Hybrid CNN and RNN Model for Histopathological Sub-Image Classification in Breast Cancer Analysis Using Self-Learning

Alaa Hussein Abdulaal[1*] , Morteza Valizadeh[2] , Riyam Ali Yassin[3] , Mehdi Chehel Amirani[4] , A. F. M. Shahen Shah[5] , Baraa M. Albaker[6] , Ammar Saad Mustaf [7]

[1,6] Department of Electrical Engineering, College of Engineering, Al-Iraqi University, Baghdad, Iraq
[2,3,4]Department of Electrical Engineering, Urmia University, West Azerbaijan, Iran
[5]Department of Electronics and Communication Engineering, Yildiz Technical University, Istanbul, Turkey
[7]Department of Missions and Cultural Relations, Al-Iraqi University, Baghdad, Iraq

*Email: Alaa.H.Abdulaal@aliraqia.edu.iq

Article Info	Abstract
	Breast cancer, a pervasive and life-threatening disease, necessitates the development of advanced classification techniques. This paper introduces a model that combines Convolutional Neural Networks with Recurrent Neural Networks to classify sub-images in breast cancer. By leveraging localized features from a pre-trained CNN and insights from the RNN, this innovative approach aims to enhance accuracy. A sub-image-based strategy is employed to capture localized characteristics more effectively. A hierarchical self-learning approach is implemented to gradually correct mislabeled images, utilizing an invariant rule informed by prior knowledge of potential labeling errors. The model incorporates VGG19, Google Net, and ResNet101 for classifying breast cancer sub-images at various magnifications (40X, 100X, 200X, and 400X) from the BreaKHis dataset. Among these, ResNet101 demonstrates a notable classification accuracy of 98.58% with CNN techniques. However, the hybrid model achieves an impressive accuracy of 99.76%. This approach is promising for advancing medical image classification, offering potential diagnosis and patient care improvements.

1. Introduction

The body gets cancer when aberrant cell proliferation accumulates and forms a tumor [1]. This development may be categorized as benign or malignant. Malignant tumors, on the other hand, are destructive and quickly developing, which may result in significant sickness in the body. Benign tumors are often less dangerous to human health since they do not spread to other body areas. The most severe and complicated genetic illness yet is cancer. Pathologists differentiate between benign and malignant cells primarily based on morphological abnormalities of the nucleus [2]. A global burden of disease (GBD) analysis found that there were 2,088,849 new instances of breast cancer (BC) in 2018, and it was responsible for 626,679 fatalities [3]. The fatality rate for women with breast cancer has dramatically risen among the 100 various forms of cancer. According to predictions, there will be over 27 million new instances of BC by 2030 [4]. These numbers demonstrate the need for and significance of the automated system in reducing the death rate for women.

Although there are other imaging techniques, pathology images were used for this investigation because they are the "gold standard" for cancer [5]. This is so that the underlying illness and its impact at the tissue level may be seen in greater detail using histopathological images. In other words, histopathology offers information about the tissue's microarchitecture. The illness often begins at the molecular level before progressing to the tissue and cellular levels [6]. When the illness first begins to harm the tissue, it is imperative to find it.

Additionally, only the histopathological image may be used to determine the specific causes of certain disorders [7]. In clinics, specialists often perform the biopsy. This procedure involves removing a portion of the patient's tissue and preparing it on a glass slide with H & E staining, which improves the cytoplasm's visibility (pink) and nuclei's visibility (purple).

The pathologists then use a microscope to examine these slides to investigate the tumor [8]. Pathologists look at the critical regions, the way cells are distributed and arranged within the tissue, the form and structure of the cells, how the nuclei are managed, and any abnormalities brought on by cancer [9]. Nevertheless, this procedure requires a lot of time, money, and knowledge in the field.

In recent years, significant advancements have been made in detecting BC through histological images. Many works have employed computer-aided diagnosis (CAD) systems to classify BC into benign and malignant types using machine-learning techniques.

Hameed et al. [10] utilized four deep learning models—VGG-16 and VGG-19, fully trained and fine-tuned—to categorize carcinoma and non-carcinoma breast tissue images. Their ensemble of fine-tuned models achieved an average accuracy and F1 score of 95.29%, with a sensitivity of 97.73% for carcinoma classification.

Gupta and Chawla [11] proposed a two-phase approach using pre-trained models (VGG-16, VGG-19, ResNet-50, and Xception) to extract features from the BreaKHis dataset. They used support vector machine and logistic regression for classification, achieving 92.5% accuracy with the ResNet-50 model, which was particularly effective at 100× and 40× magnification levels.

Jiang et al. [12] introduced BHCNet, a neural network with a small SE-ResNet module optimized by a Gauss error scheduler for fine-tuning SGD learning rates. This model demonstrated high accuracy, scoring between 98.87% and 99.34% for binary and 90.66% to 93.81% for multi-class classification. Haija and Adebanjo [13] explored transfer learning (TL) with a pre-trained ResNet-50 model on the BreaKHis dataset, achieving a notable accuracy of 99% using 75% of the data for training.

Alom et al. [14] developed an improved inception-residual CNN (IRRCNN) by combining ResNet, Inception-V4, and recurrent convolutional networks (RNN). This model outperformed individual networks with accuracies of 99.05% for binary and 98.59% for multi-class classification.

Srikantamurthy et al. [15] suggested a hybrid CNN-LSTM model for classifying BC subtypes, employing transfer learning on the BreakHis dataset. Using optimizers like Adam, RMSProp, and SGD, the CNN-LSTM model achieved 99% accuracy for binary classification and 92.5% for multi-class subtype classification. All these related works are presented in Table 1.

Table 1. Summary of the related works

Ref.	Model	Dataset	Classifier	Accuracy
[10]	CNN	WSI	Binary	95.29%
[11]	CNN-SVM	BreakHis	Binary	92.5%
[12]	BHCNe (CNN)	BreakHis	Binary	98.87%
			Multi-Class	90.66%
[13]	ResNet-50	BreakHis	Binary	99%
[14]	IRRCNN	BACH	Multi-Class	98.59%
		BreakHis	Binary	99.05%
[15]	CNN-LSTM	BreakHis	Binary	99%
			Multi-Class	92.5%

This work introduces a hybrid CNN-LSTM model for classifying benign and malignant BC histology sub-images using TL. The model combines features [16]-[21] from a pre-trained CNN based on ImageNet and an RNN. The CNN layers are initially frozen, while the RNN is trained with preprocessed data. Features from both networks are then integrated [22], and classification is performed using SoftMax [23]. The approach leverages spatial data properties for improved accuracy and employs LSTM to handle long context windows and prevent gradient issues. The model's architecture allows for simultaneous fine-tuning of CNN and RNN components, enhancing classification outcomes.

The following section will thoroughly explain the suggested hybrid CNN-LSTM model and details regarding the dataset utilized to confirm the proposed method of differentiating between benign and malignant BC. The findings from the recommended strategy will be compared with those from other approaches. Then, the benefits of the proposed technique will be compared to current methods.

2. Materials and Methods

The following section of this paper highlights the main breakthroughs and provides an overview of the proposed methodology used in sub-image classification in BC based on histopathological images.

2.1. BreaKHis Dataset

The BreaKHis dataset, which was created in the P&D Laboratory in Parana, Brazil, is used in this paper. BreaKHis was initially mentioned by Spanhol et al. [24]. The dataset consists of 7909 microscopic images of breast tumor tissue taken during clinical work between January 2014 and December 2014. The collection includes 2480 standard and 5429 malignant tissue samples. These categories were used to group the photographs in the dataset after they were taken with various magnification factors (40, 100, 200, and 400), as shown in Figs. 1 and 2. The 82 patients who provided the samples had their data anonymized [25]. Hematoxylin and eosin (HE) staining and surgical open biopsy (SOB) were used to create samples from breast tissue biopsy slides. Experienced pathologists diagnosed each slide after it had been labeled in the P&D Laboratory.
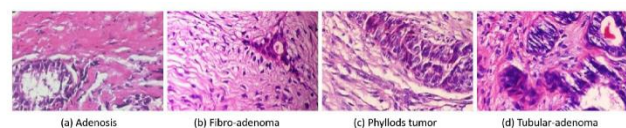

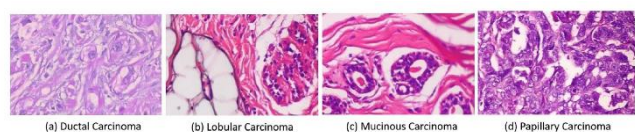**Figure 1.** Sub-classes of benign tumor images [26].

Figure 2. Sub-classes of malignant tumor images [26].

2.2. Sub-Image Creating

This paper uses the BreaKHis dataset in Tables 2 and 3 to show that images can be split into nine sub-images to make them easier to classify. The sub-image approach enables analyzing essential sections of an image, like other tissues or cancerous cells, instead of viewing the whole image as one. Abdulaal et al. [27] used this technique of sub-image classification to minimize errors in their work's accuracy.

Table 2. Image counts in the BreaKHis dataset

Class	Type	40X	100X	200X	400X
Benign	A	114	113	111	106
	F	253	260	264	237
	TA	109	121	108	115
	PT	149	150	140	130
Malignant	DC	864	903	896	788
	LC	156	170	163	137
	MC	205	222	196	169
	PC	145	142	135	138
	Total	1995	2081	2013	1820

Dividing the original image into smaller sub-images, as indicated in Fig. 3, made the classification task more directed and focused. For instance, every sub-image can capture local features and traits essential to the classification problem. This allows a more detailed analysis and improves classification accuracy.
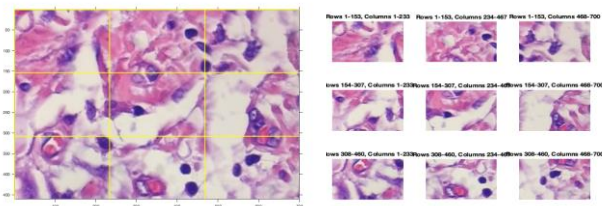

Figure 3. Sub-image operation

Table 3. Sub-image counts in the BreaKHis dataset

Class	Type	40X	100X	200X	400X
Benign	A	1026	1017	999	954
	F	2277	2340	2376	2133
	TA	981	1089	972	1035
	PT	1341	1350	1260	1170
Malignant	DC	7776	8127	8064	7092
	LC	1404	1530	1467	1233
	MC	1845	1998	1764	1521
	PC	1305	1278	1215	1242
	Total	17955	18729	18117	16380

This method has proved its efficiency in sub-image identification within the BreaKHis dataset by exhibiting the power of sub-image classification to boost the accuracy and sensitivity of breast cancer classifiers [27].

2.3. Hierarchical Self-Learning

A hierarchical self-learning approach based on literature by Abdulaal et al. [27] generated labels for new sub-images in the first case for this experiment. The idea is to gradually amend any wrong labels via a predefined rule incorporating prior information about mistakes that could have been made while labeling an initial set of sub-images. Employing such self-learning enabled us to observe consistent enhancements in accuracy in four step-wise improvements.

The first step toward obtaining the most accurate diagnosis is acquiring the labels associated with the newly acquired images. These labels were used to initiate the subsequent correction process.

Looking at the data hierarchy, the tags are refined through multiple iterations. Since the data is labeled, errors can be identified and rectified more quickly. Furthermore, these amendments were implemented across the whole dataset through interrelations and dependencies inherent within sub-images, enhancing overall precision.

The increased accuracy rates showed that the self-learning process proved highly effective after four rounds of label correction, significantly improving overall performance. Refers to how incremental changes in image annotation helped reduce instances of mislabeling, hence enhancing data set reliability.

This technique generates new sub-image-level annotated datasets, described in Table 4.

Table 4. New sub-image dataset

Type	Factor	Benign	Malignant	Total
BreaKHis		625	1370	1995
Sub-images		5625	12330	17955
Self-Learning	40X	7983	9972	17955
BreaKHis		644	1437	2081
Sub-images		5796	12933	18729
Self-Learning	100X	8137	10592	18729
BreaKHis		623	1390	2013
Sub-images		5607	12510	18117
Self-Learning	200X	8065	10052	18117
BreaKHis		588	1232	1820
Sub-images		5292	11088	16380
Self-Learning	400X	7576	8804	16380

2.4. Hybrid CNN-RNN Model

The rapid advancement of biotechnology has paved the way for developing CNNs as powerful models for image understanding. CNNs employ coordinated neurons that act as local filters across the entire input space, allowing for extracting local and deep features related to the input image. Another type of Neural Network that deals with sequence data is called RNN.

In this respect, unlike the traditional CNN architecture in which nodes and layers are placed in a forward sequence from the input layer to the output layer, the latter is inappropriate for dealing with sequential data types.

The training phase begins with pre-training the CNN on the ImageNet dataset, followed by BreaKHis data preprocessing steps. The TL strategy is used to start a new CNN with parameters of the pre-trained network. Subsequently, the RNN model is trained after freezing all the CNN layers. Once the training process is completed, the CNN layers are unfrozen, and the entire Hybrid model is trained.

Attention processes combine the features extracted by the RNN and CNN in a neural network framework. During the testing phase, preprocessed testing images are fed into the tuned Hybrid architecture, and classification outcomes are obtained using a SoftMax layer [23]. The model architecture includes multiple components: a CNN layer, an RNN layer, a merge layer, and a fully connected layer (FC) with a SoftMax output. Fig. 4 illustrates the overall model structure.

The pre-trained CNN layer employs weights acquired from pre-training on the ImageNet dataset to initialize the proposed CNN architecture. CNNs consist of pooling and convolutional layers.

The pooling layer performs a similar operation to the convolution layer but with a sliding window typically of size 2x2 and a sliding step of 2. This halves the feature map size from the previous layer, significantly reducing the feature map's size and the model's complexity. This reduction improves the speed of network training and enables the network to adapt to changes in image scales.

In this work, pre-trained models like VGG19 [28]-[31], Google Net [32]-[36], and ResNet101 [37]-[38] have been used, and the ReLU activation function has been employed. The convolution layer is the most significant part of a CNN. It performs convolutional operations with various-sized convolution windows on the feature maps of the previous layer. Typically, size 3x3 or 5x5 filters slide sequentially over the previous layer's feature map. The size of the window and the number of weight parameters in the convolution layer vary accordingly. The feature maps of the last layer are convolved through the corresponding filters, and the activation function used in the layer determines the outcome.

The RNN layer, like the CNN layer, consists of input, output, and hidden layers. The connections between the hidden layers are a significant feature of RNNs. The output of the hidden layer is fed back to the hidden layer itself, and there can also be connections between adjacent hidden nodes. This dynamic nature of the network makes RNNs closer to biological neural systems. The LSTM (Long Short-Term Memory) model is predominantly used in this work. LSTM is capable of learning long-term dependencies in data. The main difference between RNN and LSTM is that a structure has been added to determine the vital information. Input gate, output gate, and forget gate are three in it. The information fed into the LSTM network can be modified with these gates. Forget gate discards irrelevant information and only retains information that meets the algorithm's parameters.

The merging layer combines extracted features from CNN and RNN through a combinatorial model. A neural network that concentrates on only a few inputs or qualities assists in this undertaking.

The merge layer incorporates an attention mechanism to allow the model to prioritize important or relevant features during the merger. Consequently, more precise processing could be possible, finally improving overall performance.

The merge layer uses element-wise multiplication in conjunction with an attentional mechanism to form a composite feature representation capable of capturing both spatial and temporal information. A combination representation that includes the strong points of either model- a joint CNN and RNN model can be used in tasks like classification or other forms of analysis. In this case, the attentional neural network combines features extracted from CNN with those produced by RNN into merged features. The fact that this network has attention mechanisms allows it to focus on specific inputs or features to select the ones that further need processing. The
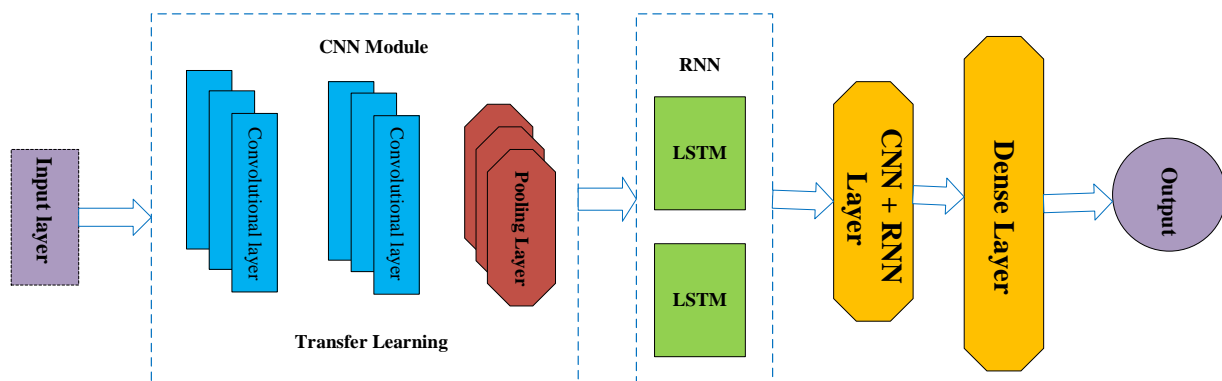

Figure 4. Hybrid model structure

number of hidden units or cells in the LSTM layer determines the size of the feature vectors. In this setup, an LSTM layer with 512 hidden units is used. Consequently, the feature vectors obtained from the LSTM layer would be 512.

The FC layer with SoftMax output receives the merged features from the RNN and CNN and produces a probability distribution over all the classes. Cross-entropy is the loss function that measures the difference between the actual and target outputs. The Weight values in the CNN branch utilize the parameters pre-trained on the ImageNet data set, and the RNN branch randomizes parameters. Those weight values are updated iteratively through a cross-entropy loss function gradient in the training process. The convolutional layer was frozen initially, and the training samples were estimated through the Adam optimizer, which requires 100 epochs for the training process. The training process is terminated following a certain number of periods. The model with a minimal verification loss value has been selected as an ultimate network.

In the architecture described, the fusion of features from the CNN and the LSTM is achieved using an attentional neural network mechanism or other fusion techniques. This fusion process combines the feature vectors obtained from the last layer of the CNN with the feature vectors from the LSTM. The attentional neural network mechanism used at the fusion step can assign different importance levels to other elements in the feature vectors generated by CNN and LSTM. This way, the model can be selective towards more informative features for better performance during fusing. Other fusion techniques can also combine CNN and LSTM features. Depending on what is necessary to accomplish the objective, these techniques include aggregation, external addition or multiplication, and others like capsule or graph neural networks. Fig. 5 depicts the flowchart of the proposed model.
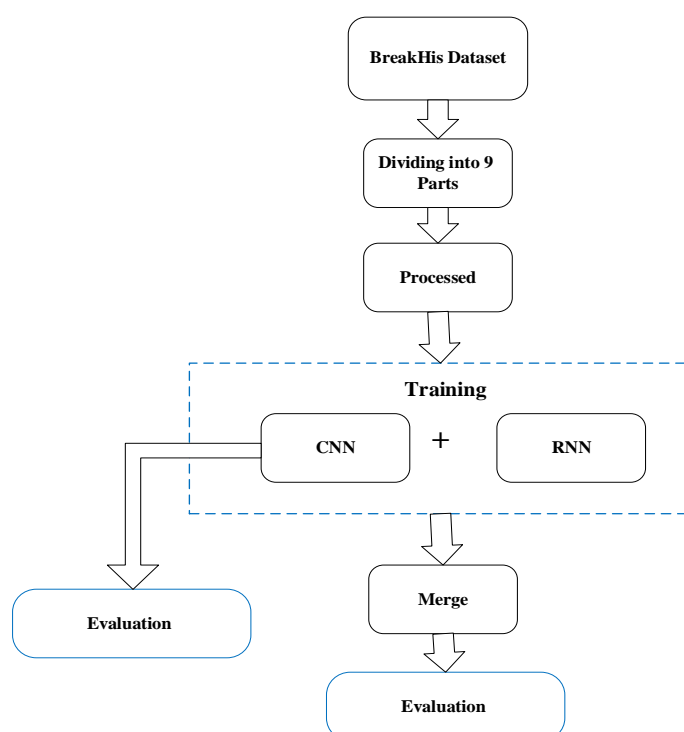

Figure 5. Flowchart of the proposed model

2.5. Resize image

To assure interoperability with the multiple networks utilized in this work, the data in the BreaKHis dataset needed to be size normalized. Data cropping and rescaling algorithms match the image sizes with the pre-trained deep neural networks. ResNet 101 utilizes 224 x 224 pixels for operation, VGG19 uses 224 x 224 pixels, and Google Net uses 224 x 224 pixels—compatibility with the particular network topologies utilized in this investigation by appropriately scaling the input images.

2.6. Transfer Learning

Transfer learning indicates the trained system's parameters' migration to the new model to help train that new architecture. In the present work, the pool layer and conv layer of the first few pre-trained ImageNet data of the CNN model have been utilized as the proposed network's base, on top of which a number of the task-specific fully-connected layers with the randomly initialized weight values have been attached [39]-[43]. To facilitate transmissions of the features, the same network layer (i.e., pooling & conv) as the CNN has been transferred to the exact location in the proposed architecture. Additionally, CNN utilizes an RGB channel as the input, and the RNN takes one channel map as the input. All those layers have been trained (i.e., tweaked) jointly on the BreaKHis data set and utilized at the original learning rate for the FC layer training.

2.7. Experimental Setup

This work employs a sub-image from BreaKHis dataset-based learning approach to train the proposed model. To conduct the tests, the BreaKHis dataset is divided into two groups: a testing set and a training set. The division was performed randomly, without replacement, with 20% of the data allocated for testing, while the remaining 80% was used for training. The data was trained and tested separately for each magnification level. Cross-validation techniques were not utilized in this work.

A dropout regularization is implemented to mitigate the risk of overfitting in the pre-trained neural networks. Dropout was applied to each network to prevent excessive reliance on specific features during training. Prior works that adopted an image-based approach have also utilized the BreaKHis dataset, including references [25]-[27] and [39]-[42].

The training process involved freezing the CNN layers and only training the final classification layer and the RNN network [15]. This training was performed using the Adaptive Moment Estimation (Adam). Afterward, all network layers were unfrozen and fine-tuned using the Adam optimizer [44] with a learning rate 0.0001. The cross-entropy loss function was utilized and optimized to fit not only the 1-hot distribution but also the uniform distribution.

The hyperparameters, including the decay factor, learning rate, and batch size, were configured as follows: the decay factor was set to 0.99. In contrast, the learning rate was set to a specific value. Additionally, the batch size was set to 128 during the training process.

3. Results

This section provides the results of tests on three DNNs trained using TL techniques. These networks were then evaluated to identify BC using histopathology images. The model was trained and tested with sub-image data as specified in Table 4. These models were built using 80% of the available data, while 20% was employed to assess their performance in binary classifications. Dividing the data ensures that the models' generalization abilities are adequately assessed and can accurately predict and separate breast cancer cases.

Self-learning is employed on sub-image data; the models will only concentrate on specific areas within the images to capture more localized features, enhancing classification accuracy.

These experiments have yielded important information about how well DNNs trained on histopathological images perform in breast cancer classification. To evaluate the model's competence at accurately recognizing and classifying breast cancer cases, metrics including accuracy, precision, recall, and F1-score are used. Breast cancer detection auto-systems can improve their classification accuracy and performance by using either sub-image data or applying transfer learning.

3.1. Evaluation Metrics

Evaluation metrics are essential for evaluating how well categorization algorithms work. In this work, several widely recognized metrics are utilized to assess the efficacy of the models, including Accuracy, Sensitivity (Recall), Specificity, Precision, ROC, and F1 score [45]-[48].

3.1.1. Accuracy

Accuracy, as shown in equation (1), measures the proportion of correctly identified samples to all samples. Greater accuracy values correspond to better model performance.

$$Accuracy = (tp + tn)/total\ samples \qquad (1)$$

3.1.2. Recall (Sensitivity)

A model's sensitivity assesses its ability to correctly identify positive samples among all true positives. Sensitivity is essential when minimizing false negatives, as shown in equation (2).

$$Sensitivity = tp/(tp + fn) \qquad (2)$$

3.1.3. Specificity

As shown in equation (3), specificity measures how well the model can identify negative samples out of all negative ones.

$$Specificity = tn/(tn + fp) \qquad (3)$$

3.1.4. Precision

Precision measures how well the model predicts the positive samples. It is helpful when one wants to minimize false positive results, as shown in equation (4).

$$Precision = tp/(tp + fp) \qquad (4)$$

3.1.5 F1 score

The F1 score is a balanced statistic combining precision and sensitivity. F1 is a measure that combines precision and recall

in one number, as shown in equation (5). This indicates an improved trade-off between accuracy and recall.

$$F1\ score = 2 * (precision * sensitivity) / (precision + sensitivity) \qquad (5)$$

3.2. Performance Evaluation of Proposed Model

This work investigates the performance of classifications by VGG19, Google Net, and Resnet101 deep neural network (DNN) classifiers. The BreaKHis images on which these classifiers have been tested had magnification factors of 40x, 100x, 200x, and 400x. The results are provided in Tables 5 and 6 and shown in Fig. 6, 7, and 8 to evaluate how these DNNs perform relative to each other. Resnet101 was the best performer and was accurate in all measures.

Table 5. Evaluation measures for the DCNN model

Type	Fact.	Accu.	Sens.	Spec.	Prec.	F1
VGG 19	40X	97.91	97.90	97.90	97.88	97.89
	100X	98.21	98.18	98.18	98.18	98.18
	200X	97.65	97.62	97.62	97.63	97.62
	400X	97.04	97.03	97.03	97.02	97.02
Google Net	40X	97.74	97.72	97.72	97.71	97.72
	100X	98.24	98.19	98.19	98.22	98.21
	200X	97.71	97.69	97.69	97.67	97.68
	400X	97.01	96.98	96.98	97	96.99
Resnet 101	40X	98.41	98.39	98.39	98.40	98.39
	100X	**98.58**	98.59	98.59	98.53	98.56
	200X	98.48	98.47	98.47	98.46	98.46
	400X	97.47	97.46	97.46	97.44	97.45

Table 6. Evaluation measures for the hybrid model

| Type	Fact.	Accu.	Sens.	Spec.	Prec.	F1
VGG 19 + LSTM	40X	98.55	98.54	98.54	98.53	98.53
	100X	98.99	98.95	98.95	98.98	98.97
	200X	98.37	98.37	98.37	98.34	98.35
	400X	98.26	98.25	98.25	98.25	98.25
Google Net + LSTM	40X	98.38	98.36	98.36	98.37	98.36
	100X	98.93	98.91	98.92	98.92	98.91
	200X	98.76	98.73	98.73	98.75	98.74
	400X	98.53	98.54	98.54	98.52	98.53
Resnet 101 + LSTM	40X	99.16	99.14	99.14	99.17	99.15
	100X	**99.76**	99.76	99.76	99.75	99.76
	200X	99.20	99.20	99.20	99.18	99.19
	400X	99.11	99.13	99.13	99.10	99.11
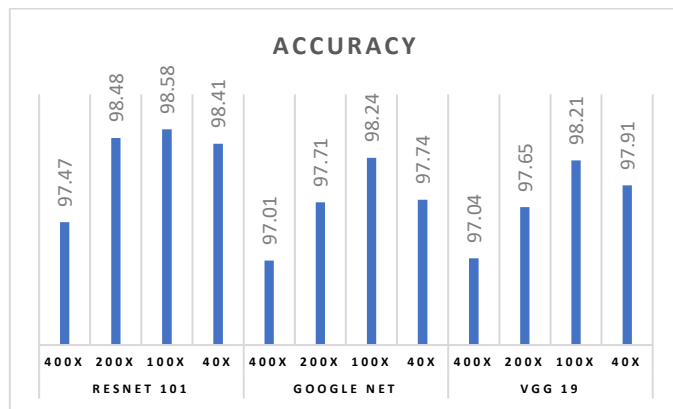
Figure 6. DNN accuracy at multiple magnifications in manufacturing
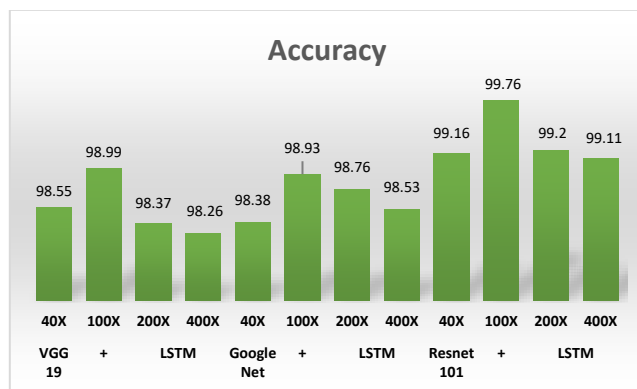

Figure 7. Hybrid accuracy at multiple magnifications in manufacturing
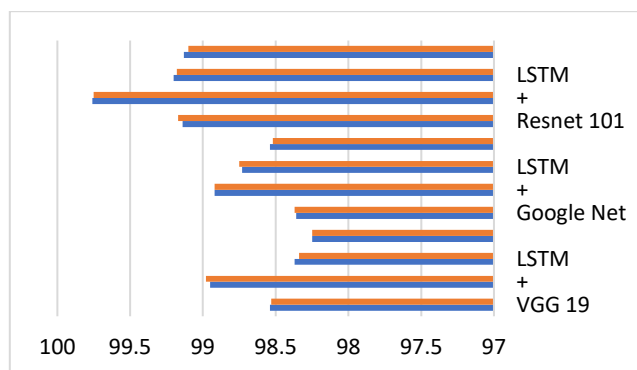

Figure 8. Precision and specificity for hybrid system

Table 7. Confusion matrices for DCNNs

Type	Fact.	TP	FP	FN	TN	A%
VGG 19	40X	1561	36	39	1955	97.91%
	100X	1593	34	33	2085	98.21%
	200X	1569	44	41	1969	97.65%
	400X	1467	48	49	1712	97.04%
Google Net	40X	1558	39	42	1952	97.74%
	100X	1592	35	31	2087	98.24%
	200X	1573	40	43	1967	97.71%
	400X	1464	51	47	1714	97.01%
Resnet 101	40X	1568	29	28	1966	98.41%
	100X	1605	22	31	2087	**98.58**%
	200X	1587	26	29	1981	98.48%
	400X	1476	39	44	1717	97.47%

Table 8. Confusion matrices for the hybrid model

Type	Fact.	TP	FP	FN	TN	A%
VGG 19 + LSTM	40X	1572	25	27	1967	98.55%
	100X	1606	21	17	2101	98.99%
	200X	1586	27	32	1978	98.37%
	400X	1487	28	29	1732	98.26%
Google Net + LSTM	40X	1567	30	28	1966	98.38%
	100X	1606	21	19	2099	98.93%
	200X	1589	24	21	1989	98.76%
	400X	1493	22	26	1735	98.53%
Resnet 101 + LSTM	40X	1579	18	12	1982	99.16%
	100X	1623	4	5	2113	**99.76**%
	200X	1600	13	16	1994	99.20%
	400X	1504	11	18	1743	99.11%

3.3. Confusion Matrices

Tables 7 and 8 are the confusion matrices for two classification models applied to the sub-image BreaKHis dataset.

These matrices give a detailed account of each model's classification results, indicating the number of true positives, false negatives, false positives, and true negatives.

The findings demonstrate that ResNet101 has attained a markedly superior accuracy rate of 99.76% compared to the other models. ROC curves, shown in Fig. 9, show the trade-off between true positive rate and false positive rate, helping to evaluate model performance across all thresholds. Precision-recall curves, shown in Fig. 10, highlight the trade-off between precision and recall for different thresholds and are helpful in imbalanced datasets. Table 9. compares the proposed hybrid CNN-LSTM with existing state-of-the-art models. Figs. 11 and 12 show the convergence loss and accuracy for the Hybrid ResNet101 + LSTM model.
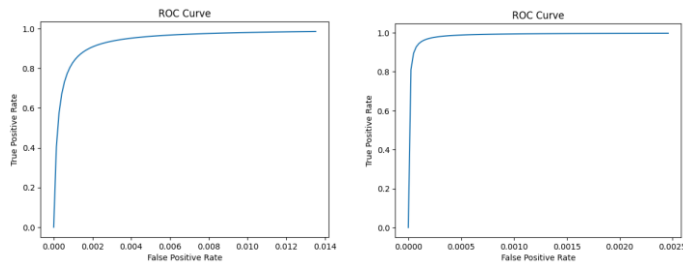
| (a) ResNet101 | (b) ResNet101+LSTM | (a) ResNet101 | (b) ResNet101+LSTM |

Figure 9. ROC curves   **Figure 10.** Precision-recall curves

Table 9. Performance comparison

| Ref.	Factors	Accuracy	Sensitivity	Specificity	Precision	F1 score
[14] 2019	40X- 400X	88.12%	86.80%	87.33%	-	-
[10] 2020	-	95.29%	95.21%	-	95.46%	95.29%
[11] 2020	40X	89.87%	86.27%	-	87.11%	86.34%
	100X	96.45%	86%	-	86.37%	85.93%
	200X	94.10%	83.11%	-	84.52%	83.08%
	400X	90.94%	80.09%	-	80.14%	80.04%
[15] 2023	40X	96.30%	95%	-	97%	-
	100X	92.60%	93%	-	92%	-
	200X	88.04%	88%	-	87%	-
	400X	92.51%	93%	-	92%	-
[25] 2024	40X	97.24%	96.69%	96.69%	96.89%	96.79%
	100X	96.88%	96.67%	96.67%	96.09%	96.37%
	200X	97.77%	97.50%	97.50%	97.29%	97.40%
	400X	98.08%	97.92%	97.92%	97.71%	97.81%
This Work 2024	40X	99.16%	99.14%	99.14%	99.17%	99.15%
	100X	99.76%	99.76%	99.76%	99.75%	99.76%
	200X	99.20%	99.20%	99.20%	99.18%	99.19%
	400X	99.11%	99.13%	99.13%	99.10%	99.11%
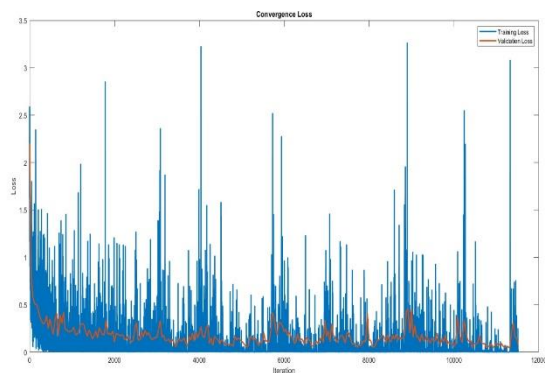
Figure 11. Convergence loss
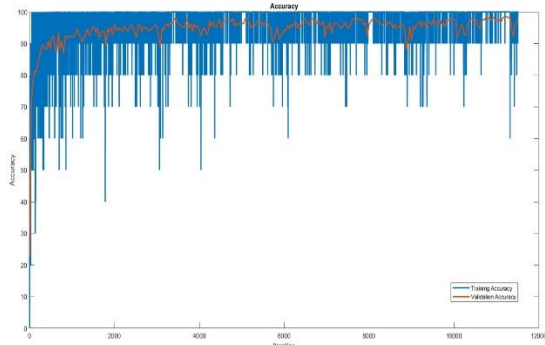
Figure 12. Accuracy

4. Conclusions

This paper presents a novel hybrid approach that combines CNN-RNN models for breast cancer classification using histopathology images. The model effectively extracts features and enhances classification accuracy by employing pre-trained CNNs such as VGG19, Google Net, and ResNet101.

The integration of CNN with RNN components allows for capturing localized features and insights, further boosting accuracy. A sub-image-based strategy and hierarchical self-learning approach were implemented to address mislabeled images, refining the dataset and facilitating robust model training. The hybrid model, achieving an accuracy of 99.76%, highlights its potential for broader tumor and disease classification applications. The paper primarily addresses binary classification, which may not capture the complexity of different breast cancer subtypes. Overall, this innovative approach marks a promising advancement in medical image classification, with significant implications for improving diagnosis and patient care.

Acknowledgment

Conflict of interest

The authors affirm no conflicts of interest concerning this manuscript's publication.

Data availability statement

This work used the BreaKHis dataset, published by Spanhol, Spanhol et al., in 2016 [24], and the sub-images BreaKHis dataset, published by Abdulaal et al. in 2024 [27].

Author Contribution Statement

Alaa Hussein Abdulaal and Morteza Valizadeh: Found the problem statement, determined the main objectives, conducted an extensive literature review, and developed the methodology.

Riyam Ali Yassin and Mehdi Chehel Amirani: Collected data for the work, designed the work, and analyzed the results.

A.F.M. Shah and Baraa M. Albaker: Developed the methodology and conducted computations and numerical experiments.

Ammar Saad Mustafa: Gave a valuable interpretation of the results.

All authors were actively engaged in debates on work findings and participated in composing the final manuscript. They collectively reviewed and revised the manuscript, incorporating feedback from each other.

References

[1] E. I. Obeagu and G. U. Obeagu, "Breast cancer: a Review of Risk Factors and Diagnosis," Medicine, vol. 103, no. 3, pp. e36905–e36905, Jan. 2024, doi: https://doi.org/10.1097/md.0000000000036905.

[2] E. G. Fischer, "Nuclear Morphology and the Biology of Cancer Cells," Acta Cytologica, vol. 64, no. 6, pp. 511–519, 2020, doi: https://doi.org/10.1159/000508780.

[3] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," CA: a Cancer Journal for Clinicians, vol. 68, no. 6, pp. 394–424, Sep. 2018, doi: https://doi.org/10.3322/caac.21492.

[4] F. A. Spanhol, L. S. Oliveira, P. R. Cavalin, C. Petitjean, and L. Heutte, "Deep Features for Breast Cancer Histopathological Image Classification," 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1868–1873, Oct. 2017, doi: https://doi.org/10.1109/smc.2017.8122889.

[5] K. Rautela, D. Kumar, and V. Kumar, "A Comprehensive Review on Computational Techniques for Breast Cancer: past, present, and Future," Multimedia Tools and Applications, pp. 1–34, Feb. 2024, doi: https://doi.org/10.1007/s11042-024-18523-2.

[6] S. Katuwal, P. Jousilahti, and E. Pukkala, "Causes of Death among Women with Breast cancer: a Follow-up Study of 50 481 Women with Breast Cancer in Finland," International Journal of Cancer, vol. 149, no. 4, pp. 839–845, May 2021, doi: https://doi.org/10.1002/ijc.33607.

[7] W. Ikezogwo et al., "Quilt-1M: One Million Image-Text Pairs for Histopathology," Advances in Neural Information Processing Systems, vol. 36, pp. 37995–38017, Dec. 2023, doi: https://doi.org/10.48550/arXiv.2306.11207.

[8] E. Balaur et al., "Colorimetric Histology Using Plasmonically Active Microscope Slides," Nature, vol. 598, no. 7879, pp. 65–71, Oct. 2021, doi: https://doi.org/10.1038/s41586-021-03835-2.

[9]   M. G. Hanna et al., "Integrating Digital Pathology into Clinical Practice," Modern Pathology, vol. 35, no. 2, pp. 152–164, Feb. 2022, doi: https://doi.org/10.1038/s41379-021-00929-0.

[10] Z. Hameed, S. Zahia, B. Garcia-Zapirain, J. Javier Aguirre, and A. María Vanegas, "Breast Cancer Histopathology Image Classification Using an Ensemble of Deep Learning Models," Sensors, vol. 20, no. 16, p. 4373, Aug. 2020, doi: https://doi.org/10.3390/s20164373.

[11] K. Gupta and N. Chawla, "Analysis of Histopathological Images for Prediction of Breast Cancer Using Traditional Classifiers with Pre-Trained CNN," Procedia Computer Science, vol. 167, pp. 878–889, 2020, doi: https://doi.org/10.1016/j.procs.2020.03.427.

[12] Y. Jiang, L. Chen, H. Zhang, and X. Xiao, "Breast Cancer Histopathological Image Classification Using Convolutional Neural Networks with Small SE-ResNet Module," PLOS ONE, vol. 14, no. 3, p. e0214587, Mar. 2019, doi: https://doi.org/10.1371/journal.pone.0214587.

[13] Q. A. Al-Haija and A. Adebanjo, "Breast Cancer Diagnosis in Histopathological Images Using ResNet-50 Convolutional Neural Network," 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Sep. 2020, doi: https://doi.org/10.1109/iemtronics51293.2020.9216455.

[14] M. Z. Alom, C. Yakopcic, Mst. S. Nasrin, T. M. Taha, and V. K. Asari, "Breast Cancer Classification from Histopathological Images with Inception Recurrent Residual Convolutional Neural Network," Journal of Digital Imaging, vol. 32, no. 4, pp. 605–617, Feb. 2019, doi: https://doi.org/10.1007/s10278-019-00182-7.

[15] M. M. Srikantamurthy, V. P. S. Rallabandi, D. B. Dudekula, S. Natarajan, and J. Park, "Classification of Benign and Malignant Subtypes of Breast Cancer Histopathology Imaging Using Hybrid CNN-LSTM Based Transfer Learning," BMC Medical Imaging, vol. 23, no. 1, Jan. 2023, doi: https://doi.org/10.1186/s12880-023-00964-0.

[16] C. Guo, B. Fan, Q. Zhang, S. Xiang, and C. Pan, "AugFPN: Improving Multi-scale Feature Learning for Object Detection," ArXiv (Cornell University), pp. 12595–12604, Dec. 2019, doi: https://doi.org/10.48550/arxiv.1912.05384.

[17] S. K. Pal, A. Pramanik, J. Maiti, and P. Mitra, "Deep Learning in multi-object Detection and tracking: State of the Art," Applied Intelligence, vol. 51, no. 9, pp. 6400–6429, Apr. 2021, doi: https://doi.org/10.1007/s10489-021-02293-7.

[18] N. Prakash, A. Manconi, and S. Loew, "Mapping Landslides on EO Data: Performance of Deep Learning Models vs. Traditional Machine Learning Models," Remote Sensing, vol. 12, no. 3, p. 346, Jan. 2020, doi: https://doi.org/10.3390/rs12030346.

[19] X. Wang, Y. Zhao, and F. Pourpanah, "Recent Advances in Deep Learning," International Journal of Machine Learning and Cybernetics, vol. 11, no. 4, pp. 747–750, Feb. 2020, doi: https://doi.org/10.1007/s13042-020-01096-5.

[20] S. Azizi, S. Kornblith, C. Saharia, M. Norouzi, and D. J. Fleet, "Synthetic Data from Diffusion Models Improves ImageNet Classification," arXiv.org, vol. 2304, no. 8466, Apr. 2023, doi: https://doi.org/10.48550/arXiv.2304.08466.

[21] M. Doostmohammadian, M. I. Qureshi, M. H. Khalesi, Rabiee, Hamid R, and U. A. Khan, "Log-Scale Quantization in Distributed First-Order Methods: Gradient-based Learning from Distributed Data," arXiv.org, vol. 2406, no. 621, 2024, doi: https://doi.org/10.48550/arXiv.2406.00621

[22] G. Liang, H. Hong, W. Xie, and L. Zheng, "Combining Convolutional Neural Network with Recursive Neural Network for Blood Cell Image Classification," IEEE Access, vol. 6, pp. 36188–36197, 2018, doi: https://doi.org/10.1109/access.2018.2846685.

[23] S. Li, Z. Song, Y. Xia, T. Yu, and T. Zhou, "The Closeness of In-Context Learning and Weight Shifting for Softmax Regression," arXiv.org, vol. 2304, no. 13276, 2023, doi: https://doi.org/10.48550/arXiv.2304.13276.

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

[25] A. H. Abdulaal, M. Valizadeh, B. M. AlBaker, R. A. Yassin, M. C. Amirani, and A. F. M. S. Shah, "Enhancing Breast Cancer Classification Using a Modified GoogLeNet Architecture with Attention Mechanism," Al-Iraqia Journal of Scientific Engineering Research, vol. 3, no. 1, Mar. 2024, doi: https://doi.org/10.58564/ijser.3.1.2024.145.

[26] S. Sharma, R. Mehra, and S. Kumar, "Optimised CNN in Conjunction with Efficient Pooling Strategy for the Multi-classification of Breast Cancer," IET Image Processing, vol. 15, no. 4, pp. 936–946, Dec. 2020, doi: https://doi.org/10.1049/ipr2.12074.

[27] A. H. Abdulaal, M. Valizadeh, M. C. Amirani, and A. F. M. Shahen Shah, "A self-learning Deep Neural Network for Classification of Breast Histopathological Images," Biomedical Signal Processing and Control, vol. 87, no. Part B, p. 105418, Jan. 2024, doi: https://doi.org/10.1016/j.bspc.2023.105418.

[28] S. Liu and W. Deng, "Very Deep Convolutional Neural Network Based Image Classification Using Small Training Sample Size," 015 3rd IAPR Asian Conference on Pattern Recognition (ACPR). IEEE, pp. 730–734, Nov. 2015, doi: https://doi.org/10.1109/ACPR.2015.7486599.

[29] S. Sukegawa et al., "Deep Neural Networks for Dental Implant System Classification," Biomolecules, vol. 10, no. 7, p. 984, Jul. 2020, doi: https://doi.org/10.3390/biom10070984.

[30] X. Wan, X. Zhang, and L. Liu, "An Improved VGG19 Transfer Learning Strip Steel Surface Defect Recognition Deep Neural Network Based on Few Samples and Imbalanced Datasets," Applied Sciences, vol. 11, no. 6, p. 2606, Mar. 2021, doi: https://doi.org/10.3390/app11062606.

[31] H. Arshad et al., "A Multilevel Paradigm for Deep Convolutional Neural Network Features Selection with an Application to Human Gait Recognition," Expert Systems, vol. 39, no. 7, p. e12541, Mar. 2020, doi: https://doi.org/10.1111/exsy.12541.

[32] J. Bai, H. Jiang, S. Li, and X. Ma, "NHL Pathological Image Classification Based on Hierarchical Local Information and GoogLeNet-Based Representations," BioMed Research International, vol. 2019, no. 1, pp. 1–13, Mar. 2019, doi: https://doi.org/10.1155/2019/1065652.

[33] J. Zhang, X. Wei, C. Che, Q. Zhang, and X. Wei, "Breast Cancer Histopathological Image Classification Based on Convolutional Neural Networks," Journal of Medical Imaging and Health Informatics, vol. 9, no. 4, pp. 735–743, May 2019, doi: https://doi.org/10.1166/jmihi.2019.2648.

[34] M. Gour, S. Jain, and T. Sunil Kumar, "Residual Learning Based CNN for Breast Cancer Histopathological Image Classification," International Journal of Imaging Systems and Technology, vol. 30, no. 3, pp. 621–635, Feb. 2020, doi: https://doi.org/10.1002/ima.22403.

[35] R. Anand, T. Shanthi, M. S. Nithish, and S. Lakshman, "Face Recognition and Classification Using GoogleNET Architecture," Advances in Intelligent Systems and Computing, vol. 1, pp. 261–269, Nov. 2019, doi: https://doi.org/10.1007/978-981-15-0035-0_20.

[36] R. Pandian, V. Vedanarayanan, D. N. S. Ravi Kumar, and R. Rajakumar, "Detection and Classification of Lung Cancer Using CNN and Google Net," Measurement: Sensors, vol. 24, p. 100588, Dec. 2022, doi: https://doi.org/10.1016/j.measen.2022.100588.

[37] M. Shafiq and Z. Gu, "Deep Residual Learning for Image Recognition: a Survey," Applied Sciences, vol. 12, no. 18, p. 8972, Sep. 2022, doi: https://doi.org/10.3390/app12188972.

[38] C. Sun, A. Shrivastava, S. Singh, and A. Gupta, "Revisiting Unreasonable Effectiveness of Data in Deep Learning Era," Proceedings of the IEEE International Conference on Computer Vision, pp. 843–852, Jul. 2017, doi: https://doi.org/10.48550/arxiv.1707.02968.

[39] Y. Gu and J. Yang, "Densely-Connected Multi-Magnification Hashing for Histopathological Image Retrieval," IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 4, pp. 1683–1691, Jul. 2019, doi: https://doi.org/10.1109/jbhi.2018.2882647.

[40] D. Sabari Nathan, R. Saravanan, J. Anbazhagan, and P. Koduganty, "Comparison of Deep Feature Classification and Fine Tuning for Breast Cancer Histopathology Image Classification," Communications in

Computer and Information Science, no. Part II, pp. 58–68, 2019, doi: https://doi.org/10.1007/978-981-13-9184-2_5.

[41] S. Sharma and R. Mehra, "Effect of layer-wise fine-tuning in magnification-dependent Classification of Breast Cancer Histopathological Image," The Visual Computer, vol. 36, no. 9, pp. 1755–1769, Oct. 2019, doi: https://doi.org/10.1007/s00371-019-01768-6.

[42] P. Xue et al., "Deep Learning in Image-based Breast and Cervical Cancer Detection: a Systematic Review and Meta-analysis," Npj Digital Medicine, vol. 5, no. 1, p. 19, Feb. 2022, doi: https://doi.org/10.1038/s41746-022-00559-z.

[43] K. Muhannad and B. M. AlBaker, "Efficient Classification Model of Pneumonia Infection Based on Deep Transfer Learning and Chest X-Ray Images," Al-Iraqia Journal of Scientific Engineering Research, vol. 1, no. 1, Sep. 2022, doi: https://doi.org/10.33193/ijser.1.1.2022.37.

[44] A. Barakat and P. Bianchi, "Convergence and Dynamical Behavior of the ADAM Algorithm for Nonconvex Stochastic Optimization," SIAM Journal on Optimization, vol. 31, no. 1, pp. 244–274, Jan. 2021, doi: https://doi.org/10.1137/19m1263443.

[45] A. H. Abdulwahhab, A. H. Abdulaal, Assad, A. A. Mohammed, and M. Valizadeh, "Detection of Epileptic Seizure Using EEG Signals Analysis Based on Deep Learning Techniques," Chaos, Solitons & fractals/Chaos, Solitons and Fractals, vol. 181, pp. 114700–114700, Apr. 2024, doi: https://doi.org/10.1016/j.chaos.2024.114700

[46] A. H. Abdulaal et al., "Unsupervised Histopathological Sub-Image Analysis for Breast Cancer Diagnosis Using Variational Autoencoders, Clustering, and Supervised Learning," Journal of Engineering and Sustainable Development, vol. 28, no. 6, pp. 729–744, Nov. 2024, doi: https://doi.org/10.31272/jeasd.28.6.6.

[47] A. H. Abdulaal et al., "Cutting-Edge CNN Approaches for Breast Histopathological Classification: the Impact of Spatial Attention Mechanisms," ShodhAI: Journal of Artificial Intelligence, vol. 1, no. 1, pp. 109–130, Oct. 2024, doi: https://doi.org/10.29121/shodhai.v1.i1.2024.14.

[48] A. H. Abdulaal et al., "Deep Learning-based Signal Identification in Wireless Communication Systems: a Comparative Analysis on 3G, LTE, and 5G Standards," Al-Iraqia Journal for Scientific Engineering Research, vol. 3, no. 3, pp. 66–70, Sep. 2024, doi: https://doi.org/10.58564/IJSER.3.3.2024.224.