; this illustrates the effectiveness of the suggested attention technique. Owing to a lack of training data and because of using single-modality datasets, classification accuracy is still limited, D. M. Vo et al. [17] used deep learning models with convolutional layers for breast cancer classification. They used an ensemble of DCNNs for feature extraction and gradient-boosting trees for classification. The combination of DCNNs and a gradient-boosting tree classifier produced better classification performance, with an accuracy of 96.4% for four classes and 99.5% for two classes. We need to resolve the issues of limited training samples and imbalanced data.

Some authors have used multimodal data fusion. Although the use of deep learning has substantially enhanced the performance of breast cancer classification, the classification accuracy obtained using only single-source data still falls short of therapeutic needs. Consequently, a combination of features from multimodal data can result in more effective and improved outcomes. Arya, Nikhilanand, et al. [6]. Proposed deep-learning-based prediction techniques in a stacked ensemble architecture to enhance breast cancer survival prediction using a multimodal dataset. They first selected important features using the mRMR method; then, features were extracted from different modalities using the CNN architecture, which were fed into the stacked layer. In step two, the RF classifier was used for prediction. This model outperforms previous multimodality and unimodality prediction algorithms, achieving an accuracy of 90.2% and an AUC value of 0.93. This study required the use of a large dataset and multimodality datasets (image modality and miRNA expression values). In another study, the same authors [18] combined gated attentive DL models with random forest classifiers to increase the accuracy of BC predictions using multimodal data. They created the “SiGaAttCNN” architecture for feature extraction from several modalities in which the extracted features are passed to the stacked layer. RF classifiers with adaptive boosting were used for prediction in stage two. A comparison of the suggested approach with other current approaches shows 5.1% greater sensitivity values, which is a considerable improvement. The suggested model’s performance may be improved by the addition of other modalities, which could result in a model that is more powerful and effective.

While also using decision-level fusion, a combination of techniques can result in more effective and improved outcomes. Yadav, Rohit, et al. [19] proposed a feature and decision fusion method for breast cancer detection. They first used the CLAHE enhancement method for low-contrast mammography images. Then they used a CNN for feature extraction. ML algorithms (SVM, DT, and RF) were used as classifiers. Finally, they used a voting classifier, which generated the final prediction. Three algorithms achieved accuracies of 92.3%, 94.3%, and 95%, respectively. The results show that RF is better than SVM and DT for BC prediction. Using the voting classifiers (SVM, DT, and RF), accuracy improved by 96.18%. Tewary, Suman, et al. [20] presented a computerized HER2 quantification model using transfer learning and statistical voting. They used five pretrained architectures for classification: VGG16, VGG19, ResNet50, MobileNetV2, and NASNetMobile. The transfer learning models demonstrated appreciable accuracy, with VGG19 displaying the best accuracy of 93% for the test images, which increased to 98% for image-based scoring employing a statistical voting system. The results demonstrate the strength of the proposed quantification process for creating automated HER2.

Several studies have been published in the literature in which attempts were made to use hybrid deep learning models to improve prediction accuracy. Yan, Rui, et al. [21] developed a new approach combining a hybrid convolutional and recurrent deep neural network for the classification of breast cancer pathology images. This method combines the advantages of CNN and RNN networks. They used a CNN to extract features and an

RNN to combine patch features for final image classification. The experimental results demonstrate that, with an average accuracy of 91.3% for the four-class classification test, this approach surpasses the state-of-the-art method. A sufficiently broad and diverse dataset is essential to increase classification accuracy. Wang, X. et al. [22] proposed a hybrid deep learning model (CNN-GRU) for the automatic detection of BC-IDC (+). The proposed model automatically implements several CNN layers and GRU to predict breast IDC (+) cancer. The approach produced an average classification of ACC of 86.21%, a PR of 8.590%, an SN of 85.71%, an F1 of (88%), and an AUC of 0.89. A comparison of the proposed model’s output with those of CNN-BiLSTM and other existing ML/DL models showed that CNN-GRU has 4–5% higher accuracy and a less time-intensive processing time.

According to previous studies, the CNN methodology has proven to be the most effective feature extraction technique, and the accuracy of the models is improved by using multimodal data and the ensemble methodology. MLP is a method that has not been utilized frequently in the diagnosis of breast cancer, as shown in Table 1. The use of deep learning and hybrid deep learning algorithms based on multimodal datasets significantly improves classification accuracy.

**Table 1.** Fusion-level models for breast cancer classification reported in the literature.

Fusion-Level Type	Algorithm	Dataset	Advantages	Disadvantages
Data-Level Fusion [23]	<ul style="list-style-type: none"> <li>■ Data augmentation</li> <li>■ Convolutional network (DenseNet)</li> </ul>	Mini-DDSM BUSI (mammography images and ultrasound images)	<ul style="list-style-type: none"> <li>• Earlier fusion and later fusion</li> <li>• The fusion of two image modalities (mammography and ultrasound)</li> </ul>	<ul style="list-style-type: none"> <li>• A limited number of medical images is available, requiring more preprocessing</li> </ul>
Data-Level Fusion [24]	<ul style="list-style-type: none"> <li>■ CNN architecture (VGG, ResNet, and DenseNet)</li> <li>■ Ensemble method</li> </ul>	SNUH and BUSI datasets (ultrasound images)	<ul style="list-style-type: none"> <li>• Uses a method for image fusion and various representations of the image content</li> <li>• Uses an ensemble of various CNN architectures</li> </ul>	<ul style="list-style-type: none"> <li>• The ROI region and tumor contour are traditionally cropped by expert-defined criteria in the B-mode US image, resulting effects of human interventions</li> </ul>
Feature-Level Fusion [6]	<ul style="list-style-type: none"> <li>■ CNN for feature extraction</li> <li>■ Random forest stack-based ensemble model for classification</li> </ul>	METABRIC (clinical, gene expression, and CNA data)	<ul style="list-style-type: none"> <li>• Uses a multimodal dataset</li> <li>• Uses a stacked ensemble</li> <li>• Manual omission to determine the scoring-level fusing coefficients</li> </ul>	<ul style="list-style-type: none"> <li>• Small, imbalanced dataset</li> <li>• Some data modalities are not available, such as image modality and miRNA expression values</li> </ul>
Feature-Level Fusion [18]	<ul style="list-style-type: none"> <li>■ SiGaAtCNN’ for feature extraction</li> <li>■ Random forest stack-based ensemble model for classification</li> </ul>	METABRIC (clinical, gene expression, and CNA data)	<ul style="list-style-type: none"> <li>• A new CNN architecture</li> <li>• Improved classification results</li> <li>• Includes the idea of sigmoid-gated attention and creates more informative features for classifications</li> </ul>	<ul style="list-style-type: none"> <li>• Small, imbalanced dataset</li> <li>• Some data modalities are not available, such as image modality and miRNA expression values</li> </ul>
Decision-Level Fusion [20]	<ul style="list-style-type: none"> <li>■ Transfer learning</li> <li>■ VGG16, VGG19, ResNet50, MobileNetV2, and NASNetMobile for classification</li> <li>■ Statistical voting</li> </ul>	Warwick University dataset	<ul style="list-style-type: none"> <li>• Using the suggested method of decision-level fusion, statistical voting increased the accuracy for the VGG19 pretrained model from 93 to 98%</li> </ul>	<ul style="list-style-type: none"> <li>• The is acquired from hospital records, and according to standard procedure, at least two specialists score the records. The results may be improved if the scores could be made available to different pathologists and the regions were marked according to all photos in the dataset rather than just an overall score</li> </ul>
Decision-Level Fusion [19]	<ul style="list-style-type: none"> <li>■ CLAHE approach</li> <li>■ CNN model for feature extraction</li> <li>■ RF, SVM, and DT classifiers</li> <li>■ Voting classifier</li> </ul>	MIAS (mammogram images)	<ul style="list-style-type: none"> <li>• Uses CLAHE and RMSHE for contrast enhancement</li> <li>• Uses decision fusion</li> </ul>	<ul style="list-style-type: none"> <li>• Multiple imaging modalities are not used, such as ultrasound, X-ray, and magnetic images</li> </ul>

### 3. Proposed Method

#### 3.1. Feature Selection

Dimensionality reduction is an important prediction and classification procedure. It helps in the improvement of classification performance by decreasing the number of attributes and removing unnecessary and unrelated attributes [25]. The METABRIC dataset used in our proposed contains approximately 24,000 genes in the gene expression profile data, approximately 26,000 genes in the CNA profile data, and 27 features in the clinical profile data. Deep learning techniques do not perform effectively due to the dataset's high dimensionality and limited data. To reduce the dimensionality of our dataset, Sun et al. employed the well-known mRMR feature selection technique. The performance of the features was then assessed using the area under the curve (AUC) value. Sun et al. searched for the best N features between 100 and 500 using a step size of 100. The final features included in our MDNNMD model are 200 genes from CNA profile data, 400 genes from gene expression profile data, and 25 clinical features from clinical profile data [18]. A full description of all features of the multidimensional data used in this research is provided in Table 2.

**Table 2.** Dataset properties.

Data Type	Complete Features	Chosen Features
Clinical information	27	25
Genetic expression information	24,368	400
Copy number	26,298	200

#### 3.2. Feature Extraction and Feature Fusion

In this work, a convolutional neural network (CNN) is used for feature extraction. Convolutional neural networks (CNNs) achieve impressive performance in cancer detection and diagnosis [26]. A CNN receives inputs from various data modalities, and each input dataset is passed through convolution layers containing a specific number of kernels or filters, generating a feature map as an output of the convolution process. This feature map is created by summing the corresponding values from the input matrix and the filter matrix after performing a straightforward element-by-element multiplication. The Glorot normal initializer [27] is used to initialize the filter matrix values, choosing random numbers with a mean equal to zero and a standard deviation in the range of  $[-\sqrt{\frac{2}{m_i+n_o}}, \sqrt{\frac{2}{m_i+n_o}}]$ . The number of input and output units is represented here by  $m_i$  and  $n_o$ , respectively, for the chosen layer. The biases used by the convolution layer are initialized with a constant value of 0.1. To conduct convolution across the full input matrix in this layer, we utilized a stride value of 2 to shift the filter with the specific number. The flexibility to use padding to regulate the feature map's size is another characteristic of the convolution layer. This CNN maintains the size of the feature map at the same level as the input data shape using padding. Furthermore, the flattening layer flattens the output of the convolution layer before being passed through a fully connected dense layer with 150 hidden units. After the dense layer, an output layer with a sigmoid activation function generates the predicted class of breast cancer prognosis. The activation functions for the dense layer and convolution layer are rectified linear (ReLU) and hyperbolic tangent (TANH), respectively.

In the CNN architecture, we employed binary cross entropy as a loss function, since predicting the prognosis of breast cancer is a binary classification task. To prevent the model from overfitting, L2 regularization is used with the loss function. The L2 regularization technique is a popular regularization method in deep learning [28]. In this method, 10-fold cross validation is used to solve the variance issue posed by the small size of the dataset. Additionally, the training set is split into a training set (80%) and a validation set (20%).

The CNN model contains a single input layer, a single convolution layer, a single flattened layer, a single fully connected dense layer, and a single output layer. Because a complex CNN based on a small dataset can overfit, the CNN design is not overly complex. Our CNN employs an additional dropout layer of 50% dropout for gene expression and CNA profiles [29]. The model's AUC value was assessed at several mini batch sizes, ranging from 8 to 128, and an optimal size of 8 was chosen because it produced the highest AUC value. Our model's learning rate is  $10^{-3}$ . Table 3 provides information on the detailed parameter configurations of CNN.

**Table 3.** Model configuration for the CNN with detailed parameters.

Number of convolutional layers	1
Filter size	15
Number of filters	4
Stride size	2
Padding in the convolutional layer	Same
Activation function	ReLU
Number of hidden layers	1
Number of hidden neurons	150
Mini batch size	8
Training epochs	20
Activation function	TANH
Loss function	binary cross entropy + L2 regularization

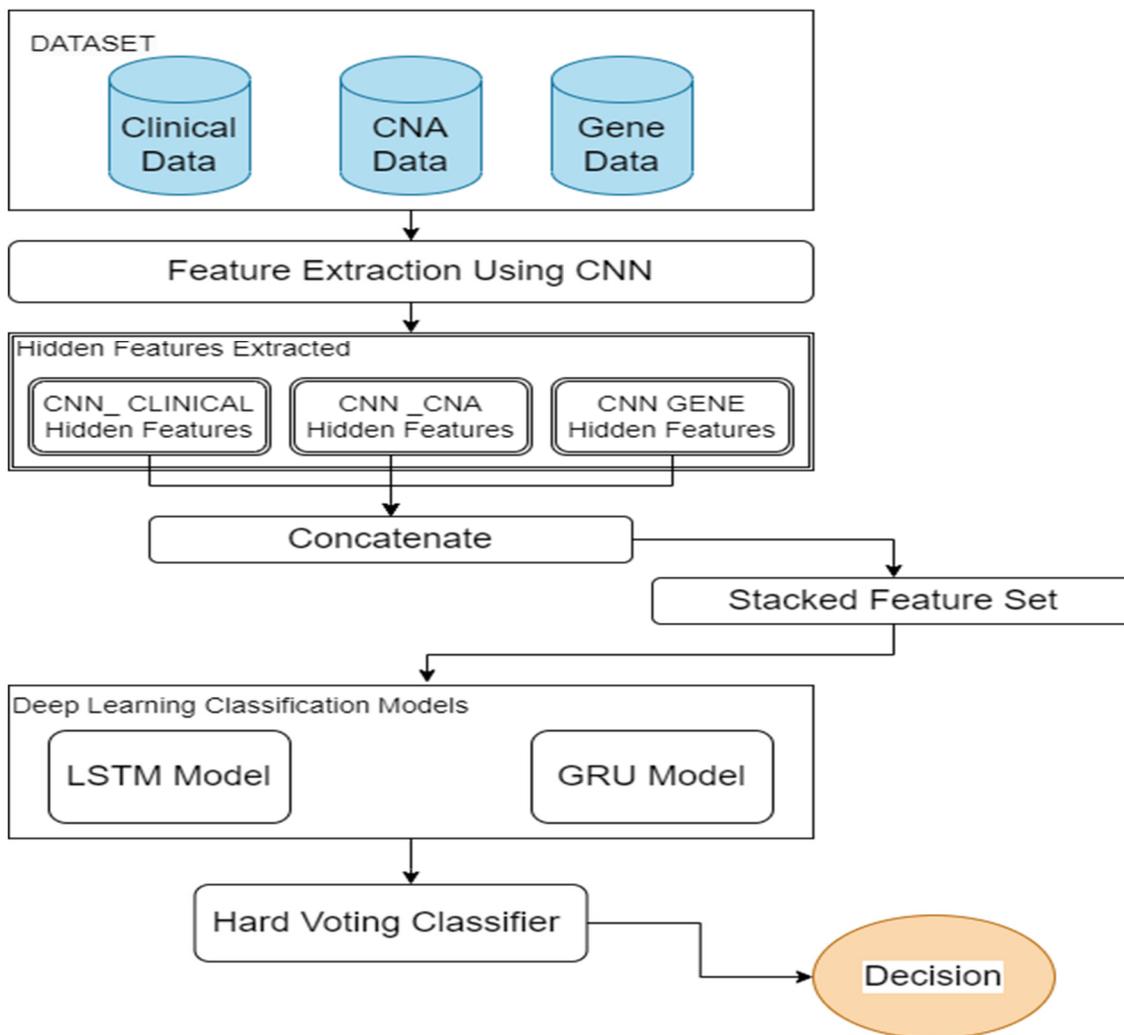
### 3.3. Deep Learning Classification Models and Decision-Level Fusion

The major goal of the proposed method is to use deep learning models to predict whether a patient has breast cancer disease or not. For these predictions, the proposed method uses a framework that is built on GRU and LSTM. Recent research has focused on how to create an intelligent system with LSTM and GRU that collects a patient's medical information and calculates the likelihood of contracting tumors using deep-learning-based network models. First, the stacked features are used to train individual modality classifiers (LSTM and GRU) [30,31]. Then, using a voting classifier, these two input classifiers are combined to make a decision. Long short-term memory (LSTM) and gated recurrent units (GRU) resolve the gradient-vanishing and exploding problems. The neural network becomes unstable and unable to learn from training data if the gradients begin to explode. By using LSTM and GRU together, networks can take advantage of the strengths of both units, i.e., the ability to learn long-term associations for the LSTM and the ability to learn from short-term patterns for the GRU. LSTM and GRU exhibited good outcomes for the majority of performance indicators, making them the most effective algorithms [32]. Figure 1 demonstrates the steps of the proposed model.

Long short-term memory (LSTM) is a subclass of recurrent neural networks (RNNs). All inputs and outputs in a typical neural network are independent of one another. RNNs have a kind of memory that they may utilize to access previous data, but this does not imply that they store data over a long period; rather, they store data from just a few prior steps. The short-term memory issue with RNNs was addressed with the development of the LSTM network. The main advantage of LSTM over RNNs is its ability to achieve long-term dependency learning [33].

Gated Recurrent Unit (GRU): The second deep learning model used for the proposed method is the gated recurrent unit (GRU). To solve the vanishing gradient problem, recurrent neural networks typically employ GRU models. With three primary gates and an internal cell state, GRU is more effective than LSTM [34]. Due to parameter reduction, GRU outperforms LSTM models in some contexts and has a higher convergence rate than LSTM.

GRU takes less computational time than LSTM. The data are secretly maintained within the GRU for security purposes. The structure of the GRU is simpler than that of LSTM, which reduces matrix multiplication, and GRU can save a lot of time without sacrificing performance. GRUs have been shown to perform better on smaller datasets [35].



**Figure 1.** The workflow of the breast cancer survival prediction process.

Voting classifier: The accuracy and efficiency of classification results have increased because of the widespread use of ensemble models. Performance can improve over time when classifiers are merged compared to utilizing individual models. To achieve superior results, this method predicts breast cancer using an ensemble learning approach [36]. We suggest using LSTM and GRU deep learning models in an ensemble voting classifier for detection of breast cancer. A customized CNN is used to extract standout features from the multimodal dataset instead of manually creating them. These collected features are concatenated and used as training data for LSTM and GRU.

The final prediction is made through voting on the results of these models. The suggested approach uses a hard voting classifier with voting criteria and integrates LSTM and GRU. This ensemble model, which is trained by the separate models, is in charge of predicting the output class label by merging the majority of projected class votes acquired for each class label. This ensemble method is typically appropriate when there are two or more classifier models that predict in a nearly identical manner. Figure 2 illustrate Architecture of ensemble hard voting classifier. In majority voting, the class label (y) is predicted by majority (plurality) voting among the classifiers (C). In this instance, the class

with the most votes is chosen. Here, we use a majority vote from each classifier to predict the class label (y).

$$y = \text{mode} \{C_1(X), C_2(X), \dots, C_n(X)\}$$

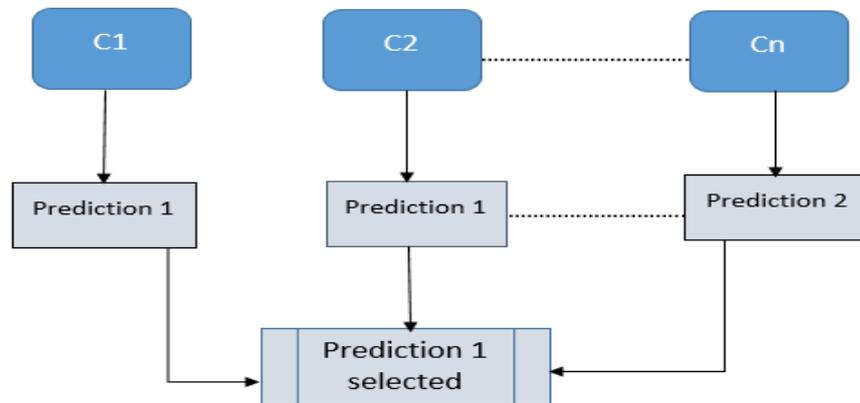


Figure 2. Architecture of ensemble hard voting classifier.

Each classifier has a probability ranging from 0 to 1. If two classifiers have a probability  $< 0.5$ , the voting classifier classifies the label as “class 0”. If two classifiers have probabilities  $> 0.5$ , the voting classifier assign the class label “class 1”. If classifier 1 has a probability  $> 0.5$  and the prediction class label is 1 and if classifier 2 has a probability  $< 0.5$  and the prediction class label is 0, the voting classifier classifies the label as “class 1” based on the majority class label because classifier 1 has more confidence than classifier 2.

The model that is suggested in this paper is based on GRU-CNN and LSTM-CNN, with four dense layers, as well as an alternative GRU and LSTM layer order. LSTM and GRU layers follow dropout layers. Dropout layers are used to prevent overfitting in this model. The proposed model has four LSTM and GRU layers with 128, 64, 32, and 16 units, respectively. A layer with a 20% dropout rate follows each of these layers. Sigmoid and ReLU activation functions are employed in the layers. Finally, the Adam optimizer is used to compile the layers over 40 epochs, with a batch size of 128. The LSTM network comprises 189,433 parameters, and the GRU network comprises a total of 157,273 parameters, which are then used to train the models to make predictions. A comprehensive description of the LSTM and GRU models is provided in Tables 4 and 5, respectively.

Table 4. Details of the LSTM model.

Layers	Number of Units	Number of Received Parameters	Resultant Shape
LSTM	128	66,560	(None, 452, 128)
Dropout	0.2	0	(None, 452, 128)
LSTM	64	49,408	(None, 452, 64)
Dropout	0.2	0	(None, 452, 64)
LSTM	32	12,416	(None, 452, 32)
Dropout	0.2	0	(None, 452, 32)
LSTM	16	3136	(None, 452, 16)
Dropout	0.2	0	(None, 452, 16)
Flatten	0	0	(None, 7232)
Dense	8	57,864	(None, 8)
Dense	4	36	(None, 4)
Dense	2	10	(None, 2)
Dense	1	3	(None, 1)

**Table 5.** Details of the GRU model.

Layers	Number of Units	Number of Received Parameters	Resultant Shape
GRU	128	50,304	(None, 452, 128)
Dropout	0.2	0	(None, 452, 128)
GRU	64	37,248	(None, 452, 64)
Dropout	0.2	0	(None, 452, 64)
GRU	32	9408	(None, 452, 32)
Dropout	0.2	0	(None, 452, 32)
GRU	16	2400	(None, 452, 16)
Dropout	0.2	0	(None, 452, 16)
Flatten	0	0	(None, 7232)
Dense	8	57,864	(None, 8)
Dense	4	36	(None, 4)
Dense	2	10	(None, 2)
Dense	1	3	(None, 1)

#### 4. Results Evaluation

In this section, we first describe the dataset used in this study. Next, the various results of our experiments are reported for feature extraction and classification. Finally, a comparison is made between our model's outcomes and those previously reported in the literature.

##### 4.1. Dataset Description

To directly evaluate the proposed model we accessed on 1 August 2022 the preprocessed version of the METABRIC dataset, which is available on GitHub (<https://github.com/USTC-Hilab/MDNNMD>). Data from 1980 valid breast cancer patients in the METABRIC trial were used to create the dataset [37]. This covers breast cancer multimodal data such as clinical data profiles, gene expression profiles, and CNA profiles. The patients were divided into two groups: long-term survivors (those who survived for more than five years), comprising 1489 patients, and short-term survivors (those who survived for less than five years), comprising 491 patients. The average patient survival was 125.1 months, and the average age at diagnosis was 61 years. A general overview of our dataset is included in Table 6.

**Table 6.** Details about the dataset.

Survival limit (years)	5
# of patients	1980
Long-term survivors	1489
Short-term survivors	491
Average age at diagnosis	61
Average survival (months)	125.1

For this binary classification model, the long-term survivors are assigned the label 0, and the short-term survivors were assigned the label 1. Table 7 shows the description of the METABRIC dataset. Missing values for gene expression profile data and CNA profile data were estimated using the weighted nearest-neighbor technique [38]. The min-max normalization method was used to normalize the features of clinical data into the range [0,1]. Regarding CNA features, we used the five discrete values of the original data (−2, −1, 0, 1, and 2). Table 8 shows the description of the clinical dataset.

**Table 7.** METABRIC dataset description.

Dataset	Data Type	Dataset Description
Clinical	Numerical/Categorical	<p>Clinical features were classified into four categories:</p> <ol style="list-style-type: none"> <li>1. Personal: age at diagnosis</li> <li>2. Clinical pathology: tumor size, tumor stage, lymph_nodes_examined_positive, neoplasm_histologic_grade, histological type, er_status, HER2_SNP6_state</li> <li>3. Treatment: type of treatment the patient received</li> <li>4. Survival: status and time</li> </ol>
CNA	Categorical	<p>Copy number aberration features describe each region within a chromosome (number of markers and type of mutation in the somatic tissues):</p> <ul style="list-style-type: none"> <li>- Location information: chrom, loc.start, loc.end, segment, call</li> <li>- Number of genes within the segment: num.mark</li> <li>- Type of mutation: call2; NEUT (neutral), HOMO (homozygous deletion), HETD (hemizygous deletion), CNV (copy number variations), GAIN (gain), AMP (amplification)</li> </ul>
Gene Expression	Numerical	48,803 EXPRESSED GENE ILLUMINA SEQUENCED HT 12 array v3

**Table 8.** CLINICAL dataset description.

S. No.	Attribute	Value Examples
1	Age at diagnosis	21 to 96 years
2	Histologic grade	1, 2, 3
3	Tumor size	1 to 182 mm
4	Tumor stage	
5	Positive examined lymph nodes	0 to 45
6	Inferred menopausal state	Pre, Post
7	ER status	Positive, negative
8	PR status	Positive, negative
9	Overall survival (months)	0 to 355
10	Histological type	Ductal/NST, lobular
11	HER2_SNP6_state	NETURAL, LOSS, GAIN
12	Treatment	Chemotherapy
13	Patients vital status	Overall survival status (0: yes, 1: no)

#### 4.2. Evaluation Criteria

To evaluate performance, we used an ROC curve to display the difference between false-positive and true-positive rates. Using the ROC curve, we computed the AUC value as a metric of classification results. Standard measures were used to evaluate the models: accuracy, sensitivity, precision, and Matthew's correlation coefficient. TP indicates true positive, FP indicates false positive, TN indicates true negative, and FN indicates false negative in a confusion matrix. The following equations define the evaluation metrics:

$$\text{accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{Mcc} = \frac{TP * TN - FP * FN}{\sqrt{(TP + FN) * (TP + FP) * (TN + FN) * (TN + FN)}} \tag{4}$$

### 5. Experimental Results

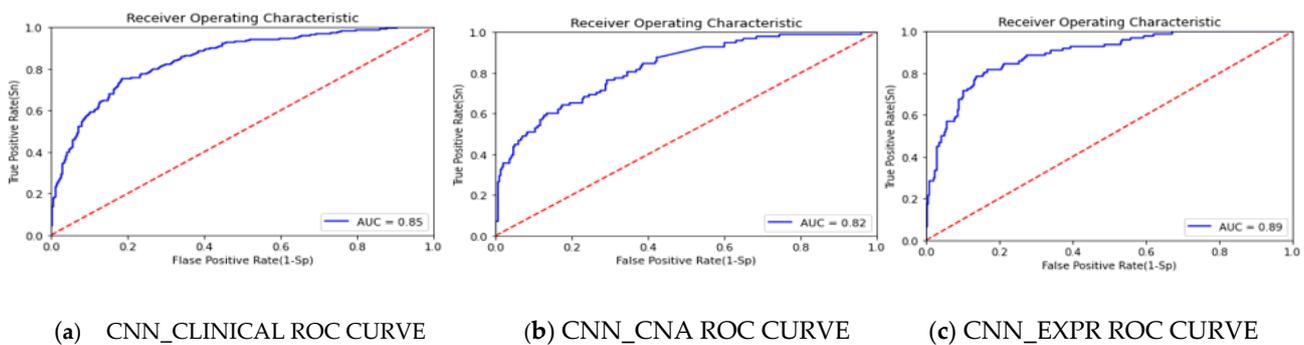
This section covers the results of applying deep learning models (LSTM and GRU) to the stacked features extracted from each trained CNN. Additionally, we discuss the performance of CNN networks on each separate unimodal data category, as well as how hard voting classifiers affect the decision. The METABRIC dataset was divided into a training set (80%) and a validation set (20%).

#### 5.1. Performance Metrics of Unimodal CNN

In this step, the hidden features were extracted from each separate unimodal dataset using a CNN model: clinical modality data, CNA modality data, and gene expression modality data. Table 9 shows the ACC and AUC of trained CNN models. These extracted features were combined to create the “stacked features”, which were then fed into the deep learning models for breast cancer survival prediction. Figure 3 shows ROC curves and AUC values of CNNs trained on each data modality.

**Table 9.** Comparison of ACC and AUC between trained CNN models.

Model	ACC	AUC
CNN_CLINICAL	80.8	85
CNN_CNA	74.3	82
CNN_EXPR	80.2	89



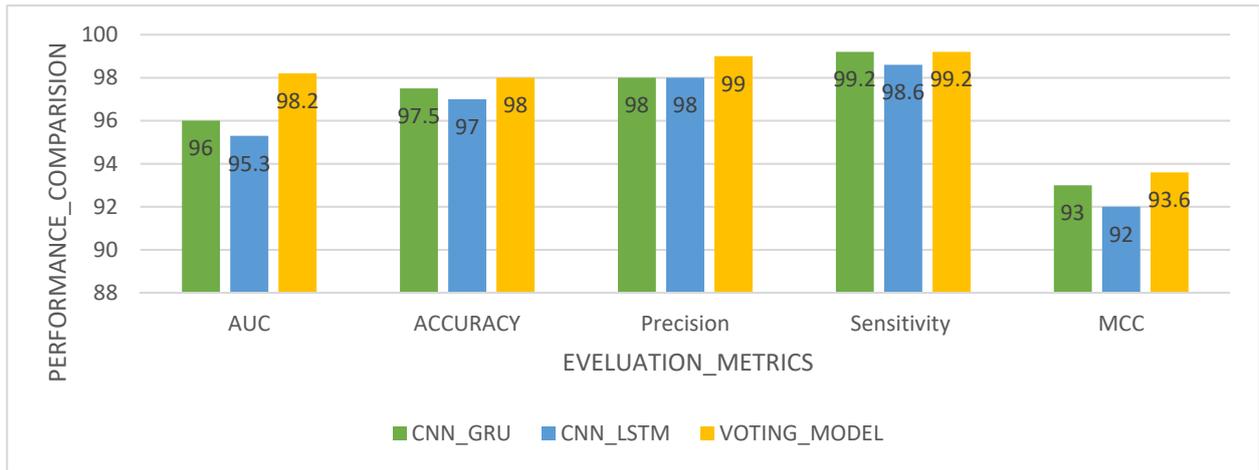
**Figure 3.** ROC curves and AUC values of CNNs trained on each data modality.

#### 5.2. Performance Metrics of LSTM, GRU, and Voting Classifier with Stacked Features

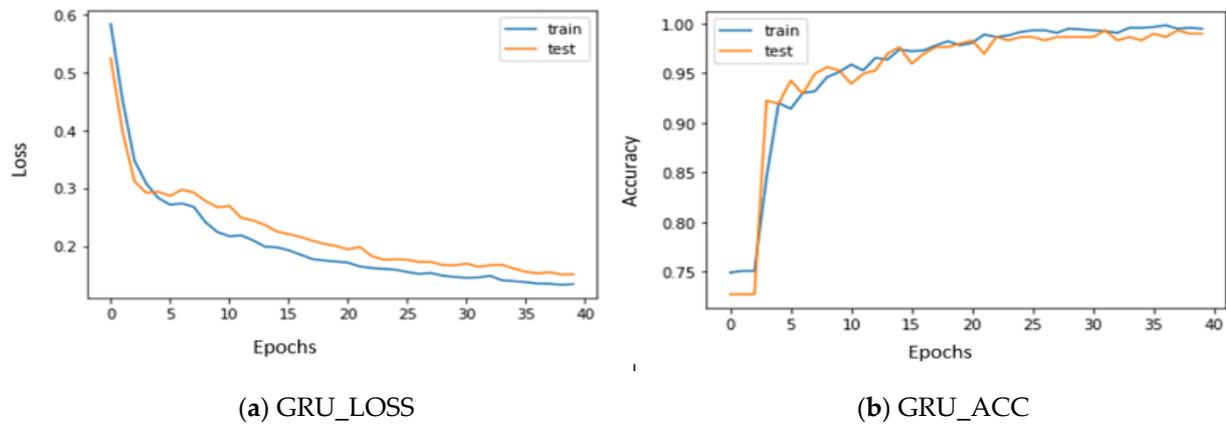
The outcomes of using stacked features with deep learning are shown in Table 10. Figure 4 shows all the classification results. Overall, GRU achieved the highest performance (AUC = 96.0%, ACC = 97.5%, PR = 98.0%, SN = 99.2%, MCC = 93.0%). Figure 5 shows the loss and the accuracy curve of GRU\_model, while LSTM achieved the lowest performance (AUC = 95.3%, ACC = 97.0%, PR = 98.0%, SN = 98.6%, MCC = 92.0%). Figure 6 shows the loss and the accuracy curve of LSTM\_model. Figure 7 shows the confusion matrix curve of the LSTM and GRU models

**Table 10.** The performance of CNN\_GRU, CNN\_LSTM, and VOTING models with stacked features.

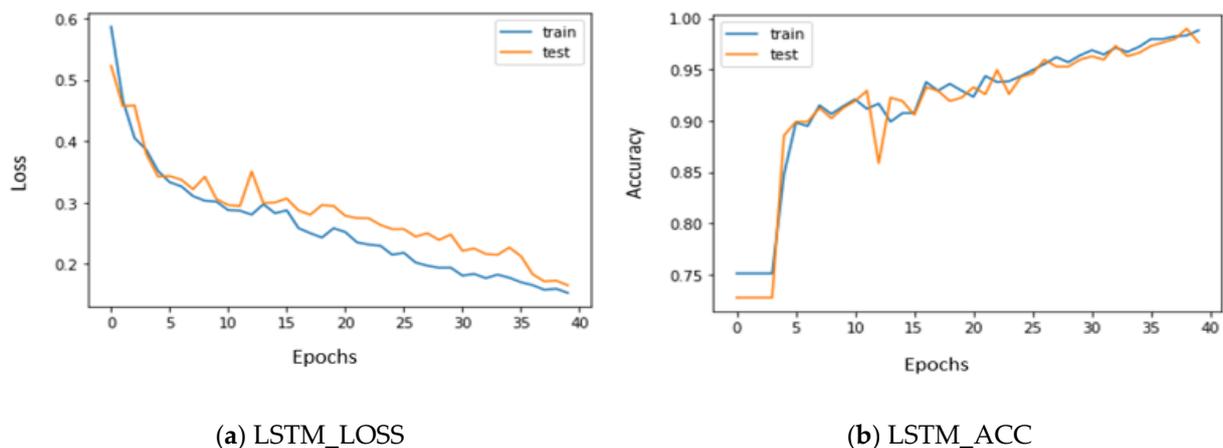
Performance Metric (%)	Classification	Dataset	CNN_GRU	CNN_LSTM	VOTING Model
AUC (%)	Binary Classification	METABRIC Dataset	96.0	95.3	98.2
Accuracy (%)			97.5	97.0	98.0
Precision (%)			98.0	98.0	99.0
Sensitivity (%)			99.2	98.6	99.2
MCC (%)			93.0	92.0	93.6



**Figure 4.** The classification performance of the CNN\_GRU, CNN\_LSTM, and VOTING models.



**Figure 5.** The loss and the accuracy curve of GRU\_model.



**Figure 6.** The loss and the accuracy curve of LSTM\_model.

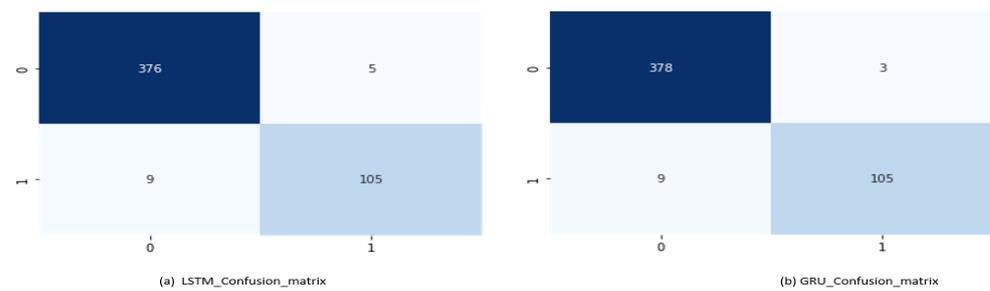


Figure 7. The confusion matrix curve of the LSTM and GRU models.

### 5.3. Decision-Level Fusion Using Hard Voting Classifier

The voting ensemble model performs well when compared to the performance of the individual models. The performance was improved by the combination of LSTM and GRU, both of which performed well on their own. According to the experimental results, the proposed voting ensemble model, LSTM + GRU, surpasses all other models and achieved the greatest accuracy of 98.0%. Figure 8 shows the confusion matrix curve of the hard voting model (GRU-LSTM).

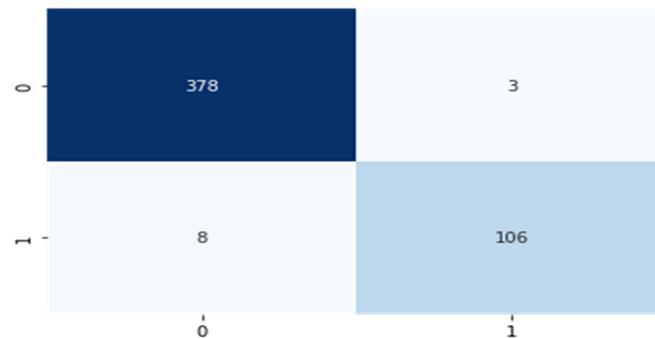


Figure 8. The confusion matrix curve of the hard voting model (GRU-LSTM).

### 5.4. Comparison of Various Classification Techniques

We contrasted five popular methods for predicting the diagnosis of breast cancer with deep learning models: SiGaAtCNN stacked RF, stacked RF, MDNNMD, SVM, and LR. The AUC values for deep learning models were determined, including LSTM and GRU. The AUC values for machine learning models were also determined, including SiGaAtCNN STACKED RF, stacked RF, MDNNMD, SVM, and LR. When compared to other prediction techniques, the GRU model achieved the highest AUC value. To compare deep learning models with different prediction techniques, alternative performance measures, including accuracy, precision, sensitivity, and MCC, were utilized. The outcomes are displayed in Table 10. According to the comparative study shown in Table 11, deep learning models outperformed all other methods in predicting the prognosis of breast cancer.

Table 11. Comparative study of performance measures of GRU, LSTM, and various classification techniques.

Model	AUC	ACC	PR	SN	MCC
Proposed Model	98.2	98.0	99.0	99.2	93.6
GRU	96.0	97.5	98.0	99.2	93.0
LSTM	95.3	97.0	98.0	98.6	92.0
SiGaAtCNN STACKED RF	95.0	91.2	84.1	79.8	76.2
STACKED RF	93.0	90.2	84.1	74.7	73.0
MDNNMD	84.5	82.6	74.9	45.0	48.6
SVM	81.0	80.5	70.8	36.5	40.7
LR	66.3	76.0	54.9	18.3	20.9

## 6. Discussion

The fundamental concept underlying this study is a hybrid deep learning model with ensemble learning using a multimodal dataset to enhance the performance of breast cancer survival prediction. As shown in Table 10, the performance measures of 98.2% AUC, 98.0% ACC, 99.0% PR, 99.2% SN, and 93.6% MCC obtained by the hard voting classifier are better than the performance measures obtained by the individual classifiers. In addition, it is evident from Figures 7 and 8 that hard voting achieved the highest value of correctly classified instances and the lowest value of incorrectly classified instances compared to the other classifiers. After building the proposed model, we analyzed the effectiveness of our algorithms. Table 11 demonstrates that the hard voting classifier and deep learning models GRU and LSTM achieved highest values compared to other machine learning methods.

This research has an advantage over previous studies in that it includes the use of hybrid deep learning models on multimodal datasets. Consequently, the performance of breast cancer survival prediction was improved. However, this research is subject to some restrictions. The first restriction relates to the generalizability of findings because our tests were conducted on a small dataset. As a result, if we consider big dataset tests with different datasets, the outcome may change. The second restriction has to do with choosing just one ensemble approach. We reported the results according to the functionality of the hard voting method. Nevertheless, we only applied two deep learning models.

## 7. Conclusions and Future Work

The most prevalent disease in the world, breast cancer, significantly contributes to the rising mortality rate among cancer patients. Breast cancer remains a concern, and additional study is needed to improve early detection and provide accurate survival predictions. Therefore, it is essential to create a quick and efficient method for predicting the prognosis of breast cancer. In this study, a hybrid deep learning model that enables decision making based on data from multiple data sources was proposed and used with two different classifiers to predict the life expectancy of breast cancer patients. Gene expression, copy number alterations, and clinical data are some of the multimodal inputs used in this model. The hidden features were extracted from each separate unimodal dataset using a CNN model. The extracted features were concatenated to form stacked features, which were then fed into deep learning classifiers (GRU and LSTM) to predict breast cancer survival. A method comprising feature fusion and decision fusion was utilized to improve the results. By applying a voting classifier, the classification of cancer was improved. The LSTM model produced an ACC of 97.0%, Prec of 98.0%, Sens of 98.6%, MCC of 92.0%, and AUC of 95.3, while the GRU model produced an ACC of 97.5%, Prec of 98.0%, Sens of 99.2%, MCC of 93.0%, and AUC 96.0. Using the voting classifier, the proposed model produced an ACC of 98.0%, Prec of 99.0%, Sens of 99.2%, MCC of 93.6%, and AUC of 98.2, which can reduce the pathologist's errors and efforts during the clinical process. This model performs better than other existing prediction techniques. A more powerful and informative model might be created by incorporating more modalities, which would also enhance the proposed model's performance metrics. We may consider incorporating more modalities in the future, such as data on miRNA expression, gene methylation, and histological images of breast cancer tissues. More deep learning models that can perform the prognosis prediction task for patients with all types of cancer can be added to this research, allowing it to be expanded even further.

**Author Contributions:** Conceptualization, M.A.A.-F. and N.A.O.; methodology, A.T.A.; software, A.T.A. and N.A.O.; validation, N.A.O.; formal analysis, A.T.A. and N.A.O. and M.A.A.-F.; investigation, N.A.O.; data curation, A.T.A.; writing—original draft preparation, A.T.A.; writing—review and editing, N.A.O.; supervision, M.A.A.-F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Ethical review and approval were waived for this study, due to the reason that we use the public datasets.

**Informed Consent Statement:** Informed consent was waived for this study due to the reason that we use the public datasets.

**Data Availability Statement:** METABRIC Dataset (<https://github.com/USTC-HILab/MDNNMD>, accessed on 1 August 2022).

**Acknowledgments:** The authors are grateful to the reviewers and editors for providing advice and guidance as to how to improve the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Richens, J.G.; Lee, C.M.; Johri, S. Improving the accuracy of medical diagnosis with causal machine learning. *Nat. Commun.* **2020**, *11*, 1–9. [CrossRef]
- Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2019. *CA A Cancer J. Clin.* **2019**, *69*, 7–34. [CrossRef]
- Liu, M.L.; Hu, L.; Tang, Y.; Wang, C.; He, Y.; Zeng, C. A Deep Learning Method for Breast Cancer Classification in the Pathology Images. *IEEE J. Biomed. Heal. Inform.* **2022**, *26*, 5025–5032. [CrossRef] [PubMed]
- Guo, W.; Liang, W.; Deng, Q.; Zou, X. A Multimodal Affinity Fusion Network for Predicting the Survival of Breast Cancer Patients. *Front. Genet.* **2021**, *12*, 709027. [CrossRef]
- Khamparia, A.; Bharati, S.; Podder, P.; Gupta, D.; Khanna, A.; Phung, T.K. Diagnosis of breast cancer based on modern mammography using hybrid transfer learning. *Multidimens. Syst. Signal Process.* **2021**, *32*, 747–765. [CrossRef] [PubMed]
- Arya, N.; Saha, S. Multi-modal classification for human breast cancer prognosis prediction: Proposal of deep-learning based stacked ensemble model. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2020**, *19*, 1032–1041. [CrossRef]
- Tufail, A.B.; Ma, Y.-K.; Kaabar, M.K.; Martínez, F.; Junejo, A.; Ullah, I. Deep learning in cancer diagnosis and prognosis prediction: A minireview on challenges, recent trends, and future directions. *Comput. Math. Methods Med.* **2021**, *2021*, 9025470. [CrossRef]
- Khan, R.; Yang, Q.; Ullah, I.; Rehman, A.U.; Bin Tufail, A.; Noor, A.; Rehman, A.; Cengiz, K. 3D convolutional neural networks based automatic modulation classification in the presence of channel noise. *IET Commun.* **2022**, *16*, 497–509. [CrossRef]
- Tufail, A.B.; Ullah, I.; Khan, W.U.; Asif, M.; Ahmad, I.; Ma, Y.-K.; Khan, R.; Ali, M.S. Diagnosis of Diabetic Retinopathy through Retinal Fundus Images and 3D Convolutional Neural Networks with Limited Number of Samples. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 1–15. [CrossRef]
- Kamruzzaman, M. Architecture of Smart Health Care System Using Artificial Intelligence. In Proceedings of the 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), London, UK, 6–10 July 2020; pp. 1–6.
- Ahmad, I.; Liu, Y.; Javeed, D.; Shamshad, N.; Sarwar, D.; Ahmad, S. A review of artificial intelligence techniques for selection & evaluation. *IOP Conf. Series: Mater. Sci. Eng.* **2020**, *853*, 012055. [CrossRef]
- Zou, J.; Huss, M.; Abid, A.; Mohammadi, P.; Torkamani, A.; Telenti, A. A primer on deep learning in genomics. *Nat. Genet.* **2019**, *51*, 12–18. [CrossRef]
- Liu, T.; Huang, J.; Liao, T.; Pu, R.; Liu, S.; Peng, Y. A Hybrid Deep Learning Model for Predicting Molecular Subtypes of Human Breast Cancer Using Multimodal Data. *Irbm* **2021**, *43*, 62–74. [CrossRef]
- Gao, J.; Li, P.; Chen, Z.; Zhang, J. A Survey on Deep Learning for Multimodal Data Fusion. *Neural Comput.* **2020**, *32*, 829–864. [CrossRef]
- Nazari, E.; Chang, H.-C.H.; Deldar, K.; Pour, R.; Avan, A.; Tara, M.; Mehrabian, A.; Tabesh, H. A comprehensive Overview of Decision Fusion Technique in Healthcare: A Systematic Scoping Review. *Iran. Red Crescent Med. J.* **2020**, *22*. [CrossRef]
- Sanyal, R.; Jethanandani, M.; Sarkar, R. DAN: Breast Cancer Classification from High-Resolution Histology Images Using Deep. *Atten. Netw.* **2020**, *1189*, 319–326. [CrossRef]
- Vo, D.M.; Nguyen, N.-Q.; Lee, S.-W. Classification of breast cancer histology images using incremental boosting convolution networks. *Inf. Sci.* **2019**, *482*, 123–138. [CrossRef]
- Arya, N.; Saha, S. Multi-modal advanced deep learning architectures for breast cancer survival prediction. *Knowledge-Based Syst.* **2021**, *221*, 106965. [CrossRef]
- Yadav, R.; Sharma, R.; Pateriya, P.K. Feature and Decision Fusion for Breast Cancer Detection. In *Data Analytics and Management*; Khanna, A., Polkowski, Z., Castillo, O., Eds.; Springer Nature Singapore: Singapore, 2022; pp. 737–747. [CrossRef]
- Tewary, S.; Mukhopadhyay, S. HER2 Molecular Marker Scoring Using Transfer Learning and Decision Level Fusion. *J. Digit. Imaging* **2021**, *34*, 667–677. [CrossRef]
- Yan, R.; Ren, F.; Wang, Z.; Wang, L.; Zhang, T.; Liu, Y. Breast cancer histopathological image classification using a hybrid deep neural network. *Methods* **2019**, *173*, 52–60. [CrossRef] [PubMed]
- Wang, X.; Ahmad, I.; Javeed, D.; Zaidi, S.A.; Alotaibi, F.M.; Ghoneim, M.E. Intelligent Hybrid Deep Learning Model for Breast Cancer Detection. *Electronics* **2022**, *11*, 2767. [CrossRef]
- Hamdy, E.; Badawy, O.; Zaghloul, M. Densely Convolutional Networks for Breast Cancer Classification with Multi-modal Image Fusion. *Int. Arab. J. Inf. Technol.* **2022**, *19*, 12. [CrossRef]

24. Moon, W.K.; Lee, Y.-W.; Ke, H.-H.; Lee, S.H.; Huang, C.-S.; Chang, R.-F. Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. *Comput. Methods Programs Biomed.* **2020**, *190*, 105361. [[CrossRef](#)]
25. Sathiyabhama, B.; Kumar, S.U.; Jayanthi, J.; Sathiya, T.; Ilavarasi, A.K.; Yuvarajan, V. A novel feature selection framework based on grey wolf optimizer for mammogram image analysis. *Neural Comput. Appl.* **2021**, *33*, 14583–14602. [[CrossRef](#)]
26. Hu, Z.; Tang, J.; Wang, Z.; Zhang, K.; Zhang, L.; Sun, Q. Deep learning for image-based cancer detection and diagnosis—A survey. *Pattern Recognit.* **2018**, *83*, 134–149. [[CrossRef](#)]
27. Glorot, X.; Bengio, Y. Understanding the Difficulty of Training Deep Feedforward Neural Networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13 May 2010; pp. 249–256.
28. Havaei, M.; Davy, A.; Warde-Farley, D.; Biard, A.; Courville, A.; Bengio, Y. Brain tumor segmentation with Deep Neural Networks. *Med Image Anal.* **2017**, *35*, 18–31. [[CrossRef](#)]
29. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
30. Dutta, S.; Mandal, J.K.; Kim, T.H.; Bandyopadhyay, S.K. Breast Cancer Prediction Using Stacked GRU-LSTM-BRNN. *Appl. Comput. Syst.* **2020**, *25*, 163–171. [[CrossRef](#)]
31. Begum, A.; Kumar, V.D.; Asghar, J.; Hemalatha, D.; Arulkumaran, G. A Combined Deep CNN: LSTM with a Random Forest Approach for Breast Cancer Diagnosis. *Complexity* **2022**, *2022*, 1–9. [[CrossRef](#)]
32. Ghosh, P.; Azam, S.; Hasib, K.M.; Karim, A.; Jonkman, M.; Anwar, A. A Performance Based Study on Deep Learning Algorithms in the Effective Prediction of Breast Cancer. In Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 18–22 July 2021; pp. 1–8. [[CrossRef](#)]
33. Jafarbigloo, S.K.; Danyali, H. Nuclear atypia grading in breast cancer histopathological images based on CNN feature extraction and LSTM classification. *CAAI Trans. Intell. Technol.* **2021**, *6*, 426–439. [[CrossRef](#)]
34. Ahmad, S.; Ullah, T.; Ahmad, I.; Al-Sharabi, A.; Ullah, K.; Khan, R.A.; Rasheed, S.; Ullah, I.; Uddin, N.; Ali, S. A Novel Hybrid Deep Learning Model for Metastatic Cancer Detection. *Comput. Intell. Neurosci.* **2022**, *2022*, 1–14. [[CrossRef](#)]
35. Shah, A.A.; Alturise, F.; Alkhalifah, T.; Khan, Y.D. Deep Learning Approaches for Detection of Breast Adenocarcinoma Causing Carcinogenic Mutations. *Int. J. Mol. Sci.* **2022**, *23*, 11539. [[CrossRef](#)] [[PubMed](#)]
36. Umer, M.; Naveed, M.; Alrowais, F.; Ishaq, A.; Hejaili, A.A.; Alsubai, S. Breast Cancer Detection Using Convolved Features and Ensemble Machine Learning Algorithm. *Cancers* **2022**, *14*, 6015. [[CrossRef](#)] [[PubMed](#)]
37. Curtis, C.; Shah, S.P.; Chin, S.-F.; Turashvili, G.; Rueda, O.M.; Dunning, M.J. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **2012**, *486*, 346–352. [[CrossRef](#)] [[PubMed](#)]
38. Lin, W.-C.; Tsai, C.-F. Missing value imputation: A review and analysis of the literature (2006–2017). *Artif. Intell. Rev.* **2020**, *53*, 1487–1509. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.