









Research Article

Classification of Breast Cancer Histopathological Images Using DenseNet and Transfer Learning

Musa Adamu Wakili ¹, Harisu Abdullahi Shehu ², Md. Haidar Sharif ³,
Md. Haris Uddin Sharif ⁴, Abubakar Umar ¹, Huseyin Kusetogullari ⁵,
Ibrahim Furkan Ince ⁶, and Sahin Uyaver ⁷

¹Abubakar Tafawa Balewa University, Bauchi 740272, Nigeria

²School of Engineering and Computer Science, Victoria University of Wellington, Wellington 6012, New Zealand

³College of Computer Science and Engineering, University of Hail, Hail 2440, Saudi Arabia

⁴School of Computer & Information Sciences, University of the Cumberland, Williamsburg, KY 40769, USA

⁵Department of Computer Science, Blekinge Institute of Technology, Karlskrona 37141, Sweden

⁶Department of Digital Game Design, Nisantasi University, 34485 Istanbul, Turkey

⁷Department of Energy Science and Technologies, Turkish-German University, 34820 Istanbul, Turkey

Correspondence should be addressed to Md. Haidar Sharif; md.sharif@uoh.edu.sa

Received 9 May 2022; Revised 19 June 2022; Accepted 30 July 2022; Published 10 October 2022

Academic Editor: Mohammed A. A. Al qaness

Copyright © 2022 Musa Adamu Wakili et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Breast cancer is one of the most common invading cancers in women. Analyzing breast cancer is nontrivial and may lead to disagreements among experts. Although deep learning methods achieved an excellent performance in classification tasks including breast cancer histopathological images, the existing state-of-the-art methods are computationally expensive and may overfit due to extracting features from in-distribution images. In this paper, our contribution is mainly twofold. First, we perform a short survey on deep-learning-based models for classifying histopathological images to investigate the most popular and optimized training-testing ratios. Our findings reveal that the most popular training-testing ratio for histopathological image classification is 70%:30%, whereas the best performance (e.g., accuracy) is achieved by using the training-testing ratio of 80%:20% on an identical dataset. Second, we propose a method named DenTnet to classify breast cancer histopathological images chiefly. DenTnet utilizes the principle of transfer learning to solve the problem of extracting features from the same distribution using DenseNet as a backbone model. The proposed DenTnet method is shown to be superior in comparison to a number of leading deep learning methods in terms of detection accuracy (up to 99.28% on BreaKHis dataset deeming training-testing ratio of 80%:20%) with good generalization ability and computational speed. The limitation of existing methods including the requirement of high computation and utilization of the same feature distribution is mitigated by dint of the DenTnet.

1. Introduction

Breast cancer is one of the most familiar invasive cancers in women worldwide. Nowadays, it is overtaking lung cancer as the world's chiefly regularly diagnosed cancer [1]. The diagnosis of breast cancer in the early stages significantly decreases the mortality rate by allowing the choice of adequate treatment. With the onset of pattern recognition and machine learning, a good deal of handcrafted or engineered

features-based studies have been proposed for classifying breast cancer histology images. In image classification, feature extraction is a cardinal process used to maximize the classification accuracy by minimizing the number of selected features [2–5]. Deep learning models have the power to automatically extract features, retrieve information, and take in the latest intellectual depictions of data. Thus, they can solve the problems of common feature extraction methods. The automated classification of breast cancer

histopathological images is one of the important tasks in CAD (Computer-Aided Detection/Diagnosis) systems, and deep learning models play a remarkable role by detecting, classifying, and segmenting prime breast cancer histopathological images. Many researchers worldwide have invested appreciable efforts in developing robust computer-aided tools for the classification of breast cancer histopathological images using deep learning. At present, in this research arena, the most popular deep learning models proposed in the literature are based on CNNs [6–66].

A pretrained CNN model, for example, DenseNet [67], utilizes dense connection between layers, reduces the number of parameters, strengthens propagation, and encourages feature reutilization. This improved parameter efficiency makes the network faster and easier to train. Nevertheless, a DenseNet [67] has an excessive connection, as all its layers have a direct connection to each other. Those lavish connections have been shown to decrease the computational and parameter efficiency of the network. In addition, features extracted by a neural network model stay in the same distribution. Therefore, the model might overfit as the features cannot be guaranteed to be sufficient enough. Besides, a CNN-training task demands a large number of training samples; otherwise, it leads to overfitting and reduces generalization ability. However, it is arduous to secure labeled breast cancer histopathological images, which severely limits the classification ability of CNN [27].

On the other hand, the use of transfer learning can expand prior knowledge about data by including information from a different domain to target future data [68]. Consequently, it is a good idea to extract data from a related domain and then transfer those extracted data to the target domain. This way, resources can be saved and the efficiency of the model can be improved during training. A great number of breast cancer diagnosis methods based on transfer learning have been proposed and implemented by distinct researchers (e.g., [57–66]) to achieve state-of-the-art performance (e.g., ACC, AUC, PRS, RES, and FIS) on different datasets. Yet, the limitations of such performance indices, algorithmic assumptions, and computational complexities are indicating a further development of smart algorithms.

In this paper, we aim to propose a novel neural-network-based approach called DenTnet (see Figure 1) for classifying breast cancer histopathological images by taking the benefits of both DenseNet [67] and transfer learning [68]. To address the cross-domain learning problems, we employ the principle of transfer learning for transferring information from a related domain to the target domain. Our proposed DenTnet is anticipated to increase the accuracy of breast cancer histopathological images classification and accelerate the learning process. The DenTnet demonstrates better performance over its alternative CNN and/or transfer-learning-based methods (e.g., see Table 1) on the same dataset as well as training-testing ratio.

To find the best performance scores of deep learning models for classifying histopathological images, contrasting training-testing ratios were applied for divergent models on the same dataset. What would be the most popular and/or

optimized training-testing ratios to classify histopathological images considering existing state-of-the-art deep learning models? There exist many surveys enriched to sufficient contemporary methods and materials with systematic deep discussion of automatic classification of breast cancer histopathological images [68–72]. Nevertheless, to the best of our knowledge, the direct or indirect indication of this question was not reported in any of the previous studies. Henceforth, we perform a succinct survey to investigate this question. Our findings include that the most popular training-testing ratio for histopathological image classification is 70%:30%, whereas the best performance (accuracy) is achieved by using the training-testing ratio of 80%:20% on the identical dataset.

In summary, the main contributions of this context are as follows:

- (i) Determine the most popular and/or optimized training-testing ratios for classifying histopathological images using the existing state-of-the-art deep learning models.
- (ii) Propose a novel approach named DenTnet that amalgamates both DenseNet [67] and transfer learning technique to classify breast cancer histopathological images. DenTnet is anticipated to achieve high accuracy and fasten the learning process due to its utilization of dense connections from its backbone architecture (i.e., DenseNet [67]).
- (iii) Determine the generalization ability of DenTnet and the superiority measure considering nonparametric statistical tests.

The rest of the paper is organized as follows: Section 2 hints some preliminaries; Section 3 surveys briefly the existing deep models for histopathological image classification and reports our findings; Section 4 depicts the architecture of our proposed DenTnet and its implementation details; Section 5 demonstrates the experimental results and comparison on BreaKHis dataset [33]; Section 6 evaluates the generalization ability of DenTnet; Section 7 discusses nonparametric statistical tests, their reported results, and reasons for superiority along with few hints of further study; and Section 8 concludes the paper.

2. Preliminaries

Breast cancer is one of the oldest known kinds of cancer first found in Egypt [73]. It is caused by the uncontrolled growth and division of cells in the breast, whereby a mass of tissue called a tumor is created. Nowadays, it is one of the most terrifying cancers in women worldwide. For example, in 2020, there were 2.3 million women diagnosed with breast cancer and 685000 deaths globally [74]. Early detection of breast cancer can save many lives. Breast cancer can be diagnosed in view of histology and radiology images. The radiology images analysis can help to identify the areas, where the abnormality is located. However, they cannot be used to determine whether the area is cancerous [75]. On the other hand, a biopsy is an examination of tissue removed

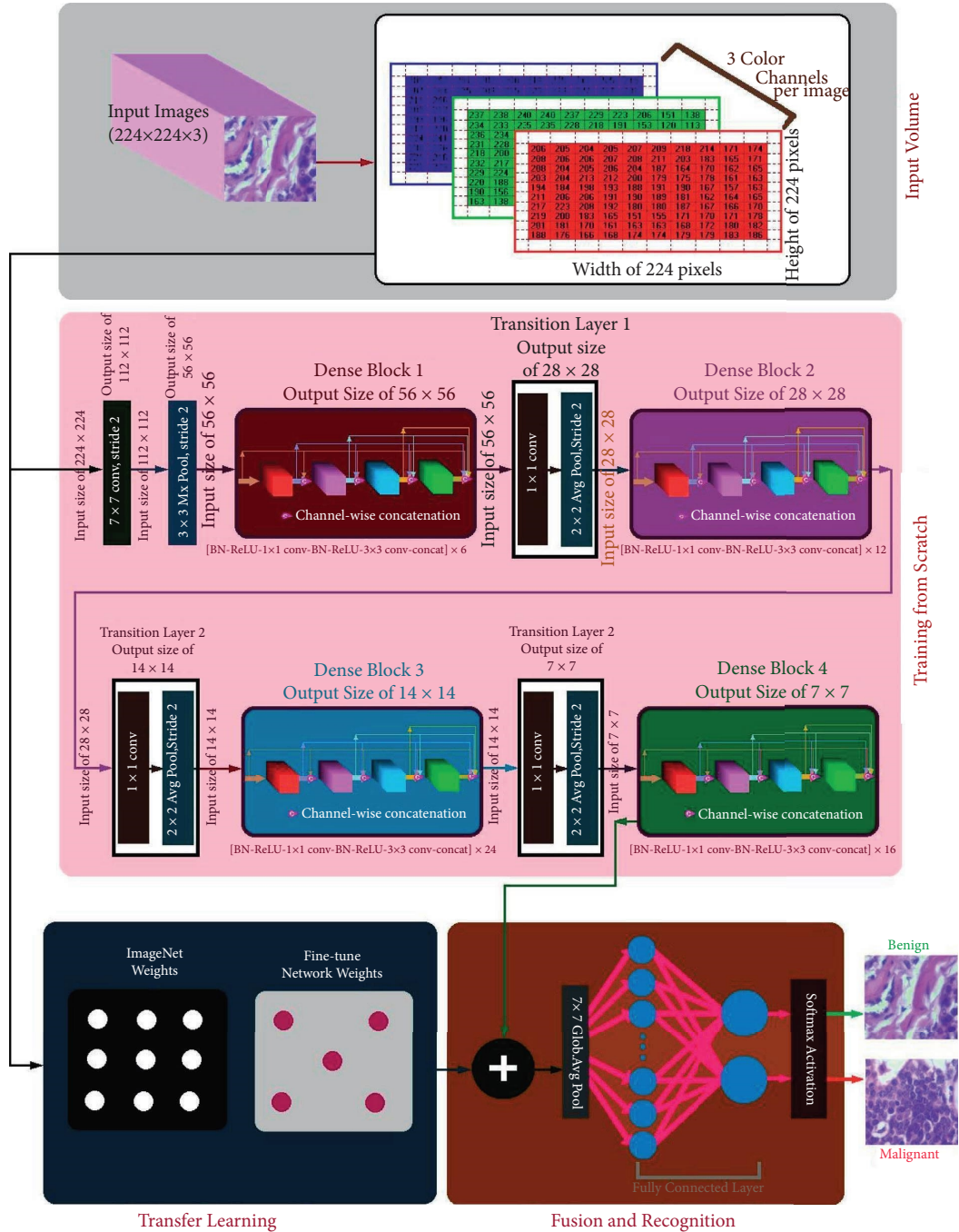


FIGURE 1: Architecture of the proposed DenTnet.

TABLE 1: Comparison of results of various methods using training-testing ratio of 80%: 20% on BreakHis [33]. The best result is shown in bold.

Year	Method	PRS	RES	F1S	AUC	ACC (%)
2020	Togacar et al. [26]	—	—	—	—	97.56
	Parvin et al. [31]	—	—	—	—	91.25
	Man et al. [36]	—	—	—	—	91.44
2021	Boumaraf et al. [63]	—	—	—	—	92.15
	Soumik et al. [60]	—	—	—	—	98.97
2022	Liu et al. [172]	—	—	—	—	96.97
	Zerouaoui and Idri [56]	—	—	—	—	93.85
	Chattopadhyay et al. [174]	—	—	—	—	96.10
	DenTnet [ours]	0.9700	0.9896	0.9948	0.99	99.28

from a living body to discover the presence, cause, or extent of a disease (e.g., cancer). Biopsy is the only reliable way to make sure if an area is cancerous [76]. Upon completion of the biopsy, the diagnosis will be based on the qualification of the histopathologists who determine cancerous regions and malignancy degree [7, 75]. If the histopathologists are not well trained, the histopathology or biopsy report can lead to an incorrect diagnosis. Besides, there might be a lack of specialists, which may cause keeping the tissue samples for up to a few months. In addition, diagnoses made by unspecialized histopathologists are sometimes difficult to replicate. As if that were not enough of a problem, at times, even expert histopathologists tend to disagree with each other. Despite notable progress being reached by diagnostic imaging technologies, the final breast cancer grading and staging are still done by pathologists using visual inspection of histological samples under microscopes.

As analyzing breast cancer is nontrivial and would get down to disagreements among experts, computerized and interdisciplinary systems can improve the accuracy of diagnostic results by reducing the processing time. The CAD can help to assist doctors in reading and interpreting medical images by locating and identifying possible abnormalities in the image [69]. It is proclaimed that the utilization of CAD to automatically classify histopathological images does not only improve the diagnostic efficiency with low cost but also provide doctors with more objective and accurate diagnosis results [77]. Consequently, there is an adamant demand for the CAD [78]. There exist several comprehensive surveys for CAD based methods in the literature. For example, Zebari et al. [71] provided a common description and analysis of existing CAD systems that are utilized in both machine learning and deep learning methods as well as their current state based on mammogram image modalities and classification methods. However, the existing breast cancer diagnosis models take issue with complexity, cost, human-dependency, and inaccuracy [73]. Furthermore, the limitation of datasets is another practical problem in this arena of research. In addition, every deep learning model demands a metric to judge its performance. Explicitly, performance evaluation metrics are the part and parcel of every deep learning model as they indicate progress indices.

In the two following subsections, we discuss the commonly used datasets for classifying histopathological images and the performance evaluation metrics of various deep learning models.

2.1. Brief Description of Datasets. Accessing relevant images and datasets is one of the key challenges for image analysis researchers. Datasets and benchmarks enable validating and comparing methods for developing smarter algorithms. Recently, several datasets of breast cancer histopathology images have been released for this purpose. Figure 2 shows a sample breast cancer histopathological image from BreKHis [33] dataset of a patient who suffered from papillary carcinoma (malignant) with four magnification levels: (a) 40x, (b) 100x (c) 200x, and (d) 400x [79]. The following list of

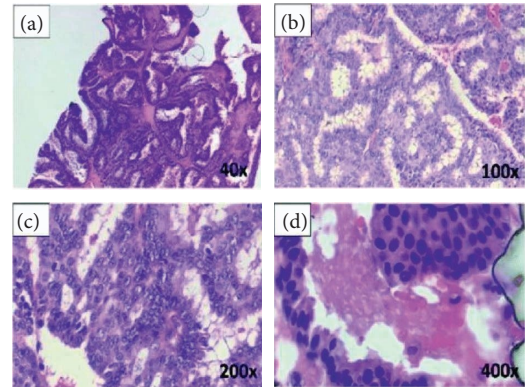


FIGURE 2: A sample breast cancer histopathological image [79] with four magnification levels of (a) 40x, (b) 100x, (c) 200x, and (d) 400x.

datasets has been used in the literature as incorporated in Table 2:

- (i) BreKHis [33] \Rightarrow It is considered as the most popular and clinical valued public breast cancer histopathological dataset. It consists of 7909 breast cancer histopathology images, 2480 benign and 5429 malignant samples, from 82 patients with different magnification factors (e.g., 40x, 100x, 200x, and 400x) [33].
- (ii) Bioimaging2015 [122] \Rightarrow The Bioimaging2015 [122] dataset contained 249 microscopy training images and 36 microscopy testing images in total, equally distributed among the four classes.
- (iii) ICIAR2018 [78] \Rightarrow This dataset, available as part of the BACH grand challenge [78], was an extended version of the Bioimaging2015 dataset [8, 122]. It contained 100 images in each of four categories (i.e., normal, benign, in situ carcinoma, and invasive carcinoma) [8].
- (iv) BACH [78] \Rightarrow The database of BACH holds images obtained from ICIAR2018 Grand Challenge [78]. It consists of 400 images with equal distribution of normal (100), benign (100), in situ carcinoma (100), and invasive carcinoma (100). The high-resolution images are digitized with the same conditions and magnification factor of 200x. In this dataset, images have a fixed size of 2048×1536 pixels [175].
- (v) TMA [99] \Rightarrow The TMA (Tissue MicroArray) database from Stanford University is a public resource with an access to 205161 images. All the whole-slide images have been scanned by a 20x magnification factor for the tissue and 40x for the cells [176].
- (vi) Camelyon [97] \Rightarrow The Camelyon (cancer metastases in lymph nodes) was established based on a research challenge dataset competition in 2016. The Camelyon organizers trained CNNs on smaller datasets for classifying breast cancer in lymph nodes and prostate cancer biopsies. The

TABLE 2: A succinct survey of deep-learning-based histopathological image classification methods. NA indicates either “not available” or “no answer” from the associated authors.

Year	Ref	Aim	Technique	Dataset	Sample	Training (%)	Testing (%)	Result	Performance AUC ACC
2016	Chan and Tuszynski [80]	To predict tumor malignancy in breast cancer	Employed binarization, fractal dimension, SVM	BreakKHis [33]	7909	50	50	ACC of 97.90%, 16.50%, 16.50%, 16.50%, and 25.30% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 39.05%
	Spanhol et al. [33]	To classify histopathological images	Employed CNN based on AlexNet [81]	BreakKHis [33]	7909	70	30	ACC of 90.0%, 88.4%, 84.6%, and 86.1% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively For single-task CNN, ACC of 83.08%, 83.17%, 84.63%, and 82.10%, obtained for 40x, 100x, 200x, and 400x magnification factors, respectively; accordingly, for multitask CNN, ACC of 81.87%, 83.39%, 82.56%, and 80.69%	NA 87.28%
	Bayram-oglu et al. [38]	To classify breast cancer histopathology images	Employed single-task CNN and multitask CNN	BreakKHis [33]	7909	70	30		NA 82.69%
	Abbas [77]	To diagnose breast masses	Applied SURF [82], LBPV [83]	DDSM [84], MIAS [85]	600	40	60	Overall 92%, 84.20%, 91.50%, and 0.91 obtained for sensitivity, specificity, ACC, and AUC, respectively	0.91 91.50%
	Song et al. [21]	To classify histopathology images	Employed a model of CNN, Fisher vector [86], SVM	BreakKHis [33], IICBU2008 [87]	8283	70	30	ACC of 94.42%, 89.49%, 87.25%, and 85.62% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 89.19%
	Wei et al. [22]	To analyze tissue images	Employed a modification of GoogleNet [88]	BreakKHis [33]	7909	75	25	ACC of 97.46%, 97.43%, 97.73%, and 97.74% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 97.59%
	Das et al. [23]	To classify histopathology images	Employed GoogleNet [88]	BreakKHis [33]	7909	80	20	ACC of 94.82%, 94.38%, 94.67%, and 93.49% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 94.34%
	Kahya et al. [89]	To identify features of breast cancer	Employed dimensionality reduction, adaptive sparse SVM	BreakKHis [33]	7909	70	30	ACC of 94.97%, 93.62%, 94.54%, and 94.42% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 94.38%
	Song et al. [24]	To classify histopathology images easily	Employed CNN-based Fisher vector [86], SVM	BreakKHis [33]	7909	70	30	ACC of 90.02%, 88.90%, 86.90%, and 86.30% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 88.03%
	Gupta and Bhavsar [90]	To classify histopathology images	Employed an integrated model	BreakKHis [33]	7909	70	30	Average ACC of 88.09% and 88.40% obtained for image and patient levels, respectively	NA 88.25%
2017	Dhungel et al. [91]	To analyze masses in mammograms	Applied multiscale deep belief nets	INbreast [92]	410	60	20	The best results on the testing set with an ACC got 95% on manual and 91% on the minimal user intervention setup ACC of 84.30%, 84.35%, 85.25% and 82.10% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	0.76 91.03%
	Spanhol et al. [34]	To classify breast cancer images	Using deep CNN	BreakKHis [33]	7900	70	30	ACC of 95.80%, 96.90%, 96.70%, and 94.9% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 96.08%
	Han et al. [35]	To study breast cancer multiclassification	Employed class structure based CNN	BreakKHis [33]	7909	50	50	ACC of 85.75%, 87.03%, and 84.18% obtained for GoogleNet [88], ResNet-50 [75], and CaffeNet [93], respectively	NA 85.65%
	Sun and Binder [39]	To assess performance of H&E stain dat.	A comparative study among ResNet-50 [75], CaffeNet [93], and GoogleNet [88]	BreakKHis [33]	7909	70	30	Overall ACC of 59.0% and 70.4% got from Back-Propagation [95] and Radial Basis [96], respectively An AUC of 97.60 (93.60, 100) obtained on par with Camelyon16 [97] test set performance	NA 64.70%
	Kaymak et al. [94]	To organize breast cancer images	Back-Propagation [95] and Radial Basis Neural Networks [96]	176 images from a hospital	176	65	35		0.97 95.00%
	Liu et al. [47]	To detect cancer metastases in images	Employed a CNN architecture	Camelyon16 [97]	110	68	32	ACC of 91.28%, 91.45%, 88.57%, and 84.58% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 88.97%
	Zhi et al. [57]	To diagnose breast cancer images	Employed a variation of VGGNet [98]	BreakKHis [33]	7909	80	20		NA 88.97%
	Chang et al. [58]	To solve the limited amount of training data	Employed CNN model from Inception [88] family (e.g., Inception V3)	BreakKHis [33]	4017	70	30	ACC of 83.00% for benign class and 89.00% for malignant class. AUC of malignant was 93.00% and AUC of benign was also 93.00%	0.93 86.00%

TABLE 2: Continued.

Year	Ref	Aim	Technique	Dataset	Sample	Training (%)	Testing (%)	Result	Performance AUC ACC
	Jannesari et al. [6]	To classify breast cancer images	Employed variations of Inception [88], ResNet [75]	BreakHis [33], 6402 images from TMA [99]	14311	85	15	With ResNets ACC of 99.80%, 98.70%, 94.80%, and 96.40% obtained for four cancer types. Inception V2 with fine-tuning all layers got ACC of 94.10%	0.99 96.34%
	Bardou et al. [7]	To classify breast cancer based on histology images	Employed CNN topology, data augmentation	BreakHis [33]	7909	70	30	ACC of 98.33%, 97.12%, 97.85%, and 96.15% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 97.36%
	Kumar and Rao [9]	To train CNN for using image classification	Employed CNN topology	BreakHis [33]	7909	70	30	ACC of 85.52%, 83.60%, 84.84%, and 82.67% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 84.16%
	Das et al. [11]	To classify breast histopathology images	Employed variation of CNN model	BreakHis [33]	7909	80	20	ACC of 89.52%, 89.06%, 88.84%, and 87.67% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 88.77%
	Nahid et al. [100]	To classify biomedical breast cancer images	Employed Boltzmann machine [101], Tamura et al. features [102]	BreakHis [33]	7909	70	30	ACC of 88.70%, 85.30%, 88.60%, and 88.40% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 87.75%
	Badejo et al. [103]	To classify medical images	Employed local phase quantization, SVM	BreakHis [33]	7909	70	30	ACC of 91.10%, 90.70%, 86.20%, and 84.30% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 88.08%
	Alireza-zadeh et al. [104]	To arrange breast cancer images	Threshold adjacency [105], quadratic analysis [106]	BreakHis [33]	7909	70	30	ACC of 89.16%, 87.38%, 88.46%, and 86.68% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 87.92%
	Du et al. [13]	To distribute breast cancer images	Employed AlexNet [81]	BreakHis [33]	7909	70	30	ACC of 90.69%, 90.46%, 90.64%, and 90.96% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 90.69%
	Gandom-kar et al. [14]	To model CNN for breast cancer image diagnosis	Employed a variation of ResNet [75] (e.g., ResNet152)	BreakHis [33]	7786	70	30	ACC of 98.60%, 97.90%, 98.30%, and 97.60% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 98.10%
2018	Gupta and Bhavsar [15]	To model CNN for breast cancer image diagnosis	Employed DenseNet [67], XGBoost classifier [107]	BreakHis [33]	7909	70	30	ACC of 94.71%, 95.92%, 96.76%, and 89.11% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 94.12%
	Ben-hammou et al. [17]	To study CNN for breast cancer images	Employed Inception V3 [88] module	BreakHis [33]	7909	70	30	ACC of 87.05%, 82.80%, 85.75%, and 82.70% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 84.58%
	Morillo et al. [108]	To label breast cancer images	Employed KAZE features [109]	BreakHis [33]	7909	70	30	ACC of 86.15%, 80.70%, 77.95%, and 72.00% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 97.20%
	Chattoraj and Vishwakarma [110]	To study breast carcinoma images	Zernike moments [111], entropies of Renyi [112], Yager [113]	BreakHis [33]	7909	70	30	ACC of 87.7%, 85.8%, 88.0%, and 84.6% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 96.53%
	Sharma and Mehra [19]	To analyze behavior of magnification independent breast cancer	Employed models of VGGNet [98] and ResNet [75] (e.g., VGG16, VGG19, and ResNet50)	BreakHis [33]	7909	90	10	Pretrained VGG16 with logistic regression classifier showed the best performance with 92.60% ACC, 95.65% AUC, and 95.95% ACC precision score for 90%–10% training-testing data splitting	0.95 94.28%
	Zheng et al. [114]	To study content-based image retrieval	Employed binarization encoding, Hamming distance [115]	BreakHis [33] and others	16309	70	30	ACC of 47.00%, 40.00%, 40.00%, and 37.00% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 41.00%
	Cascianelli et al. [20]	To study features extraction from images	Employed dimensionality reduction using CNN	BreakHis [33]	7909	75	25	ACC of 84.00%, 88.20%, 87.00%, and 80.30% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 84.88%
	Mukkamala et al. [116]	To study deep model for feature extraction	Employed PCANet [117]	BreakHis [33]	7909	80	20	ACC of 96.12%, 97.41%, 90.99%, and 85.85% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 92.59%
	Mahraban Nejad et al. [51]	To retrieve breast cancer images	Employed a variation of VGGNet [98], SVM	BreakHis [33]	7909	98	02	An average ACC of 80.00% was demonstrated from BreakHis [33]	NA 80.00%
	Rakhlil et al. [118]	To analyze breast cancer images	Several deep neural networks and gradient boosted trees classifier	BACH [78]	400	75	25	For 4-class classification task ACC was 87.2% but for 2-class classification ACC was reported to be 93.8%	0.97 90.50%
	Almasni et al. [119]	To detect breast masses	Applied regional deep learning technique	DDSM [84]	600	80	20	Distinguished between benign and malignant lesions with an overall ACC of 97%	0.96 97.00%

TABLE 2: Continued.

Year	Ref	Aim	Technique	Dataset	Sample	Training (%)	Testing (%)	Result	Performance AUC ACC
	Kassani et al. [8]	To use deep learning for binary classification of breast histology images	Usage of VGG19 [98], MobileNet [120], and DenseNet [67]	BreKHis [33], ICIAR2018 [78], PCam [121], Bioimaging2015 [122]	8594	87	13	Multimodal method got better predictions than single classifiers and other algorithms with ACC of 98.13%, 95.00%, 94.64% and 83.10% obtained for BreKHis [33], ICIAR2018 [78], PCam [121], and Bioimaging2015 [122], respectively	NA 92.72%
	Alom et al. [10]	To classify breast cancer from histopathological images	Inception recurrent residual CNN	BreKHis [33], Bioimaging2015 [122]	8158	70	30	From BreKHis [33], ACC of 97.90%, 97.50%, 97.30%, and 97.40%, obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	0.98 97.53%
	Nahid and Kong [12]	To classify histopathological breast images	Employed RGB histogram [123] with CNN	BreKHis [33]	7909	85	15	ACC of 95.00%, 96.60%, 93.500%, and 94.20% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 94.68%
	Jiang et al. [16]	To use CNN for breast cancer histopathological images	Employed CNN, Squeeze-and-Excitation [124] based ResNet [75]	BreKHis [33]	7909	70	30	The achieved accuracy between 98.87% and 99.34% for the binary classification as well as between 90.66% and 93.81% for the multiclass classification	0.99 95.67%
	Sudharshan et al. [18]	To use instance learning for image sorting	Employed CNN-based multiple instance learning algorithm	BreKHis [33]	7909	70	30	ACC of 86.59%, 84.98%, 83.47%, and 82.79% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 84.46%
2019	Gupta and Bhavsar [25]	To segment breast cancer images	Employed ResNet [75] for multilayer feature extraction	BreKHis [33]	7909	70	30	ACC of 88.37%, 90.29%, 90.54%, and 86.11% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 88.82%
	Vo et al. [125]	To extract visual features from training images	Combined weak classifiers into a stronger classifier	BreKHis [33], Bioimaging2015 [122]	8194	87	13	ACC of 95.10%, 96.30%, 96.90%, and 93.80% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 95.56%
	Qi et al. [32]	To organize breast cancer images	Employed a CNN network to complete the classification task	BreKHis [33]	7909	70	30	ACC of 91.48%, 92.20%, 93.01%, and 92.58% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 92.32%
	Talo [41]	To detect and classify diseases in images	DenseNet [67], ResNet [75] (e.g., DenseNet161, ResNet50)	KimiaPath24 [126]	25241	80	20	DenseNet161 pretrained and ResNet50 achieved ACC of 97.89% and 98.87% on grayscale and color images, respectively	NA 98.38%
	Li et al. [127]	To detect invading component in cancer images	Convolutional autoencoder-based contrast pattern mining framework	361 samples of the breast cancer	361	90	10	ACC was taken into account. The overall ACC achieved was 76.00%, whereas 77.70% was presented for FIS	NA 76.00%
	Ragab et al. [44]	To detect breast cancer from images	AlexNet [81] and SVM	DDSM [84], CBIS-DDSM [128]	2781	70	30	The deep CNN presented an ACC of 73.6%, whereas the SVM demonstrated an ACC of 87.2%	0.88 73.60%
	Romero et al. [45]	To study cancer images	A modification of Inception module [88]	HASHI [129]	151465	63	37	From deep learning networks, an overall ACC of 89.00% was demonstrated along with FIS of 90.00%	0.96 89.00%
	Minh et al. [46]	To diagnose breast cancer images	A modification of ResNet [75] and InceptionV3 [88]	BACH [78]	400	70	20	ACC with 95% for 4 types of cancer classes and ACC with 97.5% for two combined groups of cancer	0.97 96.25%

TABLE 2: Continued.

Year	Ref	Aim	Technique	Dataset	Sample	Training (%)	Testing (%)	Result	Performance AUC ACC
	Stamitsas et al. [130]	To visualize a health system for clinicians	Employed region covariance SVM, multiple instance learning [132]	FABCD [133], BreakHis [33]	7949	70	15	ACC of 91.27% and 92.00% at the patient and image level, respectively	0.98 91.64%
	Togacar et al. [26]	To analyze breast cancer images rapidly	Employed a ResNet [75] architecture with attention modules	BreakHis [33]	7909	80	20	ACC of 97.99%, 97.84%, 98.51%, and 95.88% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 97.56%
	Asare et al. [134]	To study breast cancer images	Employed self-training and self-paced learning	BreakHis [33]	7909	70	30	ACC of 93.58%, 91.04%, 93.38%, and 91.00% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 92.25%
	Gour et al. [28]	To diagnose breast cancer tumors images	Employed a modification of ResNet [75]	BreakHis [33]	7909	70	30	ACC of 90.69%, 91.12%, 95.36%, and 90.24% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	0.91 92.52%
	Li et al. [29]	To grade pathological images	Employed a modification of Xception network [135]	BreakHis [33], VLAD [136], LSC [137]	8583	60	40	ACC of 95.13%, 95.21%, 94.09%, and 91.42% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 93.96%
	Feng et al. [138]	To allocate breast cancer images	Deep neural-network-based manifold preserving autoencoder [139]	BreakHis [33]	7909	70	30	ACC of 90.12%, 88.89%, 91.57%, and 90.25% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 90.53%
	Parvin and Mehedi Hasan [31]	To study CNN models for cancer images	LeNet [140], AlexNet [81], VGGNet [98], ResNet [75], Inception V3 [88]	BreakHis [33]	7909	80	20	ACC of 89.00%, 92.00%, 94.00% and 90.00% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	0.85 91.25%
2020	Carvalho et al. [141]	To classify histological breast images	Entropies of Shannon [142], Renyi [112], Tsallis [143]	BreakHis [33]	4960	70	30	ACC of 95.40%, 94.70%, 97.60%, and 95.50% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	0.99 95.80%
	Li et al. [144]	To analyze breast cancer images	Employed global covariance pooling module [145]	BreakHis [33]	7909	70	30	ACC of 96.00%, 96.16%, 98.01%, and 95.97% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 94.93%
	Man et al. [36]	To classify cancer images	Usage of generative adversarial networks, DenseNet [67]	BreakHis [33]	7909	80	20	ACC of 97.72%, 96.19%, 86.66%, and 85.18% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 91.44%
	Kumar et al. [37]	To classify human breast cancer and canine mammary tumors	Employed a framework based on a variant of VGGNet [98] (e.g., VGGNet16) and SVM	BreakHis [33] and CMTHis [37]	8261	70	30	For BreakHis [33], ACC of 95.94%, 96.22%, 98.15%, and 94.41% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively; the same for CMTHis [37], ACC of 94.54%, 97.22%, 92.07%, and 82.84% obtained ACC was taken into account. Estimation of the standard error of mean was approximately 0.81	0.95 96.93%
	Kaushal and Singla [40]	To detect cancerous cells in images.	Employed a CNN model of self-training and self-paced learning	Total 50 images of various patients	50	90	10		NA 93.10%
	Hameed et al. [43]	To use deep learning for classification of breast cancer images	Variants of VGGNet [98] (e.g., fully trained VGG16, fine-tuned VGG16, fully trained VGG19, and fine-tuned VGG19 models)	Breast cancer images: 675 for training and 170 for testing	845	80	20	The ensemble of fine-tuned VGG16 and VGG19 models offered sensitivity of 97.73% for carcinoma class and overall accuracy of 95.29%. It also offered an F1 score of 95.29%	NA 95.29%
	Alantari et al. [48]	To detect breast lesions in digital X-ray mammograms	Adopted three deep CNN models	INbreast [92], DDSM [84]	1010	70	20	In INbreast [92] mean ACC of 89%, 93%, and 95% for CNN, ResNet50, and Inception-ResNet V2, respectively; 95%, 96%, and 98% for DDSM [146]	0.96 94.08%
	Zhang et al. [49]	To classify breast mass	ResNet [75], DenseNet [67], VGGNet [98]	CBIS-DDSM [128], INbreast [92]	3168	70	30	Overall ACC of 90.91% and 87.93% obtained from CBIS-DDSM [128] and INbreast [92], respectively	0.96 89.42%
	Hassan et al. [59]	To classify breast cancer masses	Modification of AlexNet [22] and GoogLeNet [88]	CBIS-DDSM [128], MIAS [85], INbreast [92], etc	600	75	17	With CBIS-DDSM [128] and INbreast [92] databases, the modified GoogLeNet achieved ACC of 98.46% and 92.5%, respectively	0.97 96.98%

TABLE 2: Continued.

Year	Ref	Aim	Technique	Dataset	Sample	Training (%)	Testing (%)	Result	Performance AUC ACC
	Li et al. [147]	To use high-resolution info of images	Multiview attention-guided multiple instance detection network	BreakHis [33], BACH [78], PUJH [148]	12329	70	30	Overall ACC of 94.87%, 91.32%, and 90.45% obtained from BreakHis [33], BACH [78], and PUJH [148], respectively	0.99 92.21%
	Wang et al. [27]	To divide breast cancer images	Employed a model of CNN and CapsNet [149]	BreakHis [33]	7909	70	30	ACC of 92.71%, 94.52%, 94.03%, and 93.54% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 93.70%
	Albhashish et al. [30]	To analyze VGG16 [98]	Employed a variation of VGGNet [98]	BreakHis [33]	7909	90	10	ACC of 96%, 95.10%, and 87% obtained for polynomial SVM, Radial Basis SVM, and k-nearest neighbors, respectively	NA 92.70%
	Kundale et al. [150]	To classify breast cancer from histology images	Employed SURF [82], DSIFT [151], linear coding [152]	BreakHis [33]	7909	70	30	ACC of 93.35%, 93.86%, 93.73%, and 94.00% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 93.74%
	Attallah et al. [153]	To classify breast cancer from histopathological images	Employed several deep learning techniques including autoencoder [139]	BreakHis [33], ICIAR2018 [78]	7909	70	30	For BreakHis [33], ACC of 99.03%, 99.53%, 98.08%, and 97.56% got for 40x, 100x, 200x, and 400x magnification factors, respectively; for ICIAR2018 [78], ACC was 97.93%	NA 98.43%
	Burçak et al. [154]	To classify breast cancer histopathological images	Stochastic [155], Nesterov [156], Adaptive [157], RMSprop [158], AdaDelta [159], Adam [160]	BreakHis [33]	7909	70	30	ACC was taken into account. The overall ACC of 97.00%, 97.00%, 96.00%, and 96.00% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 96.50%
2021	Hirra et al. [161]	To label breast cancer images	Patch-based deep belief network [162]	HASHI [129]	584	52	30	Images from four different data samples achieved an accuracy of 86%	NA 86.00%
	Elmannat et al. [42]	To extract eminent breast cancer image features	A combination of two deep CNNs	BACH [78]	400	60	20	The overall ACC for the subimage classification was 97.29% and for the carcinoma cases the sensitivity achieved was 99.58%	NA 97.29%
	Baker and Abu Qutaish [163]	To segment breast cancer images	Clustering and global thresholding methods	BACH [78]	400	70	30	The maximum ACC obtained from classifiers and neural network using BACH [78] to detect breast cancer	NA 63.66%
	Soumik et al. [60]	To classify breast cancer images	Employed Inception V3 [88]	BreakHis [33]	7909	80	20	ACC of 99.50%, 98.90%, 98.96% and 98.51% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 98.97%
	Brancati et al. [50]	To analyze gigapixel histopathological images	Employed CNN with a compressing path and a learning path	Camelyon16 [164], TUPAC16 [165]	892	68	32	AUC values of 0.698, 0.639, and 0.654 obtained for max-pooling, average pooling, and combined attention maps, respectively	0.66 NA
	Mahmoud et al. [61]	To classify breast cancer images	Employed transfer learning	Mammography images [166]	7500	80	20	Maximum ACC of 97.80% was claimed by using the given dataset [166]. Sensitivity and specificity were estimated	NA 94.45%
	Munien et al. [62]	To classify breast cancer images	Employed EfficientNet [167]	ICIAR2018 [78]	400	85	15	Overall ACC of 98.33% obtained from ICIAR2018 [78]. Sensitivity was also taken into account	NA 98.33%
	Bounarraf et al. [63]	To analyze breast cancer images	Employed ResNet [75] on ImageNet [168] images	BreakHis [33]	7909	80	20	ACC of 94.49%, 93.27%, 91.29%, 89.56% obtained for 40x, 100x, 200x, and 400x magnification factors, respectively	NA 92.15%
	Saber et al. [64]	To detect breast cancer	Employed transfer learning technique	MIAS [85]	322	80	20	Overall ACC, PRS, FIS, and AUC of 98.96%, 97.35%, 97.66%, and 0.995, respectively, got from MIAS [85]	0.995 98.96%

TABLE 2: Continued.

Year	Ref	Aim	Technique	Dataset	Sample	Training (%)	Testing (%)	Result	Performance AUC	ACC
2022	Ameh Joseph et al. [169]	To classify breast cancer images	Employed handcrafted features and dense layer	BreakHis [33]	7909	90	10	ACC of 97.87% for 40x, 97.60% for 100x, 96.10% for 200x, and 96.84% for 400x demonstrated from BreakHis [33]	NA	97.08%
	Reshma et al. [52]	To detect breast cancer	Employed probabilistic transition rules with CNN	BreakHis [33]	7909	90	10	ACC, PRS, RES, FIS, and GMN of 89.13%, 86.23%, 81.47%, 85.38%, and 85.17% demonstrated from BreakHis [33]	NA	89.13%
	Huang et al. [53]	To detect nuclei on breast cancer	Employed mask-region-based CNN	H&E images of patients	537	80	20	PRS, RES, and FIS of 91.28%, 87.68%, and 89.44% demonstrated from the used dataset	NA	95.00%
	Chhipa et al. [170]	To learn efficient representations	Employed magnification prior contrastive similarity	BreakHis [33]	7909	70	30	Maximum mean ACC of 97.04% and 97.81% were got from patient and image levels, respectively using BreakHis [33]	NA	97.42%
	Zou et al. [171]	To classify breast cancer images	Employed channel attention module with nondimensionality reduction	BreakHis [33], BACH [78]	8309	90	10	Average ACC, PRS, RES, and FIS of 97.75%, 95.19%, 97.30%, and 96.30% obtained from BreakHis [33], respectively. ACC of 85% got from BACH [78]	NA	91.37%
	Liu et al. [172]	To classify breast cancer images	Employed autoencoder and Siamese framework	BreakHis [33]	7909	80	20	Average ACC, PRS, RES, FIS, and RTM of 96.97%, 96.47%, 99.15%, 97.82%, and 335 seconds obtained from BreakHis [33], respectively	NA	96.97%
	Jayandhi et al. [54]	To diagnose breast cancer	Employed VGG [98] and SVM	MIAS [85]	322	80	20	Maximum ACC of 98.67% obtained from MIAS [85]. Sensitivity and specificity were also calculated	NA	98.67%
	Sharma and Kumar [55]	To classify breast cancer images	Employed Xception [135] and SVM	BreakHis [33]	2000	75	25	Average ACC, PRS, RES, FIS, and AUC of 95.58%, 95%, 95%, 95%, and 0.98 obtained from BreakHis [33], respectively	0.98	95.58%
	Zerouaoui and Idri [56]	To classify breast cancer images	Employed multilayer perceptron, DenseNet201 [67]	BreakHis [33] and others	NA	80	20	ACC of 92.61%, 92%, 93.93%, and 91.73% on four magnification factors of BreakHis [33]	NA	93.85%
	Soltane et al. [65]	To classify breast cancer images	Employed ResNet [75]	323 colored lymphoma images	323	50	50	A total of 27 misclassifications for 323 samples were claimed. PRS, RES, FIS, and Kappa score were estimated	NA	91.6%
	Naik et al. [173]	To analyze breast cancer images	Employed random forest, k-nearest neighbors, SVM	699 whole-slide images	699	80	20	Random forest algorithm achieved better result for classifying benign and malignant images from 190 testing samples	NA	98.2%
	Chattopadhyay et al. [174]	To classify breast cancer images	Employed dense residual dual-shuffle attention network	BreakHis [33]	7909	80	20	Average ACC, PRS, RES, and FIS of 96.10%, 96.03%, 96.08%, and 96.02%, respectively, obtained from four different magnification levels of BreakHis [33]	NA	96.10%
Alruwaili and Gouda [66]	To detect breast cancer	Employed the principle of transfer learning, ResNet [75]	MIAS [85]	322	80	20	Best results for ACC, PRS, RES, FIS, and AUC of 89.5%, 89.5%, 90%, and 89.5% obtained from MIAS [85], respectively	NA	89.5%	

training dataset consists of 270 whole-slide images; among them 160 are normal slides and 110 slides contain metastases [97].

- (vii) PCam [121] \Rightarrow It is a modified version of the Patch Camelyon (PCam) dataset, which consists of 327680 microscopy images with 96×96 -pixel sized patches extracted from the whole-slide images with a binary label hinting the presence of metastatic tissue [8].
- (viii) HASHI [129] \Rightarrow Each image in the dataset of HASHI (high-throughput adaptive sampling for whole-slide histopathology image analysis) [129] has the size of 3002×2384 [161].
- (ix) MIAS [85] \Rightarrow The Mammographic Image Analysis Society (MIAS) database of digital mammograms [85] contains 322 mammogram images, each of which has a size of 1024×1024 pixels with PGM format [59].
- (x) INbreast [92] \Rightarrow The INbreast database has a total of 410 images collected from 115 cases (i.e., patients) indicating benign, malignant, and normal cases having sizes of 2560×3328 or 3328×4084 pixels. It contains 36 benign and 76 malignant masses [92].
- (xi) DDSM [84] \Rightarrow The DDSM [84] dataset was collected by the expert team at the University of South Florida [84]. It contains 2620 scanned film mammography studies. Explicitly, it involves 2620 breast cases (i.e., patients) categorized in 43 different volumes with average size of 3000×4800 pixels [48].
- (xii) CBIS-DDSM [128] \Rightarrow The CBIS-DDSM [128] is an updated version of the DDSM providing easily accessible data and improved region-of-interest segmentation [128, 146]. The CBIS-DDSM dataset comprises 2781 mammograms in the PNG format [49].
- (xiii) CMTHis [37] \Rightarrow The CMTHis (Canine Mammary Tumor Histopathological Image) [37] dataset comprises 352 images acquired from 44 clinical cases of canine mammary tumors.
- (xiv) FABCD [133] \Rightarrow The FABCD (Fully Annotated Breast Cancer Database) [133] consists of 21 annotated images of carcinomas and 19 images of benign tissue taken from 21 patients [130].
- (xv) IICBU2008 [87] \Rightarrow The IICBU2008 (Image Informatics and Computational Biology Unit) malignant lymphoma dataset contains 374 H&E stained microscopy images captured using bright field microscopy [21].
- (xvi) VLAD [136] \Rightarrow The VLAD (Vector of Locally Aggregated Descriptors) dataset [136] consists of 300 annotated images with resolution of 1280×960 [29].
- (xvii) LSC [137] \Rightarrow The LSC (Lymphoma Subtype Classification) [137] dataset has been prepared by pathologists from different laboratories to create a

real-world type cohort which contains a larger degree of stain and scanning variances [137]. It consists of 374 images with resolution of 1388×1040 [29].

- (xviii) KimiaPath24 [126] \Rightarrow The official KimiaPath24 [126] dataset consists of a total of 23916 images for training and 1325 images for testing. It is publicly available. It shows various body parts with texture patterns [41].

2.2. Performance Evaluation Metrics. Performance evaluation of any deep learning model is an important task. An algorithm may give very pleasing results when evaluated using a metric (e.g., *ACC*), but it may give poor results when evaluated against other metrics (e.g., *FIS*) [177]. Usually, we use classification accuracy to measure the performance of deep learning algorithms. But it is not enough to determine the model perfectly. For truly judge any deep learning algorithm, we can use nonidentical types of evaluation metrics including classification *ACC*, *AUC*, *PRS*, *RES*, *FIS*, *RTM*, and *GMN*.

- (i) *ACC* \Rightarrow It is normally defined in terms of error or inaccuracy [178]. It can be calculated using the following equation:

$$ACC = \frac{(100)(t_n + t_p)}{t_p + t_n + f_p + f_n}, \quad (1)$$

where t_n is true negative, t_p is true positive, f_p is false positive, and f_n is false negative. Sometimes, *ACC* and the percent correct classification (*PCC*) can be used interchangeably.

- (ii) *PRS* \Rightarrow Its best value is 100 and the worst value is just 0. It can be formulated using the following equation:

$$PRS = \frac{(100)(t_p)}{t_p + f_p}. \quad (2)$$

- (iii) *RES* \Rightarrow It should ideally be 100 (the highest) for a good classifier. It can be calculated using the following equation:

$$RES = \frac{(100)(t_p)}{t_p + f_n}. \quad (3)$$

- (iv) *AUC* \Rightarrow It is one of the most widely used metrics for evaluation [177–179]. The *AUC* of a classifier equals the probability that the classifier ranks a randomly chosen positive sample higher than a randomly chosen negative sample. The *AUC* varies in value from 0 to 1. If the predictions of a model are 100% wrong, then its *AUC* = 0.00; but if its predictions are 100% correct, then its *AUC* = 1.00.
- (v) *FIS* \Rightarrow It is the harmonic mean between precision and recall. It is also called the F-score or F-measure. It is used in deep learning [177]. It conveys the balance between the precision and the recall. It

also tells us how many instances it classifies correctly. Its highest possible value is 1, which indicates perfect precision and recall. Its lowest possible value is 0, when either the precision or the recall is zero. It can be formulated as

$$F1S = \frac{2}{1/PRS + 1/RES} = \frac{t_p}{t_p + f_p + f_n/2}, \quad (4)$$

where PRS is the number of correct positive results divided by the number of positive results predicted with the classifier and RES is the number of correct positive results divided by the number of all relevant samples.

- (vi) $RTM \Rightarrow$ Estimating the RTM complexity of algorithms is mandatory for many applications (e.g., embedded real-time systems [180]). The optimization of the RTM complexity of algorithms in applications is highly expected. The total RTM can prove to be one of the most important determinative performance factors in many software-intensive systems.
- (vii) $GMN \Rightarrow$ It indicates the central tendency or typical value of a set of numbers by considering the product of their values instead of using their sum. It can be used to attain a more accurate measure of returns than the mean or arithmetic mean or average. The GMN for any set of numbers $x_1, x_2, x_3, \dots, x_m$ can be defined as

$$GMN = \left(\prod_{i=1}^m x_i \right)^{1/m} = \sqrt[m]{x_1 x_2 x_3 \dots x_m}. \quad (5)$$
- (viii) $MCC \Rightarrow$ The Matthews correlation coefficient (MCC) is used as a measure of the quality of binary classifications, introduced by biochemist Brian W. Matthews in 1975.
- (ix) $\kappa \Rightarrow$ The metric of Cohen's kappa (κ) can be used to evaluate binary classifications.

3. A Succinct Survey of State of the Art

This section deals with a summary of existing studies apposite for the classification of breast cancer histopathological images followed by a short discussion and our findings.

3.1. Summary of Previous Studies. Table 2 provides a short summary of previous studies carried out to classify breast cancer from images. Experimental results of miscellaneous deep models in the literature on publicly available datasets demonstrated various degrees of accurate cancer prediction scores. However, as AUC and ACC are the most important metrics for breast cancer histopathological images classification [49], the experimental results in Table 2 take them into account as the performance indices.

3.2. Key Techniques and Challenges. The CNNs can be regarded as a variant of the standard neural networks. Instead of using fully connected hidden layers, the CNNs introduce the structure of a special network, which comprises so-called alternating convolution and pooling layers. They were first introduced for overcoming known problems of fully connected deep neural networks when handling high dimensionality structured inputs, such as images or speech. From Table 2, it is noticeable that CNNs have become state-of-the-art solutions for breast cancer histology images classification. However, there are still challenges even when using the CNN-based approaches to classify pathological breast cancer images [16], as given below:

- (i) Risk of overfitting \Rightarrow The number of parameters of CNN increases rapidly depending on how large the network is, which may lead to poor learning.
- (ii) Being cost-intensive \Rightarrow To get a huge number of labeled breast cancer images is very expensive.
- (iii) Huge training data \Rightarrow CNNs need to be trained using a lot of images, which might not be easy to find considering that collecting real-world data is a tedious and expensive process.
- (iv) Performance degradation \Rightarrow Various hyperparameters have a significant influence on the performance of the CNN model. The model's parameters need to be tuned properly to achieve a desirable result [75], which usually is not an easy task.
- (v) Employment difficulty \Rightarrow In the process of training CNN model, it is usually inevitable to rearrange the learning rate parameters to get a better performance. This makes it arduous for the algorithm to use in real-life applications by nonexpert users [181].

Many methods had been proposed in the literature considering the aforementioned challenges. In 2012, AlexNet [81] architecture was introduced for ImageNet Challenge having error rate of 16%. Later various variations of AlexNet [81] with denser network were introduced. Both AlexNet [81] and VGGNet [98] were the pioneering works that demonstrated the potential of deep neural networks [182]. AlexNet was designed by Alex Krizhevsky [81]. It contained 8 layers; the first 5 were convolutional layers, some of them followed by max-pooling layers, and the last 3 were fully connected layers [81]. It was the first large-scale CNN architecture that did well on ImageNet [183] classification. AlexNet [81] was the winner of the ILSVRC [183] classification, the benchmark in 2012. Nevertheless, it was not very deep. SqueezeNet [184] was proposed to create a smaller neural network with fewer parameters that could be easily fit into computer memory and transmitted over a computer network. It achieved AlexNet [81] level accuracy on ImageNet with 50x fewer parameters. It was compressed to less than 0.5 MB (510x smaller than AlexNet [81]) with model compression techniques. The VGG [98] is a deep CNN used to classify images. The VGG19 is a variant of VGG which consists of 19 layers (i.e., 16 convolution layers and 3 fully connected layers, in addition to 5 max-pooling layers and 1 SoftMax layer) [98]. There exist many variants of

VGG [98] (e.g., VGG11, VGG16, VGG19, etc.). VGG19 has 19.6 billion FLOPs (floating point operations per second). VGG [98] is easy to implement but slow to train. Nowadays, many deep-learning-based methods are implemented on influential backbone networks; among them, both DenseNet [67] and ResNet [75] are very popular. Due to the longer path between the input layer and the output layer, the information vanishes before reaching its destination. DenseNet [67] was developed to minimize this effect. The key base element of ResNet [75] is the residual block. DenseNet [67] concentrates on making the deep learning networks move even deeper as well as simultaneously making them well organized to train by applying shorter connections among layers. In short, ResNet [75] adopts summation, whereas DenseNet [67] deals with concatenation. Yet, the dense concatenation of DenseNet [67] creates a challenge of demanding high GPU (Graphics Processing Unit) memory and more training time [182]. On the other hand, the identity shortcut that balances training in ResNet [75] curbs its representation dimensions [182]. Compendiously, there is a dilemma in the alternative between ResNet [75] and DenseNet [67] for many applications in terms of performance and GPU resources [182].

3.3. Our Findings. Although various deep learning models in Table 2 often achieved pretty good scores of AUC and ACC, the models demand a large amount of data but breast cancer diagnosis always suffers from a lack of data. To adopt

artificial data is a tentative solution of this issue, but the determination of the best hyperparameters is extremely difficult. Besides efficient deep learning models, the datasets themselves have some limitations, for example, overinterpretation, which cannot be diagnosed using typical evaluation methods based on the ACC of the model. Deep learning models trained on popular datasets (e.g., BreakHis [33]) may suffer from overinterpretation. In overinterpretation, deep learning models make confident predictions based on details that do not make any sense to humans (e.g., promiscuous patterns and image borders). When deep learning models are trained on datasets, they can make apparently authentic predictions based on both meaningful and meaningless subtle signals. This effect, eventually, can reduce the overall classification performance of deep models. Most probably, this is one of the reasons why any state-of-the-art deep learning model in the literature for classifying breast cancer histopathological images (see Table 2) could not show an ACC of 100%.

In addition, the training-testing ratio can regulate the performance of a deep model for image classification. We wish to determine the most popular and/or optimized training-testing ratios for classifying histopathological images using Table 2. To this end, we have calculated the usage frequency of the training-testing ratio (i.e., percentage of the number of papers that used the same ratio) by considering data in Table 2 and the following equation:

$$\text{Usage frequency (\%)} = \frac{(\text{Total number of papers use da training - testing ratio})(100)}{\text{Sum of papers both use da ndu nuse dt hesamet training - testing ratio}}. \quad (6)$$

Figure 3 demonstrates the frequency of usage of training-testing ratio considering data in Table 2. From Figure 3, it is noticeable that the most popular training-testing ratio for histopathological image classification is 70%: 30%. The second-best used training-testing ratio is 80%: 20%, followed by 90%: 10%, 75%: 25%, 50%: 50%, and so on. Figure 4 presents the GMN of ACC for the most frequently used training-testing ratios considering data in Table 2. It shows a different history; in terms of ACC, the rate of 80%: 20% became the best option for the training-testing ratio to classify histopathological images. Explicitly, the GMN of ACC formed like a Gaussian shaped curve and the ratio of 80%: 20% owned its highest peak. To cut a long story short, by considering ACC, the training-testing ratio of 80%: 20% became the finest and the optimal choice for classifying histopathological images.

4. Methods and Materials

This section explains in detail our proposed DenTnet model and its implementation. Figure 5 demonstrates a general flowchart of our methodology to classify breast cancer histopathological images automatically.

4.1. Architecture of Our Proposed DenTnet. The architecture of our proposed DenTnet is shown in Figure 1, which consists of four different blocks, namely, the input volume, training from scratch, transfer learning, and fusion and recognition.

4.1.1. Input Volume. The input is a 3D RGB (three-dimensional red, green, and blue) image with a size of 224×224 , that is, $224 \times 224 \times 3$.

4.1.2. Training from Scratch. Initially, features are extracted from the input images by feeding the input to the convolutional layer. The convolution (conv) layers contain a set of filters (or kernels) parameters, which are learned throughout the training. The size of the filters is usually smaller than the actual image, where each filter convolves with the image and creates an activation map. Thereafter, the pooling layer progressively decreases the spatial size of the representation for reducing the number of parameters in the network. Instead of differentiable functions such as sigmoid and tanh, the network utilizes the ReLU as an activation function. Finally, the extracted features or the output of the last layer from the training from scratch block is then

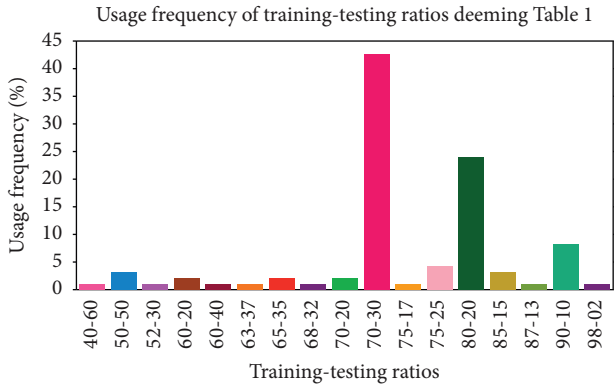


FIGURE 3: Determination of the most popular training-testing ratios using data from Table 2.

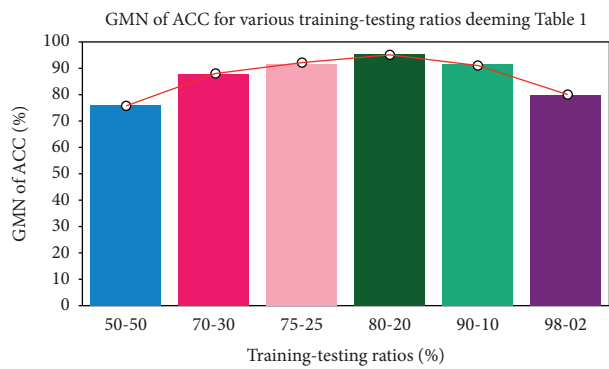


FIGURE 4: GMN of ACC for the most popular training-testing ratios deeming data from Table 2.

amalgamated with the features extracted from the transfer learning approach. Figure 1 includes the design of the DenseNet [67] architecture used to extract the feature using the learning-from-scratch approach.

4.1.3. Transfer Learning. In transfer learning, given that a domain \mathcal{D} consists of feature space \mathcal{X} and a marginal probability distribution $P(X)$, where $X = x_1, x_2, \dots, x_n \in \mathcal{X}$, and a task \mathcal{T} consists of a label space \mathcal{Y} and an objective predictive function $f: \mathcal{X} \rightarrow \mathcal{Y}$, the corresponding label $f(x)$ of a new instance x is predicted by function f , where the new tasks denoted by $\mathcal{T} = \mathcal{Y}$, $f(x)$ are learned from the training data consisting of pairs x_i and y_i , where $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$. When utilizing the learning-from-scratch approach, the extracted features stay in the same distribution. To solve this problem, we amalgamated both learning-from-scratch and the transfer learning approach. The learned parameters are further fine-tuned by retraining the extracted features. This is anticipated to expand the prior knowledge of the network about the data, which might improve the efficiency of the model during training, thereby accelerating the learning speed and also increasing the accuracy of the model. As shown in Figure 1, there is a connection between the blocks of the input volume and transfer learning. The transfer learning approach extracted features from the ImageNet [168] weights. The weight is the parameters (including trainable and nontrainable) learned from the

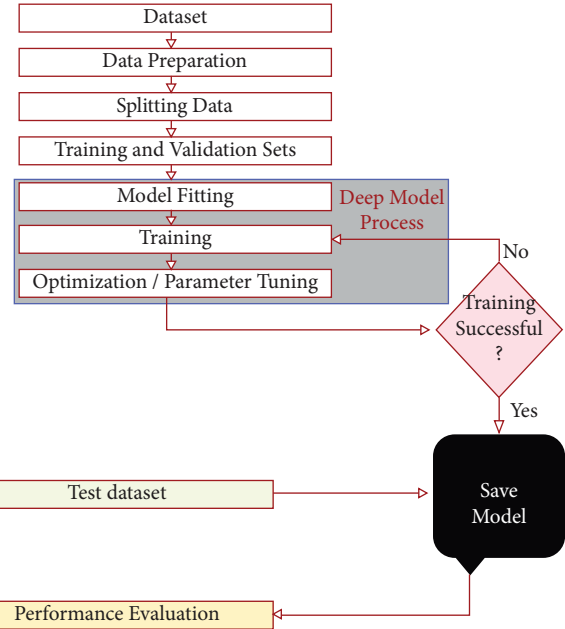


FIGURE 5: Flowchart of our methodology to classify breast cancer histopathological images.

ImageNet [168] dataset. Since transfer learning involves transferring knowledge from one domain to another, we have utilized the ImageNet weight as the models developed in the ImageNet [168] classification competition are measured against each other for performance. Henceforth, the ImageNet weight provides a measure of how good a model is for classification. Besides, the ImageNet weight has already showed a markedly high accuracy [185]. The extracted features are then used by the network before being passed to the fusion and recognition block, where the features are amalgamated with the extracted features from the learning-from-scratch block for recognition.

4.1.4. Fusion and Recognition. The extracted features based on the ImageNet weights are then amalgamated with the features extracted by the block of training from scratch. A global average pooling is performed. Dropout technique helps to prevent a model from overfitting. It is used with dense fully connected layers. The fully connected layer compiles the data extracted by previous layers to form the final output. The last step passes the features through the fully connected layer, which then uses SoftMax to classify the class of the input images.

4.2. Implementation Details

4.2.1. Data Preparation. We have adopted data augmentation, stain normalization, and image normalization strategies to optimize the training process. Hereby, we have explained each of them briefly.

4.2.2. Data Augmentation. Due to the limited size of the input samples, the training of our DenTnet was prone to overfitting, which caused low detection rate. One solution to

alleviate this issue was the data augmentation, which generated more training data from the existing training set. Dissimilar data augmentation techniques (e.g., horizontal flipping, rotating, and zooming) were applied to datasets for creating more training samples.

4.2.3. Stain Normalization. The breast cancer tissue slices are stained by H&E to differentiate between nuclei stained with purple color and other tissue structures stained with pink and red color to help pathologists analyze the shape of nuclei, density, variability, and overall tissue structure [186]. The H&E staining variability between acquired images exists due to the different staining protocols, scanners, and raw materials. This is a common problem with histological image analysis. Therefore, stain normalization of H&E-stained histology slides was a key step for reducing the color variation and obtaining a better color consistency prior to feeding input images into the DenTnet architecture. Different techniques are available for stain normalization in histological images. We have considered Macenko technique [187] due to its promising performance in many studies to standardize the color intensity of the tissue. This technique was based on a singular value decomposition. A logarithmic function was used to adaptively transform color concentration of the original histopathological image into its optical density (OD) image as $OD = -\log(I/I_0)$, where OD hints the matrix of optical density values, I belongs to the image intensity in red-green-blue space, and I_0 addresses the illuminating intensity incident on the histological sample.

4.2.4. Intensity Normalization. Intensity normalization was another important preprocessing step. Its primary aim was to get the same range of values for each input image before feeding to the DenTnet. It also speeded up the convergence of DenTnet. Input images were normalized to the standard normal distribution by min-max normalization (i.e., using one of the most popular ways to normalize data) to the intensity range of $[0, 1]$, which can be computed as

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}}, \quad (7)$$

where x , x_{min} , and x_{max} indicate pixel, minimum, and maximum intensity values of the input image, respectively.

4.2.5. Hardware and Software Requirements. DenTnet was implemented using the TensorFlow and Keras framework [188, 189] and coded in Python using Jupyter Notebook on a Kaggle Private Kernel. The experiment was performed on a machine with the following configuration: Intel® Xeon® CPU @ 2.30 GHz with 16 CPU Cores, 16 GB RAM, and NVIDIA Tesla P100 GPU. We implemented and trained everything on the cloud using Kaggle GPU hours.

4.2.6. Training and Testing Setup. The dataset was divided in a 80%: 20% ratio, where 80% was used for training and the remaining 20% was used for testing. The data used for testing were kept isolated from the training set and never seen by the

model during training. To evaluate the images classification, we have computed the recognition rate at the image level over the two different classes: (i) correctly classified images and (ii) the total number of images in the test set.

4.2.7. Training Procedure. In the training of a neural network, a measure of error is required to compute the error between the targeted output and the computed output of training data known as the loss function. An optimization algorithm is needed to minimize this function. We have considered Adam optimizer [190] with numerical stability constant $epsilon = None$, $decay = 0.0$, and $AMSGrad = True$. Table 3 presents the hyperparameter values of the proposed deep learning model. Learning rate (also referred to as step size) signifies the proportion to which weights are updated. A smaller value (e.g., 0.000001) slows down the learning process during training, whereas a larger value (e.g., 0.400) results in faster learning. We have considered a learning rate of 0.001. The exponential decay rates of the first and second moments were estimated to be 0.60 and 0.90, respectively. To update the weights, the number of epochs was set to 50 with 3222 steps per epoch and a batch size of 32. For the BreakHis [33] dataset, we had a training sample of 103104 images, with 12288 validation samples and 697 testing samples. The training process used 10-fold cross-validation, where one of the samples was used to validate the data and the remaining 9 samples were used to train the DenTnet model. The fully connected layer used 1024 filters with a dropout rate of 0.50. Finally, the last layer used two filters with a SoftMax layer to classify the image into two classes (e.g., benign and malignant). We have used categorical cross-entropy as the objective function to quantify the difference between two probability distributions. The whole training process took more than 4 hours for the breast cancer tissue images.

5. Experimental Results and Comparison on BreakHis Dataset

This section demonstrates the experimental results achieved from classifying the breast cancer histopathology (i.e., BreakHis [33]) images using our proposed DenTnet model.

Figure 6 shows the performance curves obtained during the training of DenTnet using BreakHis [33] dataset. A normalized confusion matrix for the classification of breast cancer test set images is illustrated in Figure 7(a). The main reason for confusion between benign and malignant breast tissues is their similar textures or expression. Henceforth, careful description of texture is required to remove the confusion between the two classes. For binary classification, 5 images only were misclassified, indicating that DenTnet achieved the highest and best ACC of 99.28%. Figures 7(b) and 7(c) demonstrate the ROC curve and precision-recall curve for classification of benign and malignant images from BreakHis [33] dataset, respectively. AUC of 0.99, sensitivity of 97.73%, and specificity 100% have been reported. Table 4 lists the complete classification report of DenTnet. It achieved an ACC of 99.28%.

TABLE 3: List of hyperparameter values for the proposed deep learning model.

Model	Hyperparameters							
	Beta_1	Beta_2	Learning rate	Epoch	Batch size	Epsilon	Decay	AMSGrad
DenTnet	0.60	0.90	0.001	50	32	None	0.0	True

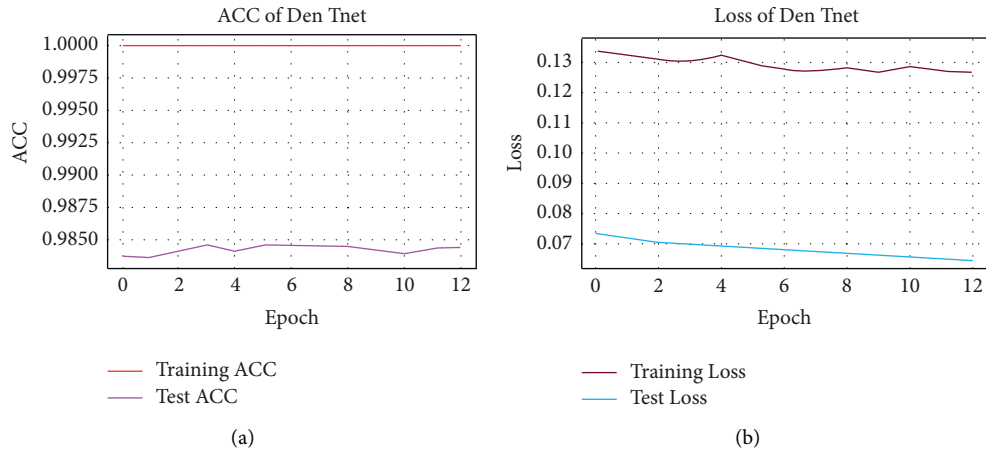


FIGURE 6: (a) hints ACC and (b) shows loss charts of DenTnet during training.

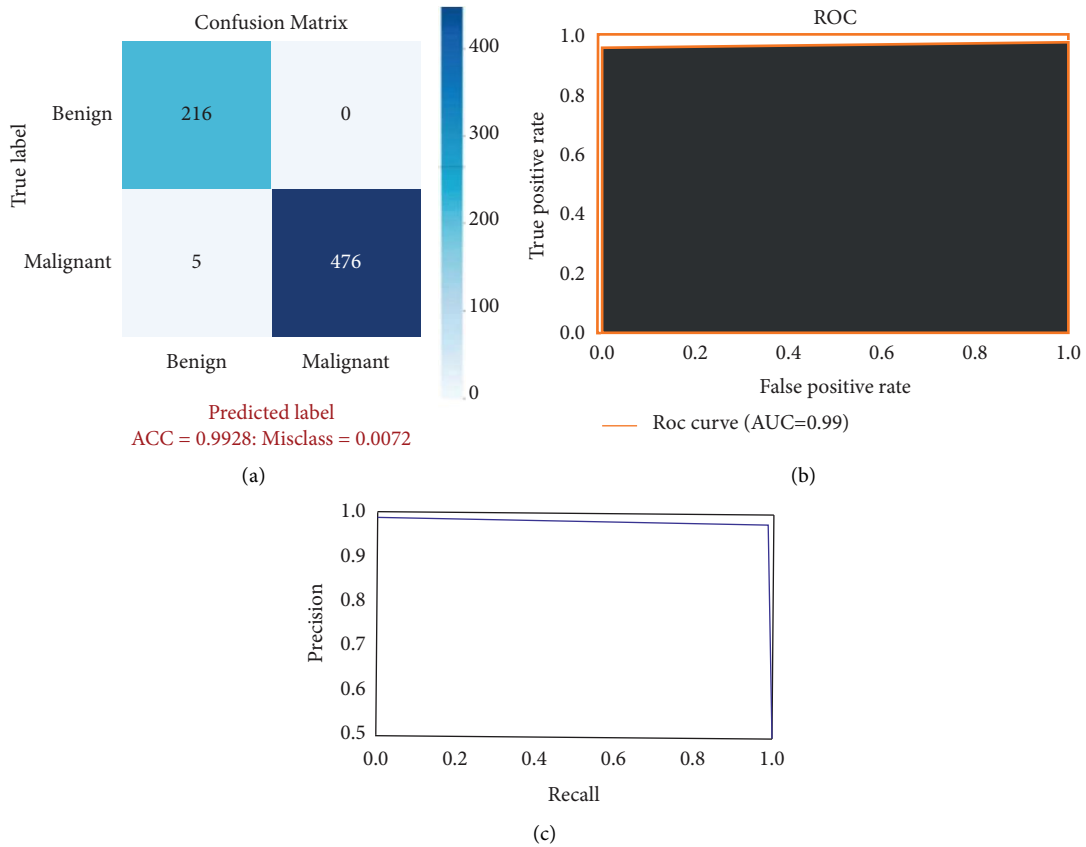


FIGURE 7: (a) hints confusion matrix for benign and malignant classification, (b) shows ROC curve, and (c) demonstrates precision-recall curve.

TABLE 4: Classification results by counting all evaluation criteria.

Type	PRS	RES	F1S	Support
Benign	0.98	1.00	0.99	216
Malignant	1.00	0.99	0.99	481
Micro mean	0.99	0.99	0.99	697
Macro mean	0.99	0.99	0.99	697
Weighted mean	0.99	0.99	0.99	697

Table 1 compares the results obtained by several methods. The methods of Togacar et al. [26], Parvin et al. [31], Man et al. [36], Soumik et al. [60], Liu et al. [172], Zerouaoui and Idri [56], and Chattopadhyay et al. [174] were centered on mainly CNN models, but they were tested against the same training-testing ratio of 80%: 20% on the BreaKHis dataset [33]. However, Boumaraf et al. [63] suggested a transfer-learning-based method deeming the residual CNN ResNet-18 as a backbone model with block-wise fine-tuning strategy and obtained a mean ACC of 92.15% applying a training-testing ratio of 80%: 20% on BreaKHis dataset [33]. From Table 1, it is notable that DenTnet [ours] achieved the best ACC on the same ground.

6. Generalization Ability Evaluation of Proposed DenTnet

What would be the performance of the proposed DenTnet compared with other types of cancer or disease datasets? To evaluate the generalization ability of DenTnet, this section presents the experimental result obtained not only from the dataset of BreaKHis [33] but also from additional datasets of Malaria [191], CovidXray [192], and SkinCancer [193].

6.1. Datasets Irrelevant to Breast Cancer. The three following datasets are not related to breast cancer. Herewith, their primary aim is to evaluate the generalization ability of our proposed method DenTnet:

- (i) Malaria [191] \Rightarrow This dataset contains a total of 27558 infected and uninfected images for malaria.
- (ii) SkinCancer [193] \Rightarrow This dataset contains balanced images from benign skin moles and malignant skin moles. The data consist of two folders, each containing 1800 pictures (224×244) from the two types of mole.
- (iii) CovidXray [192] \Rightarrow Corona (COVID-19) virus affects the respiratory system of healthy individual. The chest X-ray is one of the key imaging methods to identify the coronavirus. This dataset contains chest X-ray of healthy versus pneumonia (Corona) infected patients along with few other categories including SARS (Severe Acute Respiratory Syndrome), *Streptococcus*, and ARDS (Acute Respiratory Distress Syndrome) with a goal of predicting and understanding the infection.

Figure 8 specifies some sample images from Malaria [191], SkinCancer [193], and CovidXray [192] datasets.

6.2. Experimental Results Comparison. Using four datasets in the experiment, DenTnet has been compared with six widely used and well-known deep learning models, namely, AlexNet [81], ResNet [75], VGG16 [98], VGG19 [98], Inception V3 [88], and SqueezeNet [184]. To evaluate and analyze the performance of DenTnet, four different cases are considered. The first case is the evaluation of different deep learning methods, which are trained and tested on BreaKHis [33] dataset. The second case studies the performance of the deep-learning-based classification methods that are trained and tested on Malaria [191] dataset. The third case is to train and test the deep learning models on SkinCancer [193] dataset. The final one is to understand and analyze the performance of the deep learning models on CovidXray [192] dataset. The overall results are tabulated in Tables 5–9. Besides, the RTM in seconds of various datasets using the deep learning models is shown in Table 10.

According to the results in terms of GMN of ACC, RES, F1S, and AUC as shown in Tables 5–9, respectively, the proposed DenTnet architecture provides the best scores as compared to AlexNet [81], ResNet [75], VGG16 [98], VGG19 [98], Inception V3 [88], and SqueezeNet [184]. On the other hand, DenTnet gets the third best result. Moreover, in most of the cases, AlexNet [81] obtains the lowest results.

6.3. Performance Evaluation. The deepening of deep models makes their parameters rise rapidly, which may lead to overfitting of the model. To take the edge off the overfitting problem, predominantly a large number of dataset images are required as the training set. Considering a small dataset, it is possible to reduce the risk of overfitting of the model by reducing the parameters and augmenting the dataset. Accordingly, DenTnet used fewer parameters along with the dense connections in the construction of the model, instead of the direct connections among the hidden layers of the network. As DenTnet used fewer parameters, it attenuated the vanishing gradient descent and strengthened the feature propagation. Consequently, the proposed DenTnet outperformed its alternative state-of-the-art methods. Yet, its runtime was a bit longer in Malaria [191] and SkinCancer [193] datasets as compared to ResNet [75]. The main reason why the DenTnet model may require more time is that it uses many small convolutions in the network, which can run slower on GPU than compact large convolutions with the same number of GFLOPS. Still, DenTnet includes fewer parameters compatibility when compared to ResNet [75]. Henceforth, it is more efficient in solving the problem of overfitting. In general, all of the used algorithms suffered

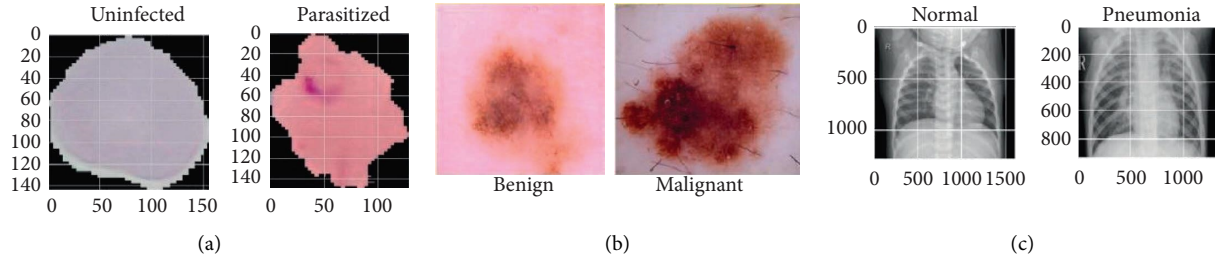


FIGURE 8: (a), (b), and (c) specify images of Malaria [191], SkinCancer [193], and CovidXray [192] datasets, respectively.

TABLE 5: ACC of various methods deeming four different datasets.

Models	ACC of various datasets				GMN of ACC	
	BreaKHis [33]	Malaria [191]	SkinCancer [193]	CovidXray [192]	Success	Failure
AlexNet [81]	0.9268	0.9738	0.8714	0.8526	0.9049	0.0951
ResNet [75]	0.9857	0.9832	0.9045	0.8990	0.9422	0.0578
VGG16 [98]	0.9785	0.9806	0.8501	0.8576	0.9145	0.0855
VGG19 [98]	0.9785	0.9811	0.8512	0.9279	0.9328	0.0672
Inception V3 [88]	0.9784	0.9879	0.8587	0.8998	0.9296	0.0704
SqueezeNet [184]	0.9756	0.9498	0.8288	0.8016	0.8858	0.1142
DenTnet [ours]	0.9928	0.9865	0.9157	0.8942	0.9463	0.0537

TABLE 6: PRS of various methods deeming four different datasets.

Models	PRS of various datasets				GMN of PRS	
	BreaKHis [33]	Malaria [191]	SkinCancer [193]	CovidXray [192]	Success	Failure
AlexNet [81]	0.9317	0.9656	0.8417	0.8744	0.9021	0.0979
ResNet [75]	0.9937	0.9793	0.9167	0.8667	0.9377	0.0623
VGG16 [98]	0.9936	0.9888	0.9055	0.8533	0.9334	0.0666
VGG19 [98]	0.9814	0.9753	0.8083	0.9872	0.9348	0.0652
Inception V3 [88]	0.9829	0.9713	0.8512	0.9796	0.9446	0.0554
SqueezeNet [184]	0.9854	0.9778	0.8871	0.7799	0.9036	0.0964
DenTnet [ours]	0.9700	0.9848	0.9258	0.8641	0.9350	0.0650

TABLE 7: RES of various methods deeming four different datasets.

Models	RES of various datasets				GMN of RES	
	BreaKHis [33]	Malaria [191]	SkinCancer [193]	CovidXray [192]	Success	Failure
AlexNet [81]	0.9647	0.9812	0.9154	0.8880	0.9366	0.0634
ResNet [75]	0.9854	0.9867	0.9010	0.9685	0.9597	0.0403
VGG16 [98]	0.9751	0.9718	0.8250	0.9846	0.9367	0.0633
VGG19 [98]	0.9875	0.9865	0.9065	0.9059	0.9457	0.0543
Inception V3 [88]	0.9854	0.9819	0.8874	0.9491	0.9501	0.0499
SqueezeNet [184]	0.9792	0.9197	0.7861	0.9514	0.9059	0.0941
DenTnet [ours]	0.9896	0.9879	0.9208	0.9629	0.9649	0.0351

TABLE 8: F1S of various methods deeming four different datasets.

Models	F1S of various datasets				GMN of F1S	
	BreaKHis [33]	Malaria [191]	SkinCancer [193]	CovidXray [192]	Success	Failure
AlexNet [81]	0.9479	0.9734	0.8770	0.8811	0.9189	0.0811
ResNet [75]	0.9896	0.9830	0.9129	0.9147	0.9494	0.0506
VGG16 [98]	0.9843	0.9803	0.8634	0.9143	0.9342	0.0658
VGG19 [98]	0.9845	0.9809	0.8546	0.9448	0.9397	0.0603
Inception V3 [88]	0.9844	0.9724	0.8693	0.9077	0.9322	0.0678
SqueezeNet [184]	0.9823	0.9479	0.8336	0.8571	0.9031	0.0969
DenTnet [ours]	0.9948	0.9864	0.9233	0.9108	0.9531	0.0469

TABLE 9: AUC of various methods deeming four different datasets.

Models	AUC of various datasets				GMN of AUC	
	BreaKHis [33]	Malaria [191]	SkinCancer [193]	CovidXray [192]	Success	Failure
AlexNet [81]	0.90	0.97	0.87	0.85	0.8964	0.1036
ResNet [75]	0.99	0.98	0.90	0.91	0.9441	0.0559
VGG16 [98]	0.98	0.98	0.86	0.85	0.9154	0.0846
VGG19 [98]	0.97	0.98	0.85	0.91	0.9260	0.0740
Inception V3 [88]	0.97	0.97	0.89	0.87	0.9239	0.0761
SqueezeNet [184]	0.97	0.95	0.83	0.75	0.8703	0.1297
DenTnet [ours]	0.99	0.99	0.91	0.90	0.9465	0.0535

TABLE 10: RTM of various methods deeming four different datasets.

Models	RTM in seconds of various datasets				GMN of RTM
	BreaKHis [33]	Malaria [191]	SkinCancer [193]	CovidXray [192]	
AlexNet [81]	07573	4100	1413	1328	2762.8
ResNet [75]	16889	3556	0799	2683	3368.5
VGG16 [98]	13419	7698	1450	1081	3567.2
VGG19 [98]	23502	7115	1255	1294	4059.4
Inception V3 [88]	14404	7357	1329	1189	3597.3
SqueezeNet [184]	20080	4140	1339	1864	3795.3
DenTnet [ours]	11083	7102	0873	1519	3196.3

from some degree of overfitting problem on all datasets. We minimized such problems by reducing the batch size and adjusting the learning rate and the dropout rate. In some cases, the proposed DenTnet predicted fewer positive samples as compared to ResNet [75]. This is due to the lack of its conservative designation of the positive class. Thus, the GMN PRS of the proposed DenTnet was about 2% lower than that of ResNet [75].

As VGG16 [98] is easier to implement, many deep learning image classification problems benefit from the technique by using the network either as a sole model or as a backbone architecture to classify images. While VGG19 [98] is better than the VGG16 [98] model, they are both very slow to train—for example, a ResNet with 34 layers only requires 18% of operations as a VGG with 19 layers (around half the layers of the ResNet) will require [194]. Regarding AlexNet [81], the model struggled to scan all features as it is not very deep, resulting in poor performance. The SqueezeNet [184] model achieved approximately the same performance as the AlexNet [81] model. VGG19 [98] and Inception V3 [88] showed almost the same level of effectiveness. Although the ResNet [75] model has proven to be a powerful tool for image classification and is usually fast, it has been shown to take a long time to train. Concisely, using all benefits of DenseNet [67] with optimization, DenTnet obtained the highest GMN ACC of 0.9463, RES of 0.9649, F1S of 0.9531, and AUC of 0.9465 from all four datasets. This implies that DenTnet has the best generalization ability compared to its alternative methods.

Often, it is important to measure that certain deep learning models are more efficient and practical as compared to their alternatives. Seemingly, it is difficult to measure such superiority from the obtained experimental results in Tables 5–10. Nonetheless, nonparametric statistical test can make a clear picture of this issue.

7. Nonparametric Statistical Analysis

Figure 9 depicts performance evaluation of various algorithms deeming the numerical values of the ineffectualness metrics and RTM from Table 11. It is noted that, for a better visualization purpose, the RTM scores in Figure 9 use log-normal distribution [195] with a mean of 10 and standard deviation of 1. However, from this graph, it is extremely hard to rank each algorithm. However, statistically, it is possible to show that one algorithm is better than its alternatives. Friedman test [196] and its derivatives (e.g., Iman-Davenport test [197]) are normally referred to as examples of the most well-known nonparametric tests for multiple comparisons. The mathematical equations of Friedman [196], Friedman’s aligned rank [198], and Quade [199] tests can be found in the works of Quade [199] and Westfall and Young [200]. Friedman test [196] takes measures in preparation for ranking of a set of algorithms with performance in descending order. But it can solely inform us about the appearance of differences among all samples of results under comparison. Henceforth, its alternatives (e.g., Friedman’s aligned rank test [198] and Quade test [199]) can give us further information. Consequently, we have performed the tests of Friedman [196], Friedman’s aligned rank [198], and Quade [199] for average rankings based on the features of our experimental study. On rejecting null-hypotheses, we have continued to use post hoc procedures to find the special pairs of algorithms that give idiosyncrasies. In the case of $1 \times N$ comparisons, the post hoc procedures make up for Bonferroni-Dunn’s [201], Holm’s [202], Hochberg’s [203], Hommel’s [204, 205], Holland and Copenhaver’s [206], Rom’s [207], Finner’s [208], and David Li’s [209] procedures, whereas the post hoc procedures of Nemenyi [210], Shaffer [211], and Bergmann-Hommel [212] are involved in $N \times N$ comparisons. The details can be found in the works of

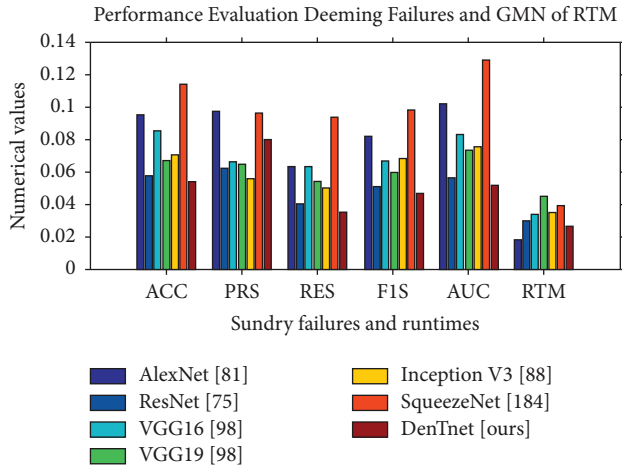


FIGURE 9: Plotting of the numerical values using data from Table 11.

Bergmann and Hommel [212], García and Herrera [213], and Hommel and Bernhard [205].

7.1. Average Ranking of Algorithms. To get the nonparametric statistical test results, Friedman [196], Friedman's aligned rank [198], and Quade [199] tests have been applied to the results of seven models in Table 11. Explicitly, statistical tests have been applied to a matrix with dimension of 7×6 , where 7 is the number of models and 6 is the number of parameters (as 6 datasets while applied to the statistical software environment [214]) in each model. Table 12 shows the average ranking computed by using Friedman [196], Friedman's aligned rank [198], and Quade [199] nonparametric statistical tests. The nonparametric Friedman [196], Friedman's aligned rank [198], and Quade [199] tests determine whether there were significant differences among various models taking data from Table 11. These tests provide the average ranking of all algorithms; that is, the best performing algorithm gets the highest rank of 1, the second-best algorithm gets the rank of 2, and so on.

Figure 10 makes a visualization of the average rankings using the data in Table 12. From Figure 10, it is noticeable that the algorithm of DenTnet [ours] became the best performing one, with the longest bars of 0.6667, 0.1395, and 0.7242 for Friedman test [196], Friedman's aligned rank test [198], and Quade test [199], respectively. This indicates that the algorithm of DenTnet [ours] gives great performance for the solution of underlying problems of classifying breast cancer histopathological images from four different datasets. Friedman [196] statistic considered reduction performance (distributed according to chi-square with 6 degrees of freedom) of 24.500000. Friedman's aligned [198] statistic considered reduction performance (distributed according to chi-square with 6 degrees of freedom) of 23.102557. Iman-Davenport [197] statistic considered reduction performance (distributed according to F-distribution with 6 and 30 degrees of freedom) of 10.652174. Quade [199] statistic considered reduction performance (distributed according to F-distribution with 6 and 30 degrees of freedom) of 5.274194. The p values computed through Friedman

statistic, Friedman's aligned statistic, Iman-Davenport statistic, and Quade statistic are 0.000422, 0.000762847204, 0.000002458229, and 0.000820133186, respectively.

Table 13 demonstrates the results obtained on post hoc comparisons of adjusted p values; $\alpha = 0.05$ and $\alpha = 0.10$. Using level of significance $\alpha = 0.05$, (i) Bonferroni-Dunn's [201] procedure rejects those hypotheses that have an unadjusted p value ≤ 0.008333 ; (ii) Holm's [202] procedure rejects those hypotheses that have an unadjusted p value ≤ 0.016667 ; (iii) Hochberg's [203] procedure rejects those hypotheses that have an unadjusted p value ≤ 0.0125 ; (iv) Hommel's [204] procedure rejects those hypotheses that have an unadjusted p value ≤ 0.016667 ; (v) Holland's [206] procedure rejects those hypotheses that have an unadjusted p value ≤ 0.016952 ; (vi) Rom's [207] procedure rejects those hypotheses that have an unadjusted p value ≤ 0.013109 ; (vii) Finner's [208] procedure rejects those hypotheses that have an unadjusted p value ≤ 0.033617 ; and (viii) Li's [209] procedure rejects those hypotheses that have an unadjusted p value ≤ 0.021422 .

7.2. Post Hoc Procedures: $1 \times N$ Comparisons. In the case of $1 \times N$ comparisons, the post hoc procedures consist of Bonferroni-Dunn's [201], Holm's [202], Hochberg's [203], Hommel's [204, 205], Holland and Copenhaver's [206], Rom's [207], Finner's [208], and David Li's [209] procedures. In these tests, multiple comparison post hoc procedures have been considered for comparing the control algorithm of DenTnet [ours] with others. The results have been shown by computing p values for each comparison. Table 14 depicts the obtained p values using the ranks computed by nonparametric Friedman [196], Friedman's aligned rank [198], and Quade [199] tests. All tests have demonstrated significant improvements of DenTnet [ours] over AlexNet [81], ResNet [75], VGG16 [98], VGG19 [98], Inception V3 [88], and SqueezeNet [184] counting each and every post hoc procedure. Besides, David Li's [209] procedure had the greatest performance, reaching the lowest p value in the comparisons.

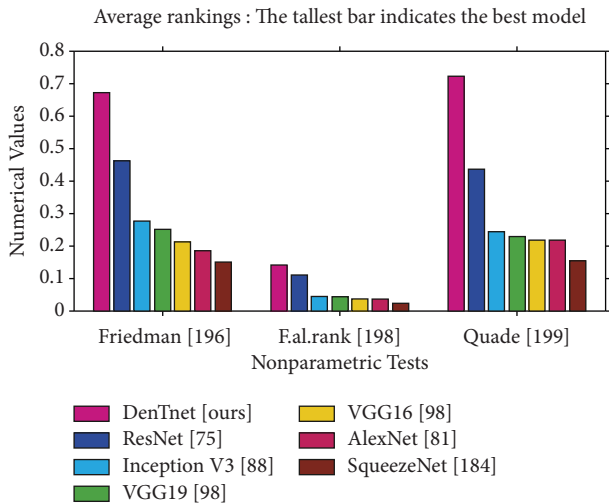
7.3. Post Hoc Procedures: $N \times N$ Comparisons. In the case of $N \times N$ comparisons, the post hoc procedures consist of Nemenyi's [210], Shaffer's [211], and Bergmann-Hommel's [212] procedures. Table 15 presents 21 hypotheses of equality among 7 different algorithms and p values achieved. Using level of significance $\alpha = 0.05$, (i) Nemenyi's [210] procedure rejects those hypotheses that have an unadjusted p value ≤ 0.002381 ; (ii) Holm's [202] procedure rejects those hypotheses that have an unadjusted p value ≤ 0.002778 ; (iii) Shaffer's [211] procedure rejects those hypotheses that have an unadjusted p value ≤ 0.002381 ; and (iv) Bergmann's [212] procedure rejects those hypotheses of AlexNet [81] versus DenTnet [ours], ResNet [75] versus SqueezeNet [184], and SqueezeNet [184] versus DenTnet [ours]. On the other hand, considering $\alpha = 0.10$, (i) Nemenyi's [210] procedure rejects those hypotheses that have an unadjusted p value ≤ 0.004762 ; (ii) Holm's [202] procedure rejects those hypotheses that have an unadjusted p value ≤ 0.005556 ; (iii)

TABLE 11: Summary of performance failure and RTM scores of miscellaneous deep learning algorithms.

Models	GMN scores of performance failure					GMN of RTM
	ACC	PRS	RES	FIS	AUC	
AlexNet [81]	0.0951	0.0979	0.0634	0.0811	0.1036	2762.8
ResNet [75]	0.0578	0.0623	0.0403	0.0506	0.0559	3368.5
VGG16 [98]	0.0855	0.0666	0.0633	0.0658	0.0846	3567.2
VGG19 [98]	0.0672	0.0652	0.0543	0.0603	0.0740	4059.4
Inception V3 [88]	0.0704	0.0554	0.0499	0.0678	0.0761	3597.3
SqueezeNet [184]	0.1142	0.0964	0.0941	0.0969	0.1297	3795.3
DenTnet [ours]	0.0537	0.0650	0.0351	0.0469	0.0535	3196.3

TABLE 12: Average ranking of each algorithm using nonparametric statistical tests. The best results are shown in bold.

Algorithms	Multiple comparison tests		
	Friedman ranking [196]	Friedman's aligned ranking [198]	Quade ranking [199]
AlexNet [81]	5.3333	26.0000	4.6189
ResNet [75]	2.1667	09.0000	2.2857
VGG16 [98]	4.6667	27.8333	4.6191
VGG19 [98]	4.0000	21.6667	4.3333
Inception V3 [88]	3.6667	22.1667	4.0952
SqueezeNet [184]	6.6667	36.6667	6.6667
DenTnet [ours]	1.5000	07.1667	1.3809
Various statistics	24.500000	23.102557	5.274194
<i>p</i> value	0.000422	0.000763	0.000820

FIGURE 10: Plotting of average rankings data from Table 12, where each value x is plotted as $1/x$ to visualize the highest ranking with the tallest bar.

Shaffer's [211] procedure rejects those hypotheses that have an unadjusted p value ≤ 0.004762 ; and (iv) Bergmann's [212] procedure rejects those hypotheses of AlexNet [81] versus DenTnet [ours], ResNet [75] versus SqueezeNet [184], and SqueezeNet [184] versus DenTnet [ours].

7.4. Critical Distance Diagram from Nemenyi [210] Test. Nemenyi [210] test is very conservative with a low power, and hence it is not a recommended choice in practice [215]. Nevertheless, it has a unique advantage of having an associated plot to demonstrate the results of fair comparison.

Figure 11 depicts the Nemenyi [210] post hoc critical distance diagrams at three distinct levels of significance α values. If the distance between algorithms is less than the critical distance, then there is no statistically significant difference between them. The diagrams in Figures 11(a) and 11(b) associated with $\alpha = 0.10$ with the critical distance of 3.3588 and with $\alpha = 0.05$ with the critical distance of 3.6768, respectively, are identical, whereas the diagram in Figure 11(c) related to $\alpha = 0.01$ with the critical distance of 4.3054 is different. Any two algorithms are considered as significantly different if their performance variation is greater than the critical distance. To this end, from Figure 11, it is noticeable that, at $\alpha = 0.01$, both SqueezeNet [184] versus DenTnet [ours] and SqueezeNet [184] versus ResNet [75] are remarkably different, while other pairs are not remarkably divergent as their performance differences are less than 4.3054. As compared to ResNet [75], DenTnet [ours] differs from SqueezeNet [184] by a greater distance. On the other hand, SqueezeNet [184] versus DenTnet [ours] and AlexNet [81] versus DenTnet [ours] are significantly different at both $\alpha = 0.10$ and $\alpha = 0.05$, whereas SqueezeNet [184] versus ResNet [75] is significantly dissimilar at those α values. Straightforwardly, DenTnet [ours] is outstandingly unlike both SqueezeNet [184] and AlexNet [81], but ResNet [75] is not outstandingly unlike AlexNet [81]. This implies that the method of DenTnet [ours] outperforms that of ResNet [75], which also agrees with the finding in Figure 10.

7.5. Reasons of Superiority. In this study, DenseNet [67] was a great choice as it was very compact and deep. It used less training parameters and reduced the risk of model

TABLE 13: Results achieved on post hoc comparisons for adjusted p values, with $\alpha = 0.05$ and $\alpha = 0.10$.

Index	Algorithms	p values	$\alpha = 0.05$		$\alpha = 0.10$	
			Holm [202]	Shaffer [211]	Holm [202]	Shaffer [211]
1	VGG19 [98] versus Inception V3 [88]	0.789268	0.050000	0.050000	0.100000	0.100000
2	ResNet [75] versus DenTnet [ours]	0.592980	0.025000	0.025000	0.050000	0.050000
3	VGG16 [98] versus VGG19 [98]	0.592980	0.016667	0.016667	0.033333	0.033333
4	AlexNet [81] versus VGG16 [98]	0.592980	0.012500	0.016667	0.025000	0.033333
5	VGG16 [98] versus Inception V3 [88]	0.422678	0.010000	0.016667	0.020000	0.033333
6	AlexNet [81] versus SqueezeNet [184]	0.285049	0.008333	0.008333	0.016667	0.016667
7	AlexNet [81] versus VGG19 [98]	0.285049	0.007143	0.007143	0.014286	0.014286
8	ResNet [75] versus Inception V3 [88]	0.229102	0.006250	0.006250	0.012500	0.012500
9	AlexNet [81] versus Inception V3 [88]	0.181449	0.005556	0.005556	0.011111	0.011111
10	ResNet [75] versus VGG19 [98]	0.141579	0.005000	0.005000	0.010000	0.010000
11	VGG16 [98] versus SqueezeNet [184]	0.108809	0.004545	0.004545	0.009091	0.009091
12	Inception V3 [88] versus DenTnet [ours]	0.082352	0.004167	0.004167	0.008333	0.008333
13	VGG19 [98] versus DenTnet [ours]	0.045021	0.003846	0.003846	0.007692	0.007692
14	ResNet [75] versus VGG16 [98]	0.045021	0.003571	0.003571	0.007143	0.007143
15	VGG19 [98] versus SqueezeNet [184]	0.032509	0.003333	0.003333	0.006667	0.006667
16	Inception V3 [88] versus SqueezeNet [184]	0.016157	0.003125	0.003333	0.006250	0.006667
17	VGG16 [98] versus DenTnet [ours]	0.011118	0.002941	0.003333	0.005882	0.006667
18	AlexNet [81] versus ResNet [75]	0.011118	0.002778	0.003333	0.005556	0.006667
19	AlexNet [81] versus DenTnet [ours]	0.002116	0.002632	0.003333	0.005263	0.006667
20	ResNet [75] versus SqueezeNet [184]	0.000309	0.002500	0.003333	0.005000	0.006667
21	SqueezeNet [184] versus DenTnet [ours]	0.000034	0.002381	0.002381	0.004762	0.004762

TABLE 14: Adjusted p values for various tests considering DenTnet [ours] as control method.

Tests	Algorithms	$1 \times N$ post hoc procedures and p values								
		Not adjusted	1-2 step-procedure		Step-down procedures			Step-up procedures		
		p values	P_{Bonf} [201]	P_{Li} [209]	P_{Holm} [202]	P_{Hol} [206]	P_{Firm} [208]	P_{Hoch} [203]	P_{Hom} [204]	P_{Rom} [207]
Friedman [196]	SqueezeNet [184]	0.000034	0.000206	0.000084	0.000206	0.000206	0.000206	0.000206	0.000206	0.000196
	AlexNet [81]	0.002116	0.012694	0.005171	0.010578	0.010533	0.006333	0.010578	0.010578	0.010060
	VGG16 [98]	0.011118	0.066705	0.026588	0.044470	0.043734	0.022112	0.044470	0.044470	0.042403
	VGG19 [98]	0.045021	0.270125	0.099595	0.135063	0.129073	0.066765	0.135063	0.123528	0.135063
	Inception V3 [88]	0.082352	0.494113	0.168281	0.164704	0.157923	0.097990	0.164704	0.164704	0.164704
	ResNet [75]	0.592980	3.557881	0.592980	0.592980	0.592980	0.592980	0.592980	0.59298	0.592980
F. al. rank [198]	SqueezeNet [184]	0.000031	0.000187	0.000152	0.000187	0.000187	0.000187	0.000187	0.000187	0.000178
	VGG16 [98]	0.003525	0.021147	0.016964	0.017623	0.017499	0.010536	0.017623	0.017623	0.016759
	AlexNet [81]	0.007837	0.047023	0.036954	0.031348	0.030982	0.015613	0.031348	0.031348	0.029891
	Inception V3 [88]	0.034193	0.205155	0.143404	0.102578	0.099110	0.050848	0.081277	0.068385	0.081277
	VGG19 [98]	0.040638	0.243830	0.165952	0.102578	0.099110	0.050848	0.081277	0.081277	0.081277
	ResNet [75]	0.795758	4.774545	0.795758	0.795758	0.795758	0.795758	0.795758	0.795758	0.795758
Quade [199]	SqueezeNet [184]	0.027879	0.167272	0.086779	0.167272	0.156038	0.156038	0.167272	0.167272	0.159049
	AlexNet [81]	0.177939	1.067632	0.377531	0.889693	0.624577	0.444463	0.517618	0.388213	0.517618
	VGG16 [98]	0.177939	1.067632	0.377531	0.889693	0.624577	0.444463	0.517618	0.388213	0.517618
	VGG19 [98]	0.219348	1.316086	0.427803	0.889693	0.624577	0.444463	0.517618	0.438695	0.517618
	Inception V3 [88]	0.258809	1.552853	0.468693	0.889693	0.624577	0.444463	0.517618	0.517618	0.517618
	ResNet [75]	0.706617	4.239701	0.706617	0.889693	0.706617	0.706617	0.706617	0.706617	0.706617

overfitting and improved the learning rate. In the dense block of DenTnet, the outputs from the previous layers were concatenated instead of using the summation. This type of concatenation helped to markedly speed up the processing

of data for large number of columns. The dense block of DenTnet contained convolution and nonlinear layers, which applied several optimization techniques (e.g., dropout and BN). DenTnet scaled to hundreds of layers, while exhibiting

TABLE 15: Adjusted p values of tests for multiple comparisons among all methods.

Index	Hypothesis	$N \times N$ post hoc procedures and p values				
		Unadjusted	Nemenyi [210]	Holm [202]	Shaffer [211]	Bergmann [212]
1	SqueezeNet [184] versus DenTnet [ours]	0.000034	0.000721	0.000721	0.000721	0.000721
2	ResNet [75] versus SqueezeNet [184]	0.000309	0.006479	0.006171	0.004628	0.004628
3	AlexNet [81] versus DenTnet [ours]	0.002116	0.044428	0.040197	0.031734	0.031734
4	AlexNet [81] versus ResNet [75]	0.011118	0.233469	0.200116	0.166763	0.111176
5	VGG16 [98] versus DenTnet [ours]	0.011118	0.233469	0.200116	0.166763	0.122293
6	Inception V3 [88] versus SqueezeNet [184]	0.016157	0.339296	0.258511	0.242354	0.177726
7	VGG19 [98] versus SqueezeNet [184]	0.032509	0.682698	0.487642	0.487642	0.292585
8	ResNet [75] versus VGG16 [98]	0.045021	0.945439	0.630292	0.495230	0.315146
9	VGG19 [98] versus DenTnet [ours]	0.045021	0.945439	0.630292	0.495230	0.405188
10	Inception V3 [88] versus DenTnet [ours]	0.082352	1.729397	0.988227	0.905874	0.494113
11	VGG16 [98] versus SqueezeNet [184]	0.108809	2.284998	1.196904	1.196904	0.652857
12	ResNet [75] versus VGG19 [98]	0.141579	2.973156	1.415789	1.415789	0.652857
13	AlexNet [81] versus Inception V3 [88]	0.181449	3.810433	1.633043	1.633043	1.270144
14	ResNet [75] versus Inception V3 [88]	0.229102	4.811140	1.832815	1.633043	1.270144
15	AlexNet [81] versus VGG19 [98]	0.285049	5.986038	1.995346	1.995346	1.270144
16	AlexNet [81] versus SqueezeNet [184]	0.285049	5.986038	1.995346	1.995346	1.425247
17	VGG16 [98] versus Inception V3 [88]	0.422678	8.876240	2.113390	2.113390	1.690712
18	AlexNet [81] versus VGG16 [98]	0.592980	12.452582	2.371920	2.371920	1.778940
19	VGG16 [98] versus VGG19 [98]	0.592980	12.452582	2.371920	2.371920	1.778940
20	ResNet [75] versus DenTnet [ours]	0.592980	12.452582	2.371920	2.371920	1.778940
21	VGG19 [98] versus Inception V3 [88]	0.789268	16.574629	2.371920	2.371920	1.778940

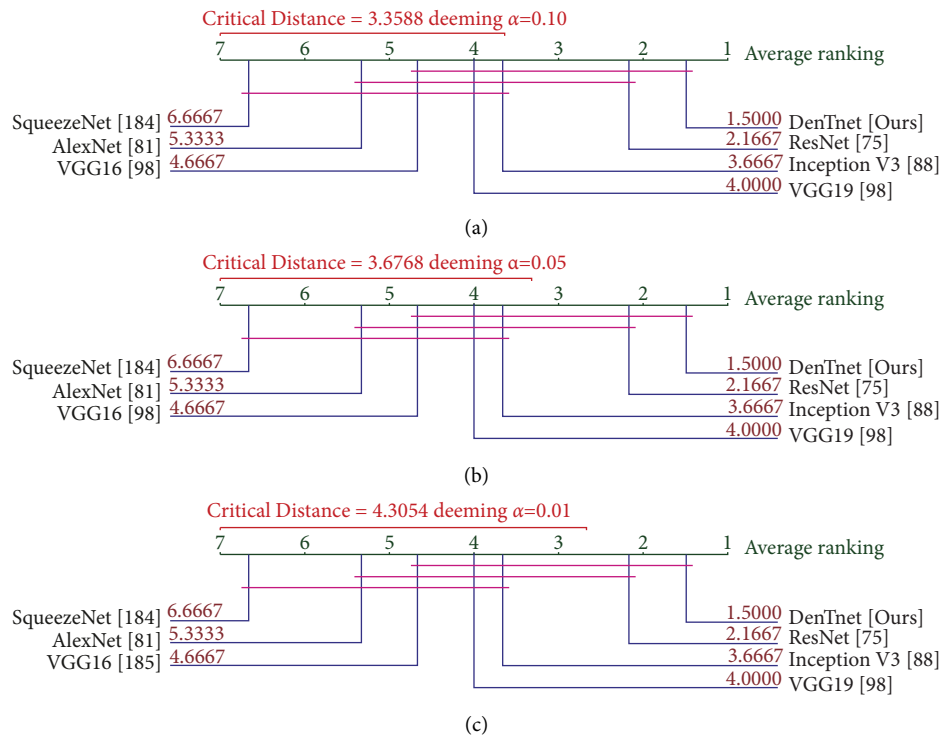


FIGURE 11: Nemenyi [210] post hoc critical distance diagrams for three α values using data in Table 11.

no optimization difficulties. Overall, this model was applied to a very large number of preprocessed augmented images from BreaKHis [33], Malaria [191], SkinCancer [193], and CovidXray [192] datasets. To the best of our knowledge, no other studies in the literature had such an edge. Additionally, the use of data augmentation approach in this study

positively affected the performance of the model due to expansion in the size of training data, which is the foremost requirement of a deep network for its proper working. Our DenTnet was well trained through various parameters' tuning. For example, in the case of BreaKHis [33], unlike other existing models, our model was trained on all the

magnifications combined (40x, 100x, 200x, and 400x) to avoid any loss of generality.

In sum and substance, based on the aforementioned experimental and nonparametric statistical test results, it is, therefore, possible to conclude that the proposed DenTnet [ours] outperformed AlexNet [81], ResNet [75], VGG16 [98], VGG19 [98], Inception V3 [88], and SqueezeNet [184] in terms of computational speed. Significantly, the accuracy achieved by the proposed DenTnet [ours] surpassed those of existing state-of-the-art models in classifying images of the BreakHis [33], Malaria [191], SkinCancer [193], and the CovidXray [192] dataset.

7.6. Limitation of Proposed Model and Methodology.

Despite these promising results, questions remain as to whether the proposed DenTnet model could be utilized to classify categorical images. Moreover, DenTnet was tested with one breast cancer dataset (i.e., BreakHis [33]) only. Although the generalization ability of DenTnet with three non-breast-cancer-related datasets was studied in Section 6, it is unknown whether DenTnet can generalize to other state-of-the-art breast cancer datasets. Future work should, therefore, investigate the efficacy and generalizability of DenTnet with datasets along with multiclass labels, as well as other publicly available breast cancer datasets (e.g., the most recently introduced MITNET dataset [216]).

The classification effect of breast cancer histopathological images of any deep learning methodology is related to the features and many studies predominantly focused on how to develop good feature descriptors and better extract features. Different from traditional handcrafted feature-based models, DenTnet can automatically extract more abstract features. Nevertheless, it is worth noting that although the proposed DenTnet has addressed the cross-domain problem by utilizing the transfer learning approach, features extracted in the methodology are solely deep-network-based features, which are extracted by feeding images directly to the model. However, feeding deep models directly with images would not generalize as the models consider color distribution of an image. It is understood that local information can be captured from color images using Local Binary Pattern (LBP) [217]. Therefore, future work can use multiple types of features by combining the features extracted by the proposed method with LBP features to address this issue.

8. Conclusion

We presented that, for classifying breast cancer histopathological images, the most popular training-testing ratio was 70%: 30%, while the best performance was indicated by the training-testing ratio of 80%: 20%. We proposed a novel approach named DenTnet to classify histopathology images using training-testing ratio of 80%: 20%. DenTnet achieved a very high classification accuracy on the BreakHis dataset. Several impediments of existing state-of-the-art methods including the requirement of high computation and the utilization of the identical feature distribution were

attenuated. To test the generalizability of DenTnet, we conducted experiments on three additional datasets (Malaria, SkinCancer, and CovidXray) with varying difficulties. Experimental results on all four datasets demonstrated that DenTnet achieved a better performance in terms of accuracy and computational speed than a large number of effective state-of-the-art classification methods (AlexNet, ResNet, VGG16, VGG19, InceptionV3, and SqueezeNet). These findings contributed to our understanding of how a lightweight model could be used to improve the accuracy and accelerate the learning process of images, including histopathology image classification on using the wild state-of-the-art datasets. Future work shall investigate the efficacy of DenTnet on datasets with multiclass labels.

Abbreviations

<i>BreaKHis</i> :	Breast cancer histopathological image classification
<i>BACH</i> :	Breast cancer histology images
<i>CNN</i> :	Convolutional neural network
<i>DenseNet</i> :	Dense convolutional Network
<i>ResNet</i> :	Residual network
<i>VGG</i> :	Visual geometry group at the University of Oxford
<i>DDSM</i> :	Digital database for screening mammography
<i>CBIS – DDSM</i> :	Curated breast imaging subset of <i>DDSM</i>
<i>SVM</i> :	Support vector machine
<i>H&E</i> :	Haematoxylin and eosin
<i>BN</i> :	Batch normalization
<i>ReLU</i> :	Rectified linear unit
<i>ROC</i> :	Receiver operating characteristic
<i>AUC</i> :	Area under the ROC curve
<i>ACC</i> :	Accuracy
<i>GMN</i> :	Geometric mean
<i>PRS</i> :	Precision score
<i>RES</i> :	Recall score
<i>F1S</i> :	F1 score
<i>RTM</i> :	Runtime.

Data Availability

The four following publicly available datasets were used in this study: BreakHis [33] (<https://www.kaggle.com/ambarish/breakhis>), Malaria [191] (<https://www.kaggle.com/iarunava/cell-images-for-detecting-malaria>), CovidXray [192] (<https://github.com/ieee8023/covid-chestxray-dataset>), and SkinCancer [193] (<https://www.kaggle.com/fanconic/skin-cancer-malignant-vs-benign>).

Conflicts of Interest

The authors have no conflicts of interest to declare.

References

- [1] Who, *Breast Cancer Now Most Common Form of Cancer: WHO Taking Action*, World Health Organization, Geneva, Switzerland, 2021.

- [2] A. A. Ewees, L. Abualigah, D. Yousri et al., "Improved Slime Mould Algorithm Based on Firefly Algorithm for Feature Selection: A Case Study on QSAR Model," *Engineering with Computers*, vol. 38, pp. 1–15, 2021.
- [3] A. A. Ewees, Z. Y. Algamal, L. Abualigah et al., "A cox proportional-hazards model based on an improved aquila optimizer with whale optimization algorithm operators," *Mathematics*, vol. 10, no. 8, p. 1273, 2022.
- [4] M. AbdElaziz, A. A. Ewees, D. Yousri, L. M. Abualigah, and M. A. A. Al-qaness, "Modified marine predators algorithm for feature selection: case study metabolomics," *Knowledge and Information Systems*, vol. 64, no. 1, pp. 261–287, 2022.
- [5] L. Abualigah, K. H. Almotairi, M. A. Al-qaness et al., "Efficient text document clustering approach using multi-search Arithmetic Optimization Algorithm," *Knowledge-Based Systems*, vol. 248, Article ID 108833, 2022.
- [6] M. Jannesari, M. Habibzadeh, H. Aboulkheyr et al., "Breast cancer histopathological image classification: a deep learning approach," in *Proceedings of the International Conference on Bioinformatics and Biomedicine*, pp. 2405–2412, BIBM, Madrid, Spain, June 2018.
- [7] D. Bardou, K. Zhang, and S. M. Ahmad, "Classification of breast cancer based on histology images using convolutional neural networks," *IEEE Access*, vol. 6, Article ID 24680, 2018.
- [8] S. H. Kassani, P. H. Kassani, M. J. Wesolowski, K. A. Schneider, and R. Deters, "Classification of histopathological biopsy images using ensemble of deep learning networks," in *Proceedings of the Annual International Conference on Computer Science and Software Engineering (CASCON)*, pp. 92–99, Markham, Ontario, July 2019.
- [9] K. Kumar and A. C. S. Rao, "Breast cancer classification of image using convolutional neural network," in *Proceedings of the International Conference on Recent Advances in Information Technology (RAIT)*, pp. 1–6, Dhanbad, India, October 2018.
- [10] M. Z. Alom, C. Yakopcic, M. 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, 2019.
- [11] K. Das, S. Conjeti, A. G. Roy, J. Chatterjee, and D. Sheet, "Multiple instance learning of deep convolutional neural networks for breast histopathology whole slide classification," in *Proceedings of the Int. Symposium on Biomedical Imaging (ISBI)*, pp. 578–581, New York, NY, USA, July 2018.
- [12] A. A. Nahid and Y. Kong, "Histopathological breast-image classification using concatenated R–G–B histogram information," *Annals of Data Science*, vol. 6, no. 3, pp. 513–529, 2019.
- [13] B. Du, Q. Qi, H. Zheng, Y. Huang, and X. Ding, "Breast cancer histopathological image classification via deep active learning and confidence boosting. Artificial neural networks and machine learning (ICANN)," in *Proceedings of the 27th international conference on artificial neural networks*, vol. 11140, pp. 109–116, Greece, May 2018.
- [14] Z. Gandomkar, P. C. Brennan, and C. Mello-Thoms, "MuDeRN: multi-category classification of breast histopathological image using deep residual networks," *Artificial Intelligence in Medicine*, vol. 88, pp. 14–24, 2018.
- [15] V. Gupta and A. Bhavsar, "Sequential modeling of deep features for breast cancer histopathological image classification," in *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops*, pp. 2254–2261, Salt Lake City, UT, USA, June 2018.
- [16] 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, Article ID 0214587, 2019.
- [17] Y. Benhammou, S. Tabik, B. Achchab, and F. Herrera, "A first study exploring the performance of the state-of-the-art CNN model in the problem of breast cancer," in *Proceedings of the International Conference on Learning and Optimization Algorithms: Theory and Applications (LOPAL)*, no. 1–47, p. 47, Rabat, Morocco, June 2018.
- [18] P. J. Sudharshan, C. Petitjean, F. A. Spanhol, L. E. Oliveira, L. Heutte, and P. Honeine, "Multiple instance learning for histopathological breast cancer image classification," *Expert Systems with Applications*, vol. 117, pp. 103–111, 2019.
- [19] S. Sharma and R. Mehra, "Breast cancer histology images classification: training from scratch or transfer learning?" *ICT Express*, vol. 4, pp. 247–254, 2018.
- [20] S. Cascianelli, R. Bello-Cerezo, F. Bianconi et al., "Dimensionality reduction strategies for CNN-based classification of histopathological images," *Intelligent Interactive Multimedia Systems and Services*, pp. 21–30, Springer International Publishing: Cham, New York, NY, USA, 2018.
- [21] Y. Song, H. Chang, H. Huang, and W. Cai, "Supervised intra-embedding of Fisher vectors for histopathology image classification," in *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 99–106, Quebec City, Canada, March 2017.
- [22] B. Wei, Z. Han, X. He, and Y. Yin, "Deep learning model based breast cancer histopathological image classification," in *Proceedings of the International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, pp. 348–353, Chengdu, April 2017.
- [23] K. Das, S. P. K. Karri, A. G. Roy, J. Chatterjee, and D. Sheet, "Classifying histopathology whole-slides using fusion of decisions from deep convolutional network on a collection of random multi-views at multi-magnification," in *Proceedings of the International Symposium on Biomedical Imaging (ISBI)*, pp. 1024–1027, Melbourne, Australia, March 2017.
- [24] Y. Song, J. J. Zou, H. Chang, and W. Cai, "Adapting Fisher vectors for histopathology image classification," in *Proceedings of the International Symposium on Biomedical Imaging (ISBI)*, pp. 600–603, Melbourne, Australia, April 2017.
- [25] V. Gupta and A. Bhavsar, "Partially-Independent framework for breast cancer histopathological image classification," in *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1123–1130, Long Beach, CA, USA, June 2019.
- [26] M. Togacar, K. B. Ozkurt, B. Ergen, and Z. Comert, "BreastNet: a novel convolutional neural network model through histopathological images for the diagnosis of breast cancer," *Physica A: Statistical Mechanics and Its Applications*, vol. 545, Article ID 123592, 2020.
- [27] P. Wang, J. Wang, Y. Li, P. Li, L. Li, and M. Jiang, "Automatic classification of breast cancer histopathological images based on deep feature fusion and enhanced routing," *Biomedical Signal Processing and Control*, vol. 65, Article ID 102341, 2021.
- [28] 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, 2020.

- [29] L. Li, X. Pan, H. Yang et al., "Multi-task deep learning for fine-grained classification and grading in breast cancer histopathological images," *Multimedia Tools and Applications*, vol. 79, no. 21-22, pp. 14509-14528, 2020.
- [30] D. Albashish, R. Al-Sayyed, A. Abdullah, M. H. Ryalat, and N. Ahmad Almansour, "Deep CNN Model Based on VGG16 for Breast Cancer Classification," in *Proceedings of the International Conference on Information Technology (ICIT)*, pp. 805-810, Amman, Jordan, July 2021.
- [31] F. Parvin and M. A. Mehedi Hasan, "A comparative study of different types of convolutional neural networks for breast cancer histopathological image classification," in *Proceedings of the 2020 IEEE Region 10 Symposium (TENSymp)*, pp. 945-948, Dhaka, Bangladesh, June 2020.
- [32] Q. Qi, Y. Li, J. Wang et al., "Label-efficient breast cancer histopathological image classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 5, pp. 2108-2116, 2019.
- [33] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using Convolutional Neural Networks," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, pp. 2560-2567, Vancouver, BC, 2016.
- [34] F. A. Spanhol, L. S. Oliveira, P. R. Cavalin, C. Petitjean, and L. Heutte, A. B. Banff, "Deep features for breast cancer histopathological image classification," in *Proceedings of the International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1868-1873, Canada, CA, USA, 2017.
- [35] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, "Breast cancer multi-classification from histopathological images with structured deep learning model," *Scientific Reports*, vol. 7, no. 1, p. 4172, 2017.
- [36] R. Man, P. Yang, and B. Xu, "Classification of breast cancer histopathological images using discriminative patches screened by generative adversarial networks," *IEEE Access*, vol. 8, Article ID 155362, 2020.
- [37] A. Kumar, S. K. Singh, S. Saxena et al., "Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer," *Information Sciences*, vol. 508, pp. 405-421, 2020.
- [38] N. Bayramoglu, J. Kannala, and J. Heikkilä, "Deep learning for magnification independent breast cancer histopathology image classification," in *Proceedings of the International Conference on Pattern Recognition (ICPR)*, pp. 2440-2445, Cancun, Mexico, December 2016.
- [39] J. Sun and A. Binder, "Comparison of deep learning architectures for H&E histopathology images," in *Proceedings of the IEEE Conference on Big Data and Analytics (ICBDA)*, pp. 43-48, Kuching, Malaysia, November 2017.
- [40] C. Kaushal and A. Singla, "Automated segmentation technique with self-driven post-processing for histopathological breast cancer images," *CAAI Trans. Intell. Technol.* vol. 5, no. 4, pp. 294-300, 2020.
- [41] M. Talo, "Convolutional neural networks for multi-class histopathology image classification," 2019, <https://arxiv.org/ftp/arxiv/papers/1903/1903.10035.pdf>.
- [42] H. Elmannai, M. Hamdi, and A. AlGarni, "Deep learning models combining for breast cancer histopathology image classification," *International Journal of Computational Intelligence Systems*, vol. 14, no. 1, pp. 1003-1013, 2021.
- [43] 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, 2020.
- [44] D. A. Ragab, M. Sharkas, S. Marshall, and J. Ren, "Breast cancer detection using deep convolutional neural networks and support vector machines," *PeerJ*, vol. 7, 2019.
- [45] F. P. Romero, A. Tang, and S. Kadoury, "Multi-level batch normalization in deep networks for invasive ductal carcinoma cell discrimination in histopathology images," in *Proceedings of the International Symposium on Biomedical Imaging (ISBI)*, pp. 1092-1095, Venice, Italy, July 2019.
- [46] H. L. Minh, M. M. Van, and T. V. Lang, "Deep feature fusion for breast cancer diagnosis on histopathology images," in *Proceedings of the International Conference on Knowledge and Systems Engineering (KSE)*, pp. 1-6, Da Nang, Vietnam, September 2019.
- [47] Y. Liu, K. Gadeballi, M. Norouzi et al., "Detecting cancer metastases on gigapixel pathology images," 2017, <https://arxiv.org/abs/1703.02442>.
- [48] M. A. Alantari, S. M. Han, and T. S. Kim, "Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms," *Computer Methods and Programs in Biomedicine*, vol. 196, Article ID 105584, 2020.
- [49] H. Zhang, R. Wu, T. Yuan et al., "DE-Ada*: a novel model for breast mass classification using cross-modal pathological semantic mining and organic integration of multi-feature fusions," *Information Sciences*, vol. 539, pp. 461-486, 2020.
- [50] N. Brancati, G. De Pietro, D. Riccio, and M. Frucci, "Gigapixel histopathological image analysis using attention-based neural networks," *IEEE Access*, vol. 9, Article ID 87552, 2021.
- [51] E. Mahraban Nejad, L. S. Affendey, R. B. Latip, I. B. Ishak, and R. Banaeeyan, "Transferred semantic scores for scalable retrieval of histopathological breast cancer images," *International Journal of Multimedia Information Retrieval*, vol. 7, no. 4, pp. 241-249, 2018.
- [52] V. K. Reshma, N. Arya, S. S. Ahmad et al., "Detection of breast cancer using histopathological image classification dataset with deep learning techniques," *BioMed Research International*, vol. 202213 pages, Article ID 8363850, 2022.
- [53] H. Huang, X. Feng, J. Jiang, P. Chen, and S. Zhou, "Mask RCNN algorithm for nuclei detection on breast cancer histopathological images," *International Journal of Imaging Systems and Technology*, vol. 32, no. 1, pp. 209-217, 2022.
- [54] G. Jayandhi, J. Leena Jasmine, and S. Mary Joans, "Mammogram learning system for breast cancer diagnosis using deep learning SVM," *Computer Systems Science and Engineering*, vol. 40, no. 2, pp. 491-503, 2022.
- [55] S. Sharma and S. Kumar, "The Xception model: a potential feature extractor in breast cancer histology images classification," *ICT Express*, vol. 8, no. 1, pp. 101-108, 2022.
- [56] H. Zerouaoui and A. Idri, "Deep hybrid architectures for binary classification of medical breast cancer images," *Biomedical Signal Processing and Control*, vol. 71, Article ID 103226, 2022.
- [57] W. Zhi, H. W. F. Yeung, Z. Chen, S. M. Zandavi, Z. Lu, and Y. Y. Chung, "Using transfer learning with convolutional neural networks to diagnose breast cancer from histopathological images," *International Conference on Neural Information Processing (ICONIP), China*, vol. 10637, pp. 669-676, 2017.
- [58] J. Chang, J. Yu, T. Han, H. Chang, and E. Park, "A method for classifying medical images using transfer learning: a pilot study on histopathology of breast cancer," in *Proceedings of the International Conference on E-Health Networking*,

- Applications and Services (Healthcom)*, pp. 1–4, Dalian, China, June 2017.
- [59] S. A. Hassan, M. S. Sayed, M. I. Abdalla, and M. A. Rashwan, “Breast cancer masses classification using deep convolutional neural networks and transfer learning,” *Multimedia Tools and Applications*, vol. 79, no. 41–42, pp. 30735–30768, 2020.
- [60] M. F. I. Soumik, A. Z. B. Aziz, and M. A. Hossain, “Improved transfer learning based deep learning model for breast cancer histopathological image classification,” in *Proceedings of the 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI)*, pp. 1–4, Rajshahi, Bangladesh, June 2021.
- [61] H. Mahmoud, A. H. Alharbi, and D. S. Khafga, “Breast cancer classification using deep convolution neural network with transfer learning,” *Intelligent Automation & Soft Computing*, vol. 29, no. 3, pp. 803–814, 2021.
- [62] C. Munien and S. Viriri, “Classification of hematoxylin and eosin-stained breast cancer histology microscopy images using transfer learning with EfficientNets,” *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–17, Article ID 5580914, 2021.
- [63] S. Boumaraf, X. Liu, Z. Zheng, X. Ma, and C. Ferkous, “A new transfer learning based approach to magnification dependent and independent classification of breast cancer in histopathological images,” *Biomedical Signal Processing and Control*, vol. 63, Article ID 102192, 2021.
- [64] A. Saber, M. Sakr, O. M. Abo-Seida, A. Keshk, and H. Chen, “A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique,” *IEEE Access*, vol. 9, pp. 71194–71209, 2021.
- [65] S. Soltane, S. Alshreef, and S. MSerag Eldin, “Classification and diagnosis of lymphoma’s histopathological images using transfer learning,” *Computer Systems Science and Engineering*, vol. 40, no. 2, pp. 629–644, 2022.
- [66] M. Alruwaili and W. Gouda, “Automated breast cancer detection models based on transfer learning,” *Sensors*, vol. 22, no. 3, p. 876, 2022.
- [67] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, Honolulu, HI, USA, July 2017.
- [68] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [69] S. Z. Ramadan, “Methods used in computer-aided diagnosis for breast cancer detection using mammograms: a review,” *Journal of Healthcare Engineering*, vol. 2020, pp. 1–21, Article ID 9162464, 2020.
- [70] Y. Benhammou, B. Achchab, F. Herrera, and S. Tabik, “BreakHis based breast cancer automatic diagnosis using deep learning: taxonomy, survey and insights,” *Neurocomputing*, vol. 375, pp. 9–24, 2020.
- [71] D. A. Zebari, D. A. Ibrahim, D. Q. Zeebaree et al., “Systematic review of computing approaches for breast cancer detection based computer aided diagnosis using mammogram images,” *Applied Artificial Intelligence*, vol. 35, no. 15, pp. 2157–2203, 2021.
- [72] K. R. Weiss, T. M. Khoshgoftaar, and D. Wang, “A survey of transfer learning,” *J. Big Data*, vol. 3, no. 1, p. 9, 2016.
- [73] M. A. Mohammed, B. Al-Khateeb, A. N. Rashid, D. A. Ibrahim, M. K. Abd Ghani, and S. A. Mostafa, “Neural network and multi-fractal dimension features for breast cancer classification from ultrasound images,” *Computers & Electrical Engineering*, vol. 70, pp. 871–882, 2018.
- [74] WHO, “Breast Cancer,” 2021, <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>.
- [75] K. He, X. Zhang, S. Ren, and J. Sun, “Identity Mappings in Deep Residual Networks,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, vol. 9908, pp. 630–645, Netherlands, Europe, 2016.
- [76] N. B. C. Foundation, *Biopsy*, The National Breast Cancer Foundation, Texas, TX, USA, 2018.
- [77] Q. Abbas, “DeepCAD: a computer-aided diagnosis system for mammographic masses using deep invariant features,” *Computers*, vol. 5, no. 4, p. 28, 2016.
- [78] G. Aresta, T. Araujo, S. Kwok et al., “BACH: grand challenge on breast cancer histology images,” *Medical Image Analysis*, vol. 56, pp. 122–139, 2019.
- [79] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [80] A. Chan and J. A. Tuszynski, “Automatic prediction of tumour malignancy in breast cancer with fractal dimension,” *Royal Society Open Science*, vol. 3, no. 12, Article ID 160558, 2016.
- [81] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems 25,” in *Proceedings of the 26th Annual Conference on Neural Information Processing Systems 2012*, pp. 1106–1114, Lake Tahoe, Nevada, USA, October 2012.
- [82] H. Bay, T. Tuytelaars, and L. V. Gool, “SURF: Speeded Up Robust Features,” in *Proceedings of the European Conference on Computer Vision ECCV*, vol. 3951, pp. 404–417, Graz, Austria, July 2006.
- [83] Z. Guo, L. Zhang, and D. Zhang, “Rotation invariant texture classification using LBP variance (LBPV) with global matching,” *Pattern Recognition*, vol. 43, no. 3, pp. 706–719, 2010.
- [84] S. F. University, “Digital Database for Screening Mammography,” 2006, <http://www.eng.usf.edu/cvprg/mammography/database.html>.
- [85] J. Suckling and J. Parker, “Mammographic Image Analysis Society (MIAS) Database v1.21 [Dataset],” 2015, <https://www.repository.cam.ac.uk/handle/1810/250394>.
- [86] C. Szegedy, W. Liu, Y. Jia et al., “Going deeper with convolutions,” in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, Boston, MA, USA, July 2015.
- [87] L. Shamir, N. Orlov, D. Mark Eckley, T. J. Macura, and I. G. Goldberg, “IICBU 2008: a proposed benchmark suite for biological image analysis,” *Medical, & Biological Engineering & Computing*, vol. 46, no. 9, pp. 943–947, 2008.
- [88] C. Szegedy, W. Liu, Y. Jia et al., “Going deeper with convolutions,” in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, Boston, MA, USA, July 2015.
- [89] M. A. Kahya, W. Al-Hayani, and Z. Y. Algamal, “Classification of breast cancer histopathology images based on adaptive sparse support vector machine,” *Journal of Applied Mathematics and Bioinformatics*, vol. 7, pp. 49–69, 2017.
- [90] V. Gupta and A. Bhavsar, “An integrated multi-scale model for breast cancer histopathological image classification with joint colour-texture features,” in *Proceedings of the International Conference on Computer Analysis of Images and Patterns (CAIP)*, vol. 10425, pp. 354–366, Ystad, Sweden, August 2017.

- [91] N. Dhungel, G. Carneiro, and A. P. Bradley, "A deep learning approach for the analysis of masses in mammograms with minimal user intervention," *Medical Image Analysis*, vol. 37, pp. 114–128, 2017.
- [92] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast: toward a full-field digital mammographic database," *Academic Radiology*, vol. 19, no. 2, pp. 236–248, 2012.
- [93] Y. Jia, E. Shelhamer, J. Donahue et al., "Caffe: Convolutional Architecture for Fast Feature Embedding," in *Proceedings of the ACM Int. Conf. on Multimedia (MM)*, pp. 675–678, Florida, FL, USA, 2014.
- [94] S. Kaymak, A. Helwan, and D. Uzun, "Breast cancer image classification using artificial neural networks. Procedia Computer Science," in *Proceedings of the International Conference on Theory and Application of Soft Computing, Computing with Words and Perception (ICSCCW)*, vol. 120, pp. 126–131, Budapest, Hungary, June 2017.
- [95] R. Hecht-Nielsen, "Theory of the backpropagation neural network," *Neural Networks*, vol. 1, pp. 445–448, 1988.
- [96] C. Bishop, "Improving the generalization properties of radial basis function neural networks," *Neural Computation*, vol. 3, no. 4, pp. 579–588, 1991.
- [97] Camelyon16, "Challenge on Cancer Metastases Detection in Lymph Node," 2016, <https://camelyon16.grand-challenge.org>.
- [98] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, San Diego, CA, USA, July 2015.
- [99] S. J. Stanford, "Tissue Microarray Database," 2021, <https://tma.im/cgi-bin/home.pl>.
- [100] A. A. Nahid, A. Mikaelian, and Y. Kong, "Histopathological breast-image classification with restricted Boltzmann machine along with backpropagation," *Biomedical Research*, vol. 29, pp. 2068–2077, 2018.
- [101] W. Lu, C. S. Leung, and J. Sum, "Analysis on noisy Boltzmann machines and noisy restricted Boltzmann machines," *IEEE Access*, vol. 9, pp. 112955–112965, 2021.
- [102] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Trans. Syst. Man Cybern.* vol. 8, no. 6, pp. 460–473, 1978.
- [103] J. A. Badejo, E. Adetiba, A. Akinrinmade, and M. B. Akanle, "Medical image classification with hand-designed or machine-designed texture descriptors: a performance evaluation," *International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO), Granada, Spain*, vol. 10814, pp. 266–275, 2018.
- [104] P. Alirezazadeh, B. Hejrati, A. Monsef-Esfahani, and A. Fathi, "Representation learning-based unsupervised domain adaptation for classification of breast cancer histopathology images," *Biocybernetics and Biomedical Engineering*, vol. 38, no. 3, pp. 671–683, 2018.
- [105] J. Spencer and K. S. John, "Random sparse bit strings at the threshold of adjacency," in *Proceedings of the Annual symposium on theoretical aspects of computer science (STACS)*, vol. 1373, pp. 94–104, Paris, France, October 1998.
- [106] C. Y. Maa and M. A. Shanblatt, "Linear and quadratic programming neural network analysis," *IEEE Transactions on Neural Networks*, vol. 3, no. 4, pp. 580–594, 1992.
- [107] F. Giannakas, C. Troussas, A. Krouska, C. Sgouropoulou, and I. Voyiatzis, "XGBoost and deep neural network comparison: the case of teams' performance," *International Conference on Intelligent Tutoring Systems (ITS), Virtual Event*, vol. 12677, pp. 343–349, 2021.
- [108] D. S. Morillo, J. Gonzalez, M. G. Rojo, and J. Ortega, "Classification of breast cancer histopathological images using KAZE features," *Int. Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO), Spain*, vol. 10814, pp. 276–286, 2018.
- [109] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "KAZE Features," in *Proceedings of the European Conference on Computer Vision (ECCV)*, vol. 7577, pp. 214–227, Florence, Italy, June 2012.
- [110] S. Chatteraj and K. Vishwakarma, "Classification of histopathological breast cancer images using iterative VMD aided Zernike moments & textural signatures," 2018, <https://arxiv.org/abs/1801.04880>.
- [111] T. Theodoridis, K. Loumpoulias, N. Vretos, and P. Daras, "Zernike pooling: generalizing average pooling using zernike moments," *IEEE Access*, vol. 9, pp. 121128–121136, 2021.
- [112] N. Lassance and F. Vrans, "Minimum Rényi entropy portfolios," *Annals of Operations Research*, vol. 299, no. 1–2, pp. 23–46, 2021.
- [113] M. Rahimi and M. Mohammadi Anjedani, "A local view on the Hudetz correction of the Yager entropy of dynamical systems," *International Journal of General Systems*, vol. 48, no. 3, pp. 321–333, 2019.
- [114] Y. Zheng, Z. Jiang, H. Zhang et al., "Size-scalable content-based histopathological image retrieval from database that consists of WSIs," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 4, pp. 1278–1287, 2018.
- [115] R. W. Hamming, "Error detecting and error correcting codes," *Bell System Technical Journal*, vol. 29, no. 2, pp. 147–160, 1950.
- [116] R. Mukkamala, P. S. Neeraja, S. Pamidi, T. Babu, and T. Singh, "Deep PCANet framework for the binary categorization of breast histopathology images," in *Proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 105–110, India, September 2018.
- [117] T. H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "PCANet: a simple deep learning baseline for image classification?" *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5017–5032, 2015.
- [118] A. Rakhlin, A. Shvets, V. Igloukov, and A. A. Kalinin, "Deep convolutional neural networks for breast cancer histology image analysis," in *Proceedings of the International Conference on Image Analysis and Recognition (ICIAR)*, vol. 10882, pp. 737–744, Portugal, July 2018.
- [119] M. A. Almasni, M. A. Alantari, J. M. Park et al., "Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system," *Computer Methods and Programs in Biomedicine*, vol. 157, pp. 85–94, 2018.
- [120] A. G. Howard, M. Zhu, B. Chen et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017, <https://arxiv.org/abs/1704.04861>.
- [121] B. S. Veeling, J. Linmans, J. Winkens, T. Cohen, and M. Welling, "Rotation equivariant CNNs for digital pathology," *Int. Conf. on Medical Image Computing and Computer Assisted Intervention (MICCAI), Spain*, vol. 11071, pp. 210–218, 2018.
- [122] A. Pego and P. Aguiar, "Bioimaging Challenge 2015 Breast Histology Dataset," 2015, <http://www.bioimaging2015.ineb.up.pt/dataset.html>.

- [123] R. Lenz and P. L. Carmona, "Transform Coding of RGB-Histograms," in *Proceedings of the International Conference on Computer Vision Theory and Applications (VISAPP)*, pp. 117–124, Lisboa, Portugal, November 2009.
- [124] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation networks," in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7132–7141, Salt Lake City, UT, USA, May 2018.
- [125] D. M. Vo, N. Q. Nguyen, and S. W. Lee, "Classification of breast cancer histology images using incremental boosting convolution networks," *Information Sciences*, vol. 482, pp. 123–138, 2019.
- [126] M. Babaie, S. Kalra, A. Sriram et al., "Classification and Retrieval of Digital Pathology Scans," in *Proceedings of the A New Dataset. Conference on Computer Vision and Pattern Recognition Workshops*, pp. 760–768, CVPR Workshops, Honolulu, HI, USA, July 2017.
- [127] X. Li, M. Radulovic, K. Kanjer, and K. N. Plataniotis, "Discriminative pattern mining for breast cancer histopathology image classification via fully convolutional autoencoder," *IEEE Access*, vol. 7, pp. 36433–36445, 2019.
- [128] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Scientific Data*, vol. 4, no. 1, Article ID 170177, 2017.
- [129] C. Roa, "Data from: High-Throughput Adaptive Sampling for Whole-Slide Histopathology Image Analysis (HASHI) via Convolutional Neural Networks: Application to Invasive Breast Cancer Detection," 2018, <https://datadryad.org/stash/dataset/doi:10.5061/dryad.1g2nt41>.
- [130] P. Stanitsas, A. Cherian, V. Morellas, R. Tejpaul, N. Papanikolopoulos, and A. Truskinovsky, "Image descriptors for weakly annotated histopathological breast cancer data," *Frontiers in Digital Health*, vol. 2, Article ID 572671, 2020.
- [131] M. A. Asan and A. Ozsoy, "cuRCD: region covariance descriptor CUDA implementation," *Multimedia Tools and Applications*, vol. 80, no. 13, pp. 19737–19751, 2021.
- [132] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez, "Solving the multiple instance problem with axis-parallel rectangles," *Artificial Intelligence*, vol. 89, no. 1-2, pp. 31–71, 1997.
- [133] A. H. Fischer, K. A. Jacobson, J. Rose, and R. Zeller, "Hematoxylin and eosin staining of tissue and cell sections," *CSH Protocols*, vol. 2008, 2008.
- [134] S. K. Asare, F. You, and O. T. Nartey, "A semisupervised learning scheme with self-paced learning for classifying breast cancer histopathological images," *Computational Intelligence and Neuroscience*, vol. 2020, pp. 1–16, Article ID 8826568, 2020.
- [135] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1800–1807, Honolulu, HI, USA, May 2017.
- [136] K. Dimitropoulos, P. Barmpoutis, C. Zioga, A. Kamas, K. Patsiaoura, and N. Grammalidis, "Grading of invasive breast carcinoma through Grassmannian VLAD encoding," *PLoS One*, vol. 12, no. 9, Article ID e0185110, 2017.
- [137] A. Janowczyk and A. Madabhushi, "Grading of invasive breast carcinoma through Grassmannian VLAD encoding," *Journal of Pathology Informatics*, vol. 7, Article ID 27563488, 2016.
- [138] Y. Feng, L. Zhang, and J. Mo, "Deep manifold preserving autoencoder for classifying breast cancer histopathological images," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 17, no. 1, pp. 91–101, 2020.
- [139] E. J. Zaferani, M. Teshnehlab, and M. Vali, "Automatic personality traits perception using asymmetric auto-encoder," *IEEE Access*, vol. 9, pp. 68595–68608, 2021.
- [140] Y. LeCun, B. E. Boser, J. S. Denker et al., "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.
- [141] R. H. Carvalho, A. S. Martins, L. A. Neves, and M. Z. do Nascimento, "Analysis of features for breast cancer recognition in different magnifications of histopathological images," in *Proceedings of the International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 39–44, Brazil, August 2020.
- [142] M. H. Sharif and C. Djeraba, "An entropy approach for abnormal activities detection in video streams," *Pattern Recognition*, vol. 45, no. 7, pp. 2543–2561, 2012.
- [143] A. Toomaj and H. A. Atabay, "Some new findings on the cumulative residual Tsallis entropy," *Journal of Computational and Applied Mathematics*, vol. 400, Article ID 113669, 2022.
- [144] J. Li, J. Zhang, Q. Sun et al., "Breast cancer histopathological image classification based on deep second-order pooling network," in *Proceedings of the Int. Joint Conf. On Neural Networks (IJCNN)*, pp. 1–7, London, U K, November 2020.
- [145] P. Li, J. Xie, Q. Wang, and Z. Gao, "Towards Faster Training of Global Covariance Pooling Networks by Iterative Matrix Square Root Normalization," in *Proceedings of the Conference on computer vision and pattern recognition (CVPR)*, pp. 947–955, Salt Lake City, UT, USA, September 2018.
- [146] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep learning to improve breast cancer detection on screening mammography," *Scientific Reports*, vol. 9, no. 1, Article ID 12495, 2019.
- [147] G. Li, C. Li, G. Wu, D. Ji, and H. Zhang, "Multi-view attention-guided multiple instance detection network for interpretable breast cancer histopathological image diagnosis," *IEEE Access*, vol. 9, pp. 79671–79684, 2021.
- [148] R. Yan, F. Ren, Z. Wang et al., "Breast cancer histopathological image classification using a hybrid deep neural network," *Methods*, vol. 173, pp. 52–60, 2020.
- [149] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules Advances in Neural Information Processing Systems 30," in *Proceedings of the Annual Conference on Neural Information Processing Systems*, pp. 3856–3866, Long Beach, CA, USA, April 2017.
- [150] J. Kundale and S. Dhage, "Classification of breast cancer using histology images: handcrafted and pre-trained features based approach," *IOP Conference Series: Materials Science and Engineering*, vol. 1074, no. 1, Article ID 012008, 2021.
- [151] A. R. H. Khayyat, X. Sun, and P. L. Rosin, "Improved DSIFT descriptor based copy-rotate-move forgery detection. Image and video technology - 7th pacific-rim symposium (PSIVT), auckland," *New Zealand*, vol. 9431, pp. 642–655, 2015.
- [152] J. Wang, J. Yang, K. Yu, F. Lv, T. S. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3360–3367, San Francisco, USA, January 2010.
- [153] O. Attallah, F. Anwar, N. M. Ghanem, and M. A. Ismail, "Histo-CADx: duo cascaded fusion stages for breast cancer diagnosis from histopathological images," *PeerJ Computer Science*, vol. 7, p. e493, 2021.

- [154] K. C. Burçak, Ö. K. Baykan, and H. Uguz, “A new deep convolutional neural network model for classifying breast cancer histopathological images and the hyperparameter optimisation of the proposed model,” *The Journal of Supercomputing*, vol. 77, no. 1, pp. 973–989, 2021.
- [155] H. Iimori, G. T. F. de Abreu, O. Taghizadeh, R. A. Stoica, T. Hara, and K. Ishibashi, “A stochastic gradient descent approach for hybrid mmWave beamforming with blockage and CSI-error robustness,” *IEEE Access*, vol. 9, pp. 74471–74487, 2021.
- [156] A. Botev, G. Lever, and D. Barber, “Nesterov’s accelerated gradient and momentum as approximations to regularised update descent,” in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, pp. 1899–1903, Anchorage, AK, USA, August 2017.
- [157] A. Byerly and T. Kalganova, “Homogeneous vector capsules enable adaptive gradient descent in convolutional neural networks,” *IEEE Access*, vol. 9, pp. 48519–48530, 2021.
- [158] N. Shi, D. Li, M. Hong, and R. Sun, “RMSprop converges with proper hyper-parameter,” in *Proceedings of the International Conference on Learning Representations (ICLR)*, Virtual Event, Austria, February 2021.
- [159] Z. Qu, S. Yuan, R. Chi, L. Chang, and L. Zhao, “Genetic optimization method of pantograph and catenary comprehensive monitor status prediction model based on adadelta deep neural network,” *IEEE Access*, vol. 7, pp. 23210–23221, 2019.
- [160] I. K. M. Jais, A. R. Ismail, and S. Q. Nisa, “Adam optimization algorithm for wide and deep neural network,” *Knowledge Engineering and Data Science*, vol. 2, no. 1, pp. 41–46, 2019.
- [161] I. Hirra, M. Ahmad, A. Hussain et al., “Breast cancer classification from histopathological images using patch-based deep learning modeling,” *IEEE Access*, vol. 9, pp. 24273–24287, 2021.
- [162] G. E. Hinton, S. Osindero, and Y. W. Teh, “A fast learning algorithm for deep belief nets,” *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [163] Q. B. Baker and A. Abu Qutaish, “Evaluation of histopathological images segmentation techniques for breast cancer detection,” in *Proceedings of the International Conference on Information and Communication Systems (ICICS)*, pp. 134–139, Valencia, Spain, May 2021.
- [164] B. Ehteshami Bejnordi, M. Veta, P. Johannes van Diest et al., “Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer,” *JAMA*, vol. 318, no. 22, pp. 2199–2210, 2017.
- [165] M. Veta, Y. J. Heng, N. Stathonikos et al., “Predicting breast tumor proliferation from whole-slide images: the TUPAC16 challenge,” *Medical Image Analysis*, vol. 54, pp. 111–121, 2019.
- [166] M. Z. D. Nascimento, A. S. Martins, L. A. Neves, R. P. Ramos, E. L. Flôres, and G. A. Carrijo, “Classification of masses in mammographic image using wavelet domain features and polynomial classifier,” *Expert Systems with Applications*, vol. 40, no. 15, pp. 6213–6221, 2013.
- [167] M. Tan and Q. V. Le, “EfficientNet: rethinking model scaling for convolutional neural networks,” in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, vol. 97, pp. 6105–6114, Long Beach, California, USA, December 2019.
- [168] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. F. Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 248–255, Miami, Florida, USA, April 2009.
- [169] A. Ameh Joseph, M. Abdullahi, S. B. Junaidu, H. Hassan Ibrahim, and H. Chiroma, “Improved multi-classification of breast cancer histopathological images using handcrafted features and deep neural network (dense layer),” *Intelligent Systems with Applications*, vol. 14, Article ID 200066, 2022.
- [170] P. C. Chhipa, R. Upadhyay, G. G. Pihlgren, R. Saini, S. Uchida, and M. Liwicki, “Magnification prior: a self-supervised method for learning representations on breast cancer histopathological images,” 2022, <https://arxiv.org/abs/2203.07707>.
- [171] Y. Zou, J. Zhang, S. Huang, and B. Liu, “Breast cancer histopathological image classification using attention high-order deep network,” *International Journal of Imaging Systems and Technology*, vol. 32, no. 1, pp. 266–279, 2022.
- [172] M. Liu, Y. He, M. Wu, and C. Zeng, “Breast histopathological image classification method based on autoencoder and siamese framework,” *Information*, vol. 13, no. 3, p. 107, 2022.
- [173] D. A. Naik, R. M. Mohana, G. Ramu, Y. S. Lalitha, M. SureshKumar, and K. V. Raghavender, “Analyzing histopathological images by using machine learning techniques,” *Applied Nanoscience*, pp. 1–7, 2022.
- [174] S. Chattopadhyay, A. Dey, P. K. Singh, and R. Sarkar, “DRDA-Net: dense residual dual-shuffle attention network for breast cancer classification using histopathological images,” *Computers in Biology and Medicine*, vol. 145, Article ID 105437, 2022.
- [175] F. Shahidi, S. Mohd Daud, H. Abas, N. A. Ahmad, and N. Maarop, “Breast cancer classification using deep learning approaches and histopathology image: a comparison study,” *IEEE Access*, vol. 8, pp. 187531–187552, 2020.
- [176] C. Kampf, I. Olsson, U. Ryberg, E. Sjostedt, and F. Ponten, “Production of tissue microarrays, immunohistochemistry staining and digitalization within the human protein atlas,” *Journal of Visualized Experiments*, vol. 31, no. 63, p. 3620, 2012.
- [177] H. A. Shehu, M. H. Sharif, M. H. U. Sharif et al., “Deep sentiment analysis: a case study on stemmed Turkish twitter data,” *IEEE Access*, vol. 9, pp. 56836–56854, 2021.
- [178] M. H. Sharif, “An eigenvalue approach to detect flows and events in crowd videos,” *Journal of Circuits, Systems, and Computers*, vol. 26, no. 07, Article ID 1750110, 2017.
- [179] H. A. Shehu, M. H. Sharif, and R. A. Ramadan, “Distributed mutual exclusion algorithms for intersection traffic problems,” *IEEE Access*, vol. 8, pp. 138277–138296, 2020.
- [180] P. Jungkass and M. Berekovic, “Static allocation of basic blocks based on runtime and memory requirements in embedded real-time systems with hierarchical memory layout,” in *Proceedings of the Second Workshop on Next Generation Real-Time Embedded Systems*, vol. 87, no. 1–3, pp. 3–14, Budapest, Hungary, February 2021.
- [181] I. Loshchilov and F. Hutter, “SGDR: stochastic gradient descent with warm restarts,” in *Proceedings of the International Conference on Learning Representations (ICLR)*, pp. 1–16, Toulon, France, September 2017.
- [182] C. Zhang, P. Benz, D. M. Argaw et al., “ResNet or DenseNet? Introducing Dense Shortcuts to ResNet,” in *Proceedings of the Winter Conference on Applications of Computer Vision (WACV)*, pp. 3549–3558, HI, USA, December 2021.
- [183] I. L. S. V. R. C. ImageNet, “Large Scale Visual Recognition Challenge (ILSVRC),” 2010, <https://image-net.org/challenges/LSVRC>.

- [184] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size," *CoRR*, 2016, <https://arxiv.org/abs/1602.07360?context=cs>.
- [185] S. Kornblith, J. Shlens, and Q. V. Le, "Do Better ImageNet Models Transfer Better?" in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2661–2671, Canada CA, USA, March 2019.
- [186] M. N. Gurcan, L. Boucheron, A. Can et al., "Histopathological image analysis: a review," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 147–171, 2009.
- [187] M. Macenko, M. Niethammer, J. S. Marron et al., "A method for normalizing histology slides for quantitative analysis," in *Proceedings of the International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 1107–1110, Boston, MA, USA, June 2009.
- [188] Keras, *Keras API*, 2021.
- [189] H. A. Shehu, R. A. Ramadan, and M. H. Sharif, "Artificial intelligence tools and their capabilities," *PLOMS AI*, p. 1, 2021.
- [190] D. P. Kingma, J. Ba, and Adam, "A Method for Stochastic Optimization," in *Proceedings of the International Conference on Learning Representations (ICLR)*, Y. Bengio and Y. LeCun, Eds., San Diego, CA, USA, May 2015.
- [191] N. I. H. Malaria, "Datasets of National Institutes of Health (NIH)," 2021, <https://www.kaggle.com/iarunava/cell-images-for-detecting-malaria>.
- [192] Kaggle, "CoronaHack - chest X-ray-dataset," 2021, <https://github.com/ieee8023/covid-chestxray-dataset>.
- [193] Kaggle and Skin Cancer, "Malignant vs. Benign," 2021, <https://www.kaggle.com/fanconic/skin-cancer-malignant-vs-benign>.
- [194] H. A. Shehu, W. Browne, and H. Eisenbarth, "An Adversarial Attacks Resistance-Based Approach to Emotion Recognition from Images Using Facial Landmarks," in *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp. 1307–1314, Naples, Italy, August 2020.
- [195] M. H. Sharif and C. Djeraba, "A simple method for eccentric event espial using mahalnobis metric. Progress in pattern recognition, image analysis, computer vision, and applications," in *Proceedings of the 14th iberoamerican conference on pattern recognition (CIARP)*, vol. 5856, pp. 417–424, Guadalajara, Mexico, April 2009.
- [196] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *Journal of the American Statistical Association*, vol. 32, no. 200, pp. 675–701, 1937.
- [197] R. L. Iman and J. M. Davenport, "Approximations of the critical region of the fbietkan statistic," *Communications in Statistics - Theory and Methods*, vol. 9, no. 6, pp. 571–595, 1980.
- [198] J. L. Hodges and E. L. Lehmann, "Rank methods for combination of independent experiments in analysis of variance," *The Annals of Mathematical Statistics*, vol. 33, no. 2, pp. 482–497, 1962.
- [199] D. Quade, "Using weighted rankings in the analysis of complete blocks with additive block effects," *Journal of the American Statistical Association*, vol. 74, no. 367, pp. 680–683, 1979.
- [200] P. Westfall and S. Young, *Resampling-based Multiple Testing: Examples and Methods for P-Value Adjustment*, John Wiley & Sons, New Jersey, NY, USA, 2004.
- [201] O. J. Dunn, "Multiple comparisons among means," *Journal of the American Statistical Association*, vol. 56, no. 293, pp. 52–64, 1961.
- [202] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, vol. 6, pp. 65–70, 1979.
- [203] Y. Hochberg, "A sharper Bonferroni procedure for multiple tests of significance," *Biometrika*, vol. 75, no. 4, pp. 800–802, 1988.
- [204] G. Hommel, "A stagewise rejective multiple test procedure based on a modified Bonferroni test," *Biometrika*, vol. 75, no. 2, pp. 383–386, 1988.
- [205] G. Hommel and G. Bernhard, "A rapid algorithm and a computer program for multiple test procedures using logical structures of hypotheses," *Computer Methods and Programs in Biomedicine*, vol. 43, no. 3–4, pp. 213–216, 1994.
- [206] B. S. Holland and M. D. Copenhaver, "An improved sequentially rejective Bonferroni test procedure," *Biometrics*, vol. 43, no. 2, pp. 417–423, 1987.
- [207] D. M. Rom, "A sequentially rejective test procedure based on a modified Bonferroni inequality," *Biometrika*, vol. 77, no. 3, pp. 663–665, 1990.
- [208] H. Finner, "On a monotonicity problem in step-down multiple test procedures," *Journal of the American Statistical Association*, vol. 88, no. 423, pp. 920–923, 1993.
- [209] J. David Li, "A two-step rejection procedure for testing multiple hypotheses," *Journal of Statistical Planning and Inference*, vol. 138, no. 6, pp. 1521–1527, 2008.
- [210] P. Nemenyi, *Distribution-free Multiple Comparisons*, PhD thesis, Princeton University, New Jersey, NY, USA, 1963.
- [211] J. P. Shaffer, "Modified sequentially rejective multiple test procedures," *Journal of the American Statistical Association*, vol. 81, no. 395, pp. 826–831, 1986.
- [212] G. Bergmann and G. Hommel, "Improvements of general multiple test procedures for redundant systems of hypotheses," in *Multiple Hypotheses Testing*, P. Bauer and G. Hommel, Eds., pp. 100–115, Springer, New York, NY, USA, 1988.
- [213] S. García and F. Herrera, "An extension on "Statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons," *Journal of Machine Learning Research*, vol. 9, pp. 2677–2694, 2008.
- [214] G. University, "Soft Computing and Intelligent Information Systems," 2020, <https://sci2s.ugr.es/sicidm>.
- [215] B. Calvo and G. Santafé, "Scmamp: statistical comparison of multiple algorithms in multiple problems," *Rice Journal*, vol. 8, no. 1, pp. 248–256, 2016.
- [216] S. Cayir, G. Solmaz, H. Kusetogullari et al., "MITNET: a novel dataset and a two-stage deep learning approach for mitosis recognition in whole slide images of breast cancer tissue," *Neural Computing & Applications*, pp. 1–15, 2022.
- [217] Q. U. Ain, H. Al-Sahaf, B. Xue, and M. Zhang, "Generating knowledge-guided discriminative features using genetic programming for melanoma detection," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 5, no. 4, pp. 554–569, 2021.