Detection of Metastatic Breast Cancer from Whole-Slide Pathology Images Using an Ensemble Deep-Learning Method

Jafar Abdollahi[a,b] (iD), Niyousha Davari[c] (iD), Yasin Panahi[d] (iD), Mossa Gardaneh[b*] (iD)

[a]Department of Computer Engineering, Ardabil Branch, Islamic Azad University, Ardabil, Iran

[b]Genomedicx, Richmond Hill, Ontario, Canada

[c]Department of Life Science, Faculty of New Science and Technology, University of Tehran, Tehran, Iran

[d]Pharmacology & Toxicology department, School of pharmacy, Ardabil University of Medical Sciences, Ardabil, Iran

ARTICLE INFO

ABSTRACT

Background: Metastasis is the main cause of death toll among breast cancer patients. Since current approaches for diagnosis of lymph node metastases are time-consuming, deep learning (DL) algorithms with more speed and accuracy are explored for effective alternativest.

Methods: A total of 220025 whole-slide pictures from patients' lymph nodes were classified into two cohorts: testing and training. For metastatic cancer identification, we employed hybrid convolutional network models. The performance of our diagnostic system was verified using 57458 unlabeled images that utilized criteria that included accuracy, sensitivity, specificity, and P-value.

Results: The DL-based system that was automatically and exclusively capable of quantifying and identifying metastatic lymph nodes was engineered. Quantification was made with 98.84% accuracy. Moreover, the precision of VGG16 and Recall was 92.42% and 91.25%, respectively. Further experiments demonstrated that metastatic cancer differentiation levels could influence the recognition performance.

Conclusion: Our engineered diagnostic complex showed an elevated level of precision and efficiency for lymph node diagnosis. Our innovative DL-based system has a potential to simplify pathological screening for metastasis in breast cancer patients.

INTRODUCTION

Breast cancer (BC) is the most prevalent cancer among women and raises a fundamental challenge in public health globally. Breast cancer continues to be the most commonly diagnosed cancer and the second leading cause of cancer deaths among U.S. women.[1] The rate of BC incidence shows no declining prospect and in 2021, an estimated 281,550 new cases of invasive BC were expected to be diagnosed in women in the U.S. including 2650 new invasive cases, with an estimated death toll of 43600 women.[2]

*Address for correspondence:
Mossa Gardaneh, M.D.,
Genomedicx, Richmond Hill, Ontario, Canada
Tel: +1437-344-0309
Email: mossabenis65@gmail.com

Despite our profound understanding of biological mechanisms behind BC progression that led to development of various diagnostic and therapeutic approaches, the widespread incidences of BC and its subsequent heavy tolls necessitate early and accurate detection of BC. Besides, the emergence of personalized medicine drastically increases the load of work for pathologists and further complicates the histopathologic detection of cancer. Therefore, it is important that diagnostic protocols equally concentrate on the accuracy and efficiency of their performance.

Recognition of lymph node metastases (LNMets) is essential for pathological staging, prognosis, and adoption of appropriate treatment strategy in BC patients.[3] Preoperative prediction of LNMets provides

invaluable information that helps specify complementary therapy and expand surgical plans, thereby simplifying pre-treatment decisions. Currently, intraoperative sentinel lymph node biopsy is the first line diagnostic test for LNMets that includes surgery for biopsy preparation, histopathological processing and meticulous examination by an experienced pathologist. The pathologist should visually scan hard-to-detect large regions suspected of being cancerous in search of malignant areas.[4] This procedure is tedious, time-consuming, and inefficient in finding small malignant areas. Also, with many false-positive feedbacks, the diagnostic accuracy of this procedure is questionable.[5,6] Investigations are aimed at substituting this inefficient and poorly sensitive method with more efficient strategies. Invasion prediction of tumor-infiltrating lymphocytes,[7] the use of fluorescent probes,[8,9] one-step nucleic acid amplification,[10] and circulating microRNA detection)[11] are some of these emerging intraoperative strategies.

Parallel improvements are under way in digital and artificial intelligence-based approaches in biomedicine that are gaining momentum. In fact, the remarkable progress made in the quality of whole-slide images paves the ground for clinical utilization of digital images in anatomic pathology. These improvements alongside the advancements made in digital image analysis have made it possible for computer-aided diagnostics in pathology in order to enhance clinical care and histopathologic interpretation.[12] Recently, AI-based technologies have shown powerful performance in different automated image-recognition usages,[13] including image analysis of mammography,[14] magnetic resonance,[15,16] and ultrasound.[17,18,19] Lately, multiple DL-based algorithms including convolutional neural networks (CNNs) have been established for LNMets detection,[19,20,21,22] among other applications in the field of pathology.

The aim of our current study is to develop a DL-based algorithm in order to specify metastatic areas located in small-scale image patches that are obtained from bigger digitally-based pathology scans of lymph nodes.

METHODS

In this study, we applied a hybrid method of CNN-based classification for classifying our images. Two well-known deep and pre-trained CNN models (Resnet50, VGG16) and two fully-trained CNNs (Mobile-net, Google net) were employed for transfer learning and full training. In order to train the CNN, we used open-source DL-based Tensorflow and Keras libraries. The models' performance was analyzed in terms of accuracy, sensitivity, specificity, receiver operating characteristic curves, areas under the receiver operating characteristic curve, and heat maps.

Datasets

We utilized the Camelyon16 dataset consisting of 400 hematoxlyin and eosin (H&E) whole-slide images of lymph nodes, with metastatic regions labeled. The images are in Portable network graphics (PNG) format and can be downloaded here: "https://www.kaggle.com/c/histopathologic-cancer-detection/data." Furthermore, the data has two folders related to testing and training images as well as a training labels file. There are 220k training images, with a roughly 60/40 split between negatives and positives, and 57k evaluation images

In the dataset, scientists are endowed with plenty of small pathological images to categorize. An image ID is assigned to each file. The ground truth for the images located in the train folder is provided by a file named train_labels.csv. Indeed, scholars forecast the labels that are for the images within the test folder. To be more specific, a positive label demonstrates that the patch center region (32x32 px) includes minimum one pixel of tumor-associated tissue. Moreover, tumor tissue located in the patch's outer area does not affect the label.

Preprocessing

One of the essential elements for categorizing histological images is pre-processing. The dataset images are fairly large, while CNNs are normally designed in order to receive significantly smaller inputs. Hence, the images' resolution needs to be diminished to take in the input while keeping the key features. The dataset size is considerably lower than what is commonly needed to train a DL model appropriately; therefore, data augmentation is employed to raise the unique data amount in the set. In fact, this method greatly contributes to avoiding overfitting that is a phenomenon by which the model absorbs the training data properly, albeit completely unable to categorize and generalize unseen images.[23, 24]

Data Augmentation

The combinations of approaches supplied by the Keras library were examined to see their influence on overfitting and their contribution to enhancing categorization accuracy. Analyzing histological images is rotationally invariant, meaning that it does not take into account the angle at which a microscopy image is viewed. Consequently, employing rotation augmentation for the image should not affect the architecture training negatively.[23]

Ensemble Deep-Learning Approach for Detecting Metastatic Breast Cancer: Proposed Method

Innovation: Layer-wise fine-tuning and different weight initialization schemes.

In this study, we propose an autonomous classification method for BC classification. We used two pre-trained methods (VGG16, Resnet50) and two fully-trained ones (Google net and Mobile net) for our classification study. The models have been previously trained using the ImageNet database, which can be retrieved from the image-classification library of TensorFlow-Slim (http://tensorflow.org). Finally, we compared the outputs of the algorithms used in the pre-trained period and those of fully trained period in order to adequately evaluate the function of all models applied for the classification of breast histopathologic images into benign (B) and malignant (M) in the precise diagnosis of BC metastasis.
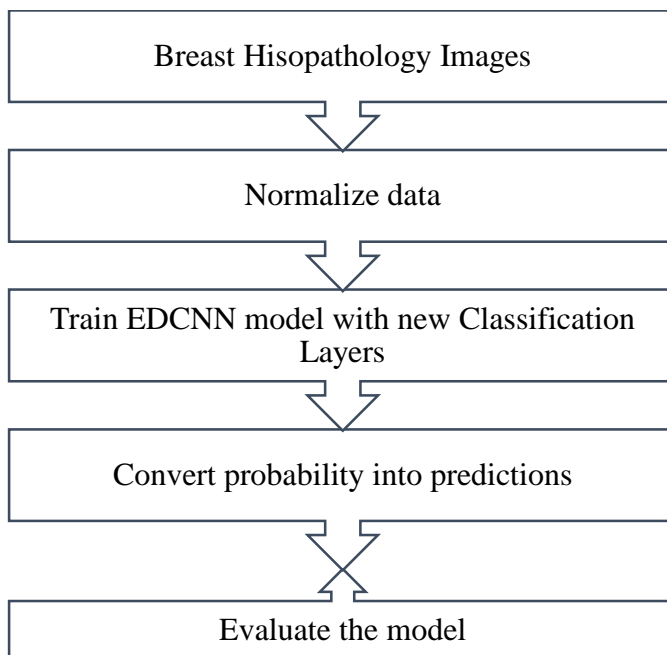
Pre-Training and Fine-tuning deep learning algorithms

In AI, pre-training is when a model is trained on one task to help it create parameters that may be applied to subsequent tasks. Humans are the ones who came up with the notion of pre-training. We do not have to learn anything from scratch because of an intrinsic skill. Instead, we transfer and reuse our previous knowledge to better interpret new information and perform a range of new activities.

Pre-training in AI imitates how humans process new information. That is, model parameters from previously learnt tasks are used to initialize model parameters for future tasks. In this approach, past knowledge aids new models in effectively performing new tasks based on previous experience rather than starting from blank. Simply explained, a pre-trained model is one that has already been taught to handle a comparable issue by someone else. Instead of beginning from scratch to tackle a comparable problem, you start with a model that has already been trained on another problem.

Fine-tuning deep learning techniques, on the other hand, will aid in improving the accuracy of a new neural network model by incorporating data from an existing neural network and utilizing it as an initialization point to speed up and simplify the training process. Although fine-tuning is useful for training new deep learning algorithms, it can only be employed when the dataset of an existing model and the dataset of the new deep learning model are similar.[45,46,47]

The basic steps of image classification using a deep CNN are demonstrated in Figure 1. In the model, the selected image is passed through a set of convolution layers with a fully connected layer, pooling layer, softmax layer, and layer filter.


```
┌─────────────────────────────────────┐
│   Breast Hisopathology Images        │
└─────────────────────────────────────┘
                  ↓
┌─────────────────────────────────────┐
│   Normalize data                     │
└─────────────────────────────────────┘
                  ↓
┌─────────────────────────────────────┐
│   Train EDCNN model with new         │
│   Classification Layers              │
└─────────────────────────────────────┘
                  ↓
┌─────────────────────────────────────┐
│   Convert probability into predictions│
└─────────────────────────────────────┘
                  ↓
┌─────────────────────────────────────┐
│   Evaluate the model                 │
└─────────────────────────────────────┘
```

Figure 1. Training process of the experiments

We trained our Ensemble deep convolutional neural network utilizing a training whole-slide pipeline in order to specify whether a selected image includes metastasis. The pipeline compromises model update and data preparation. In this classification system, low-level input image features are extracted by the first convolution layer, while semantic features are extracted by subsequent layers. The output is produced within the convolution layer and via employing dot product when a kernel slides over input. This process is accompanied by bias, thereby applying a nonlinear activation function such as Rectified linear activation is performed. The convolution layer output is transmitted to the pooling layer in order to decrease the image dimensionality and maintain the essential image information simultaneously. A series of pooling and

convolution layers extract high-level image features. Thereafter, fattening the feature map to a 1D vector and feeding it to a fully connected network are carried out. More specifically, a fully connected network includes several hidden layers possessing biases and weights. The neural networks described above employs an activation function which is nonlinear while permitting backpropagation. In contrast, backpropagation is not supported by the other activation function that is linear since the function derivative is not associated with inputs. Essentially, the neural network performance is not likely to enhance with an increase in hidden layers unless we employ a nonlinear function. Lastly, an activation function like softmax or sigmoid is performed to categorize the object based on zero to one probability.

Computer Hardware and Software

We used Google Colab, a Multi-graphics processing unit (GPU) tool, to run our experiments. We used a variety of modules in this kernel. Most of these modules provided much more functionality than we needed, but they were all handy. Software requirements are outlined in Table (1). Below is a detailed description of these modules:

- Glob – used for readily finding matching file names

- NumPy - the math module with applications in random number generation, Fourier transforms, linear algebra, etc.

- Pandas - a powerful module for data structures and analysis

- Keras - a high-level DL Application Programming Interface (API). In our case, it was employed to ease TensorFlow usage

- CV2 – used for image processing (we utilized it for loading images)

- TQDM - a minimalistic yet powerful and easy-to-use progress bar

- Matplotlib - a plotting module

Table 1. Software Requirements

Distribution	Anaconda Navigator and Google Colab
API	Keras
Library	Tensor Flow, OpenCV
Packages	Matplotlib, numpy, pandas, scikitLearn
Language	Python 3.7
IDE	Jupyter Notebook
GPU Architecture	Google Colab
Applications	LabelImg, TensorBoard

Performance Metrics

The performance of trained model has to be assessed on unseen data, a.k.a test dataset, using DL. The analysis of algorithms can be affected by performance metrics selection. In essence, this process assists in determining reasons for misclassifications; thus, they can be corrected via employing essential measures.[25,26,27,28]

Accuracy: demonstrates the "correct predictions made" number by the class, which is divided by the "total predictions made" number by the equivalent class.

Sensitivity: Real positive rate = The model will be positive if the result for a person in a few percentages of cases is positive. Sensitivity is calculated via the below formula.

Specificity: Real negative rate = The model will be negative if the result for a person in a few percentages of cases is negative. Specificity is calculated via the below formula.

Recall: the recall criterion states the ratio of "number of correctly categorized text data" within a specific class to the overall data number that must be categorized in the same class.

Precision: measures the ratio of the "correctly made predictions" division related to specific class samples to the "total predictions" number associated with the same class samples (this number includes the sum of all true and false predictions).[48,49,50]
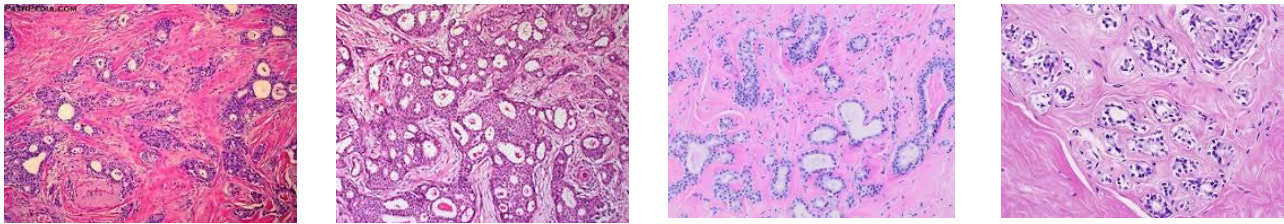
RESULTS

In the present research, coherent results related to the application of BC categorization in histopathological imaging modality were demonstrated. Figure 2 illustrates examples of images from various classes of the lymph node images we analyzed.
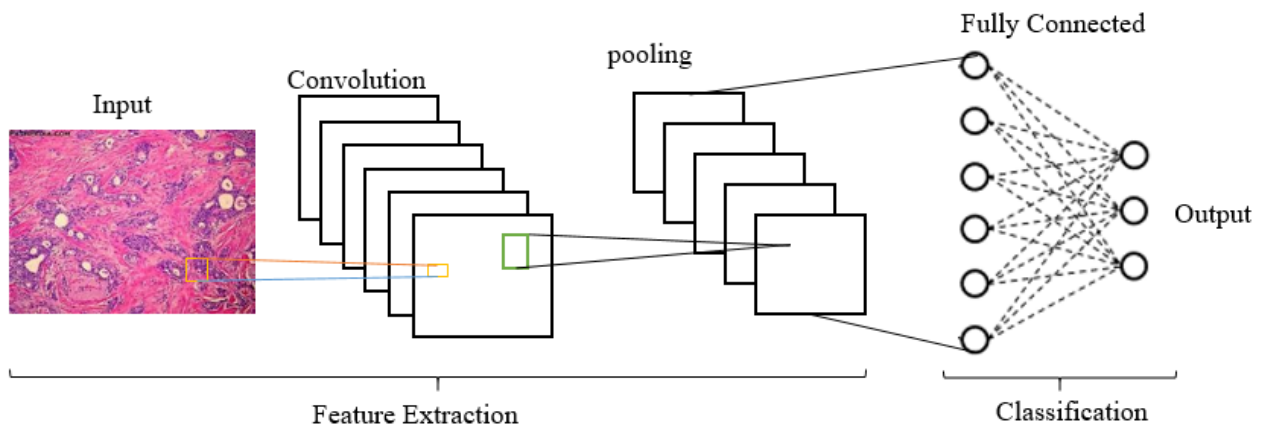
Herein, a balanced dataset was used for both fine-tuning and full training CNNs. We applied a training whole-slide pipeline in order to identify images that carry metastasis. Figure 3 represents basic classification steps using CNN.

Analysis of the pre-trained and fine-tune network

In order to measure the pre-trained and fine-tuned networks' classification performance, the whole dataset was divided into testing and training groups of data. Indeed, separating the dataset into the testing and training data is a usual practice in the field of neural networks employed for performance assessment. Moreover, to identify the size impact of testing-training data on network performance, we used four splitting fashions for testing-training data (60%, 40%).

Figure 2. Example images of lymph nodes used for analysis

Figure 3. Basic steps of classification using CNN

Using these splitting fashions, a series of experiments were conducted for four model networks. The elapsed time for each test was roughly ten to twenty hours. Thereafter, the experimental results were calculated with regard to f1 score, recall, and precision for each class, respectively. Then, to make an easy comparison, both classes' average results were calculated for each test. Additionally, classification performance was evaluated employing the Area Under the Curve (AUC) parameter and a receiver operating characteristics (ROC) analysis in order to validate the outcomes. Along with this procedure, computing the average precision score (APS) for performance measurement was done.

Results of Analyzing the full training and transfer learning

Table 2 shows the parameters used for the models. In this study, criteria 60% and 40% were used jointly to train and test the models with a learning rate of 0.0001 with epochs10. The outcomes achieved from the full training and transfer learning of Google net, Mobile net and VGG16, Resnet50 on the Breast cancer are presented.

Table 2. Hyper parameter for transfer learning and full-training

	Model	Train	Test	Batch size	Class mode	Learning rate	Loss	Epochs
1	VGG16	132015	88010	32	binary	0.0001	Binary crossentropy	10
2	Mobile-net	132015	88010	64	binary	0.0001	Binary crossentropy	10
3	ResNet50	132015	88010	32	binary	0.0001	Binary crossentropy	10
4	Google-net	132015	88010	64	binary	0.0001	Binary crossentropy	10

Analyzing the models' performance of the full training and transfer learning

Histopathological Images' dataset is summarized in Table 3. The models' performance was analyzed within the context of breast histopathological images' categorization, divided into malignant (M) and benign (B). Tab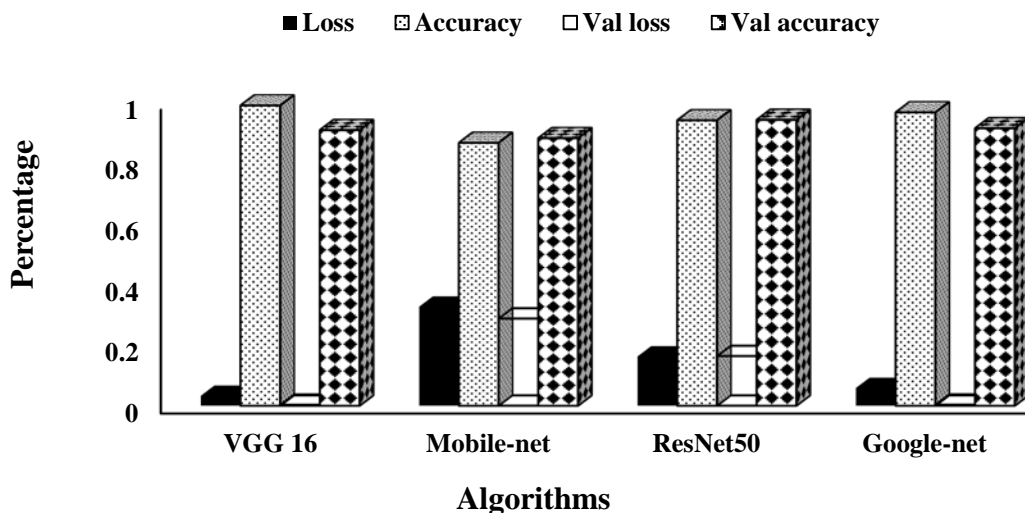le 3 indicates that the pre-trained and fine-tuned VGG16 network effectively performed better than the ResNet50 network, whilst the performance of google-net and that of ResNet50 could be compared with each other. Furthermore, in fully trained networks, mobile-net yielded poor results related to the dataset of 'Breast Histopathological Images.

Table 3. Results obtained from the transfer learning and full-training

	Model	Loss	Accuracy	Val loss	Val accuracy
1	VGG 16	0.0326	0.9884	0.0030	0.9075
2	Mobile-net	0.3256	0.8655	0.2864	0.8810
3	ResNet50	0.1621	0.9397	0.1633	0.9413
4	Google-net	0.0584	0.9658	0.0034	0.9135

Figure 4 shows Loss, Accuracy, Val loss, and Val accuracy of the four classification models when applied to the dataset. Many of the models that were built on the dataset had a high-level performance. However, the VGG16 method achieved the highest recall and highest accuracy of any model applied. Therefore, experimental results procured for the dataset show that the VGG16 method outperformed the other algorithms in terms of all metrics. We believe that our cancer prediction framework could assist oncologists to predict cancer incidence with high accuracy.

Figure 4. Results obtained from the transfer learning and full-training
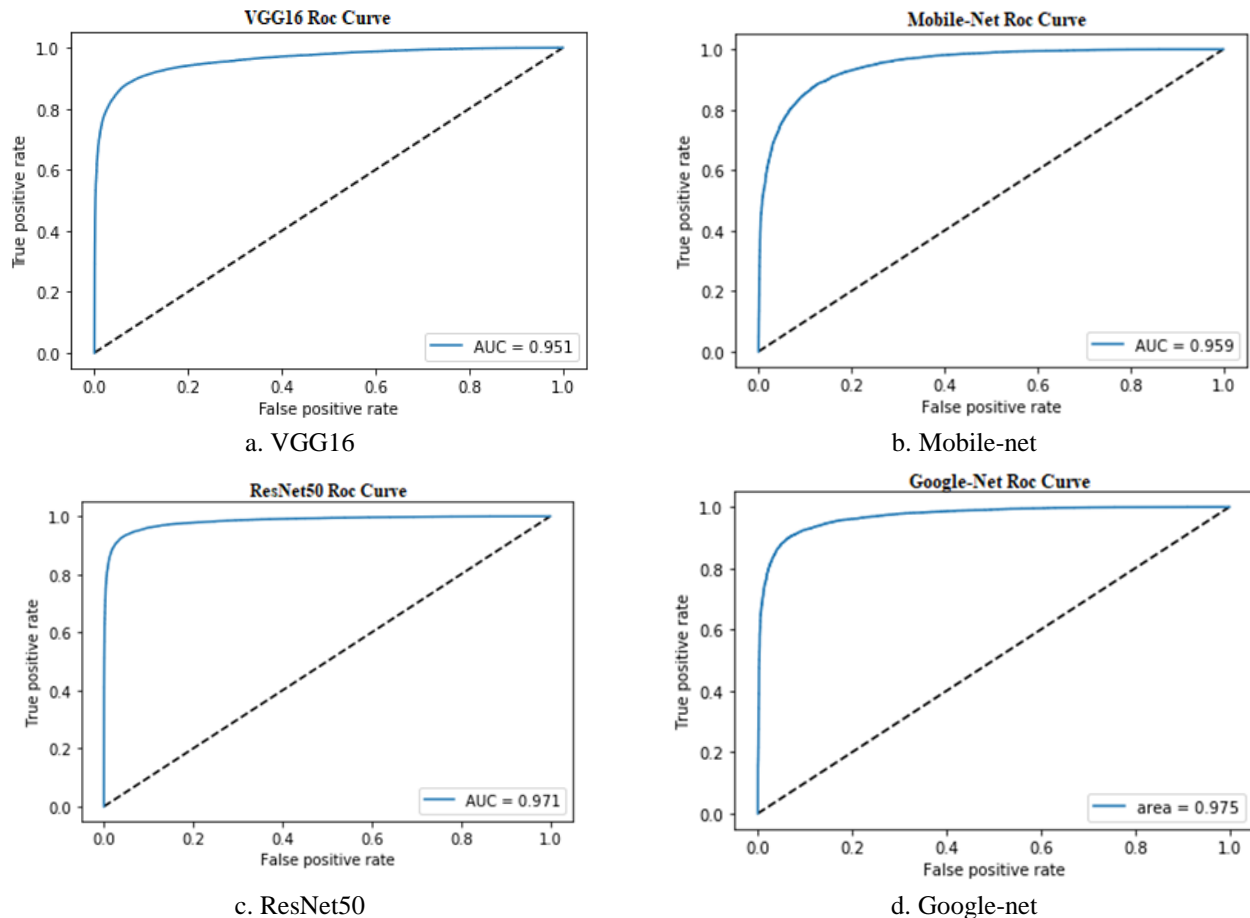
Moreover, we assessed the fully trained network performance, and google-net displayed an outstanding performance compared to ResNet50 networks. We realized that google-net and VGG16 networks were clearly biased to a specific class, based on the result obtained from Table 2. The solid evidence for the mentioned claim could be the recall, a.k.a. sensitivity, value. Within fully trained google-net and VGG16 networks, the recall amount was concurrently extremely high for one specific class and super low for other classes. Conversely, being equally sensitive to both of the classes, the ResNet50 network could better work in comparison with google-net and VGG16 networks.

Analysis of AUC and ROC in the full training and transfer learning models

Since the dataset size greatly affects the CNNs performance, four splitting fashions for testing-training data (60%–40) were employed so as to evaluate model performance regarding testing-training data size. With respect to this matter, AUC and ROC analyses were applied to compare all networks' performance, as demonstrated in Figure 5. In this figure, the AUC and ROC curves of the fully trained and pre-trained networks for the abovementioned four splits were compared. We found that pre-trained ResNet50 (AUC-98.04%) and VGG 16 (AUC 96.01%) perform better than the fully-trained mobile-net (AUC 97.01%) and google-net (AUC 95.00%).

In Figure 5, we only provide test set accuracies to prevent clutter. In each graph, we plotted the performance of transfer learning and that of full training against the number of iterations. First, we found that transfer learning outperforms full training. In the VGG16 model, the classification accuracy of the test set for transfer learning and that of full training is comparable. Second, we also found that fine-tuned models converge much earlier than their fully trained counterparts, showing that transfer learning needs less training time to achieve maximum performance. Although there is a great difference between natural/facial images and biomedical images, transfer learning gives better results than learning from scratch.

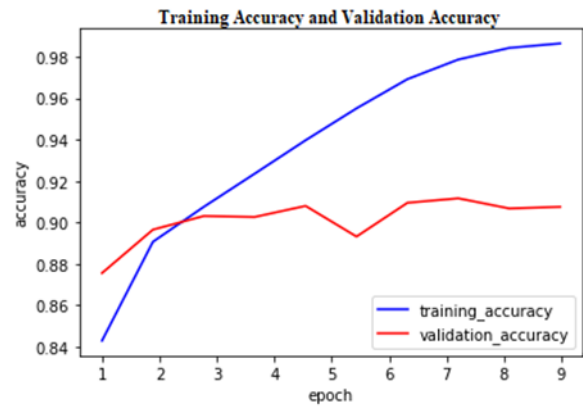a. VGG16

b. Mobile-net

c. ResNet50

d. Google-net

Figure 5. ROC analysis for breast cancer classification in (a) Fine-tuned pre-trained VGG16, (b) Fully trained mobile-net, (c) Fine-tuned pre-trained ResNet50 (d) Fully-trained google-net

All four networks performed well; however, the condition was somehow different for full training. In this study, VGG19 could perform well, but the performance of google-net and VGG16 networks was roughly analogous. The mobile-net deviation from the common trend was because of its higher sensitivity towards malignant and benign conditions during 60%–40% splitting separately. To sum up, it has been illustrated that the utilization of the transfer learning method leads to a remarkable performance regarding the CNNs with full training in the histopathological imaging modality, even in the case of a limited-size training dataset. In Table 3, we accurately compared our results with other research outcomes.

Figure 6 illustrates categorization accuracies related to the VGG16 test set (Figure 6a), mobile-net (Figure 6b), ResNet50 (Figure 6c), and google-net (Figure 6d). In each of these figures, we plotted the fine-tuning, full training, and transfer learning performance against the iterations' number. Since the sizes of the patch were smaller within VGG16, the needed iterations' number was higher in order to train them for ten epochs. Firstly, we could observe that

transfer learning performed better than full training, an outcome achieved from the comparisons. The aforesaid difference could be because of the feature space (source task) size and the network depth. Secondly, we viewed that the models being fine-tuned converged so much sooner than their fully trained counterparts, showing that transfer learning needs less time for training to obtain maximum performance. The categorization accuracies of fine-tuned VGG16, mobile-net, ResNet50 and google-net were 0.9884%, 0.8655%, 0.9397%, and 0.9658% after the first epoch. Deeper architectures like VGG16 converged more quickly within fine-tuned settings, indicating the ability to handle more intricate features. Besides, the utmost 0.9884% precision was obtained in the fine-tuned VGG16 model.
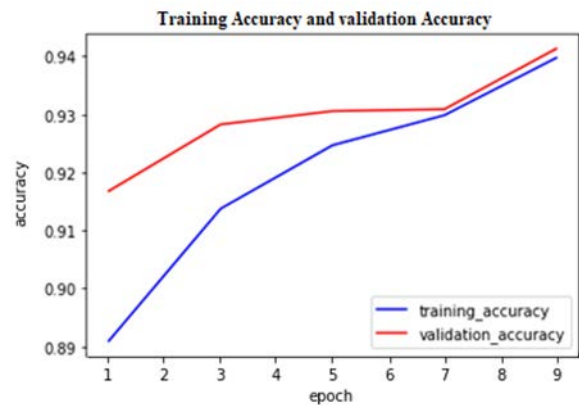
The experimental outcomes show that the learned features from deep CNNs that were trained on generic recognition works can be generalized to biomedical-related tasks and can be employed in order to fine-tune the new tasks possessing small datasets. Also, the comparison of accuracy, precision, and recall of models are outlined in Table 4.
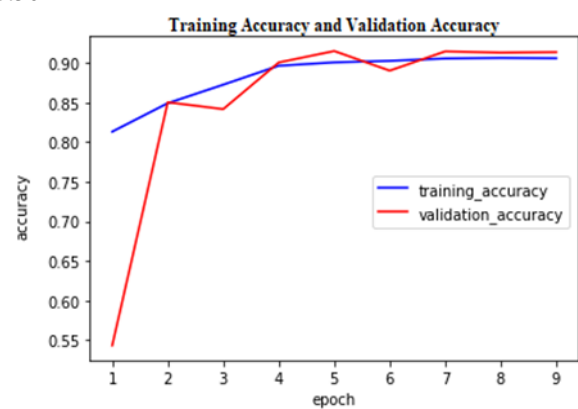
a. VGG16

b. Mobile-net

c. ResNet50

d. Google-net

Figure 6. Comparison of transfer learning with fine-tuning and full training for networks (a) VGG16, (b) mobile-net, (c) ResNet50, and (d) google-net
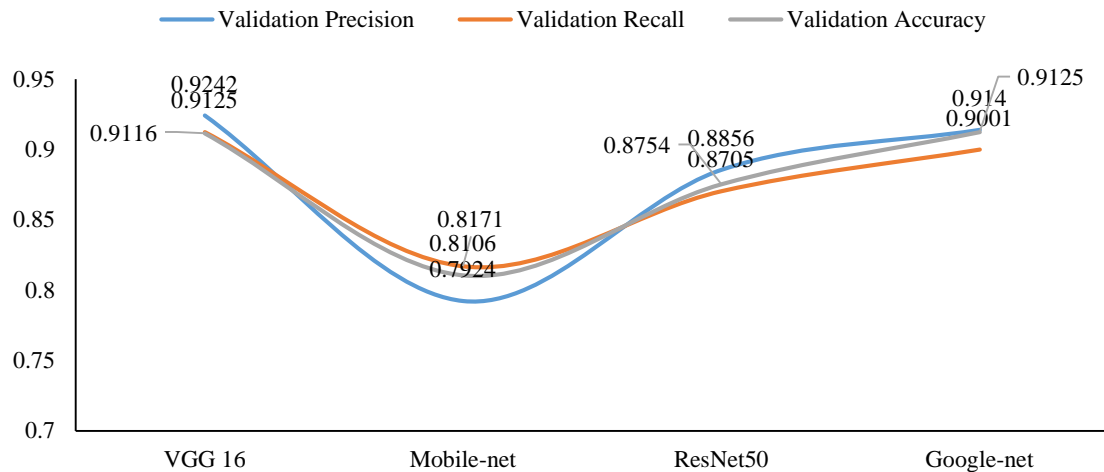
Table 4. Comparison of accuracy, precision and recall of models

	Model	Validation Precision	Validation Recall	Validation Accuracy
1	VGG 16	0.9242	0.9125	0.9116
2	Mobile-net	0.7924	0.8171	0.8106
3	ResNet50	0.8856	0.8705	0.8754
4	Google-net	0.9140	0.9001	0.9125

Figure 7 shows the Precision, Recall, and Accuracy of the four classification models when applied to the dataset. The VGG16, and google net method achieved the highest precision of any model and had the highest accuracy of any model applied. Thus, experimental results procured show that the VGG16 and google net method outperformed the other algorithms in terms of all metrics.

Figure 7. Comparison of accuracy, precision and recall of models implemented on validation data

Table 5. The comparison of data from various studies

	Date	Authors	Approach	Acc
1	2018	Sulaiman Vesal et al [30]	Inception-V3	97.08
2	2018	Shallu et al [31]	VGG16 with logistic regression	92.60%
3	2018	Hongliu Cao et al [32]	Five DL model	87.10%
4	2018	Carlos A. Ferreira [33]	Inception Resnet V2	0.76
5	2019	MuhammedTalo [34]	ResNet-50 pre-trained model	98.87%
6	2020	Laith Alzubaidi et al [35]	Utilized a transfer learning technique	97.4%
7	2020	YusufCelik [36]	DenseNet-161	91.57%
8	2021	Chanaleä Munien [37]	EfficientNet	98.33%
9	2020	Rishi Rai et al [38]	InceptionV3 and Xception	90.86
10	2020	Xiangchun Yu et al [39]	Pre-trained VGG19	85%, 75%, and 80.56%
11	2019	Md.Nuruddin Qaisar Bhuiyan et al [40]	Machine learning model	92% to 96%
12	2019	Ghulam Murtaza et al. [41]	Six ML classifiers	81.25%
13	2020	Abdulrahman Aloyayri et al [42]	ResNet18	98.73%
14	2022	You-WeiWang et al. [43]	Multi-energy level fusion model with a principal feature enhancement (PFE)	93%
15	2022	Annals of Nuclear Medicine et al. [44]	Xception	93% and 93%
16	2022	This Paper	VGG16	0.9884 %
17	2022	This Paper	Mobile-net	0.8655 %
18	2022	This Paper	ResNet50	0.9397 %
19	2022	This Paper	Google-net	0.9658 %

DISCUSSION

One of the major factors playing a role in women's mortality is BC. Nevertheless, an early diagnosis can result in cancer-associated death rates reduction. Traditional classification methods for BC are based upon handcrafted features approaches, and their performance depends on the selected features. They are also very sensitive to various sizes and intricate shapes. Histopathological images of BC possess complex shapes; hence, the classical classification techniques can be injected into DL algorithms after initial pre-processing. Employing a computer-assisted diagnosis system, scholars observe that efficiency can be enhanced, and costs will be diminished for the cancer-related diagnosis process. At present, an alternative answer to diagnosis procedure that can surmount the obstacles of classical categorization methods is DL models. Deep learning has become an active area of research, and its application in histopathology is quite new.

The principal cause of fatalities in BC patients is metastasis, referring to cancer cells spreading to different organs. Identifying cancer and determining its potential in metastasis at an early stage is essential [29]. To our knowledge, this paper is the first one delineating the overall applicability of DL approach to the diagnostic assessment of sentinel lymph nodes' whole-slide images. The potential suitability of this technique for enhancing the diagnostic process efficiency in histopathology was well-demonstrated in this research. This approach can result in adapted protocols in which pathologists conduct detailed analysis on various samples since the simple samples are already taken care of by a digital computer system. Table (5) compares the outcome of multiple relevant studies.

CONCLUSION

The comparison of four various CNN models possessing depths between three to thirteen convolutional layers was performed in this research. First of all, our empirical outcomes indicated that initializing parameters of a network with transferred features could enhance the categorization performance for any model. Nevertheless, deeper architectures that were trained on larger datasets converged rapidly. Furthermore, learning from scratch needs more training time than a pre-trained network model. Considering this matter, fine-tuned pre-trained VGG16 produced the best performance with 98.84% precision, 96.01% AUC, and 92.42 % APS for 90%–10% testing-training data splitting.

ETHICAL CONSIDERATIONS

Not applicable to this study.

ABBREVIATIONS

API, application programming interface; APS, average precision score; AUC, area under the curve; BC, breast cancer; CNNs, convolutional neural networks; DL, deep learning; LNMets, lymph node metastases; PNG, portable network graphics; ROC, receiver operating characteristics.

ACKNOWLEDGEMENTS

None.

FUNDING

This research received no external funding.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

REFERENCES

1. DeSantis CE, Fedewa SA, Goding Sauer A, Kramer JL, Smith RA, Jemal A. *Breast cancer statistics*, 2015: Convergence of incidence rates between black and white women. *CA: a cancer journal for clinicians*. 2016 Jan;66(1):31-42. doi: 10.3322/caac.21320.
2. U.S. Breast Cancer Statistics. Available from: https://www.breastcancer.org/symptoms/underst and_bc/statistics#:~:text=%20U.S.%20Breast%2 0Cancer%20Statistics%20%201%20About,rates %20have%20been%20steady%20in%20women. ..%20More%20. Updated in Feb 2021.
3. Nathanson SD, Rosso K, Chitale D, Burke M. Lymph Node Metastasis. In Introduction to Cancer Metastasis 2017 Jan 1 (pp. 235-261). *Academic Press*. doi: 10.1016/B978-0-12-804003-4.00013-X.
4. Veronesi U, Paganelli G, Viale G, Luini A, Zurrida S, Galimberti V, et al. A randomized comparison of sentinel-node biopsy with routine axillary dissection in breast cancer. *New England Journal of Medicine*. 2003 Aug 7;349(6):546-53. doi: 10.1056/NEJMoa012782.
5. Manca G, Rubello D, Tardelli E, Giammarile F, Mazzarri S, Boni G, et al. Sentinel lymph node biopsy in breast cancer: indications, contraindications, and controversies. *Clinical nuclear medicine*. 2016 Feb 1;41(2):126-33. doi: 10.1097/RLU.0000000000000985.
6. Okur O, Sagiroglu J, Kir G, Bulut N, Alimoglu O. Diagnostic accuracy of sentinel lymph node biopsy in determining the axillary lymph node metastasis. *Journal of Cancer Research and*

Therapeutics. 2020 Oct 1;16(6):1265. doi: 10.4103/jcrt.JCRT_1122_19.

7. Takada K, Kashiwagi S, Asano Y, Goto W, Kouhashi R, Yabumoto A, et al. Prediction of lymph node metastasis by tumor-infiltrating lymphocytes in T1 breast cancer. *BMC cancer*. 2020 Dec;20(1):1-3. doi: 10.1186/s12885-020-07101-y.

8. Shinden Y, Ueo H, Tobo T, Gamachi A, Utou M, Komatsu H, et al. Rapid diagnosis of lymph node metastasis in breast cancer using a new fluorescent method with γ-glutamyl hydroxymethyl rhodamine green. *Scientific reports*. 2016 Jun 9;6(1):1-7. doi: 10.1038/srep27525.

9. Fujita K, Kamiya M, Yoshioka T, Ogasawara A, Hino R, Kojima R, et al. Rapid and accurate visualization of breast tumors with a fluorescent probe targeting α-mannosidase 2C1. *ACS central science*. 2020 Oct 29;6(12):2217-27. doi: 10.1021/acscentsci.0c01189.

10. Combi F, Andreotti A, Gambini A, Palma E, Papi S, Biroli A, et al. Application of OSNA Nomogram in Patients With Macrometastatic Sentinel Lymph Node: A Retrospective Assessment of Accuracy. *Breast Cancer: Basic and Clinical Research.* 2021 May;15:11782234211014796. doi: 10.1177/11782234211014796.

11. Escuin D, López-Vilaró L, Mora J, Bell O, Moral A, Pérez I, et al. Circulating microRNAs in Early Breast Cancer Patients and Its Association With Lymph Node Metastases. Frontiers in Oncology. 2021;11. doi: 10.3389/fonc.2021.627811.

12. Steiner DF, MacDonald R, Liu Y, Truszkowski P, Hipp JD, Gammage C, et al. Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer. *The American journal of surgical pathology*. 2018 Dec;42(12):1636. doi: 10.1097/PAS.0000000000001151.

13. Lei YM, Yin M, Yu MH, Yu J, Zeng SE, Lv WZ, et al. Artificial intelligence in medical imaging of the breast. *Frontiers in Oncology*. 2021:2892. doi: 10.3389/fonc.2021.600557.

14. Geras KJ, Mann RM, Moy L. Artificial intelligence for mammography and digital breast tomosynthesis: current concepts and future perspectives. *Radiology*. 2019 Nov;293(2):246-59. doi: 10.1148/radiol.2019182627.

15. Ha R, Chang P, Karcich J, Mutasa S, Fardanesh R, Wynn RT, et al. Axillary lymph node evaluation utilizing convolutional neural networks using MRI dataset. *Journal of Digital Imaging*. 2018 Dec;31(6):851-6. doi: 10.1007/s10278-018-0086-7.

16. Li J, Zhou Y, Wang P, Zhao H, Wang X, Tang N, et al. Deep transfer learning based on magnetic resonance imaging can improve the diagnosis of lymph node metastasis in patients with rectal cancer. *Quantitative Imaging in Medicine and Surgery*. 2021 Jun;11(6):2477. doi: 10.21037/qims-20-525.

17. Sultan LR, Schultz SM, Cary TW, Sehgal CM. Machine learning to improve breast cancer diagnosis by multimodal ultrasound. In2018 *IEEE International Ultrasonics Symposium* (IUS) 2018 Oct 22 (pp. 1-4). IEEE. doi: 10.1109/ULTSYM.2018.8579953.

18. Wu T, Sultan LR, Tian J, Cary TW, Sehgal CM. Machine learning for diagnostic ultrasound of triple-negative breast cancer. *Breast cancer research and treatment*. 2019 Jan;173(2):365-73. doi: 10.1007/s10549-018-4984-7.

19. Lee YW, Huang CS, Shih CC, Chang RF. Axillary lymph node metastasis status prediction of early-stage breast cancer using convolutional neural networks. *Computers in Biology and Medicine*. 2021 Mar 1;130:104206. doi: 10.1016/j.compbiomed.2020.104206

20. Zhou LQ, Wu XL, Huang SY, Wu GG, Ye HR, Wei Q, et al. Lymph node metastasis prediction from primary breast cancer US images using deep learning. *Radiology*. 2020 Jan;294(1):19-28. doi: 10.1148/radiol.2019190372.

21. Zheng X, Yao Z, Huang Y, Yu Y, Wang Y, Liu Y, et al. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. *Nature communications*. 2020 Mar 6;11(1):1-9. doi: 10.1038/s41467-020-15027-z.

22. Wang J, Liu Q, Xie H, Yang Z, Zhou H. Boosted efficientnet: Detection of lymph node metastases in breast cancer using convolutional neural networks. *Cancers*. 2021 Jan;13(4):661. doi: 10.3390/cancers13040661.

23. Munien C, Viriri S. Classification of hematoxylin and eosin-stained breast cancer histology microscopy images using transfer learning with EfficientNets. *Computational Intelligence and Neuroscience*. 2021 Oct;2021. doi: 10.1155/2021/5580914.

24. Yu X, Chen H, Liang M, Xu Q, He L. A transfer learning-based novel fusion convolutional neural network for breast cancer histology classification. *Multimedia Tools and Applications*. 2020 Oct 9:1-5. doi: 10.1007/s11042-020-09977-1.

25. Abdollahi J, Keshandehghan A, Gardaneh M, Panahi Y, Gardaneh M. Accurate detection of breast cancer metastasis using a hybrid model of artificial intelligence algorithm. *Archives of Breast Cancer*. 2020 Feb 29:22-8. doi: 0.32768/abc.20207122-28.

26. Abdollahi J, Moghaddam BN, Parvar ME. Improving diabetes diagnosis in smart health using genetic-based Ensemble learning algorithm. Approach to IoT Infrastructure. *Future Gen Distrib Systems J*. 2019;1:23-30. doi: 10.1007/s42044-022-00100-1.

27. Abdollahi J. A review of Deep learning methods in the study, prediction and management of COVID-19. *Journal of Industrial Integration and Management* 2020. Vol. 05, No. 04, pp.453-479. doi: 10.1142/S2424862220500268.

28. Abdollahi J, Nouri-Moghaddam B. Hybrid stacked ensemble combined with genetic algorithms for diabetes prediction. *Iran Journal of Computer Science*. 2022 Mar 21:1-6. doi: 10.1007/s42044-022-00100-1.

29. Xue J, Pu Y, Smith J, Gao X, Wang C, Wu B. Identifying metastatic ability of prostate cancer cell lines using native fluorescence spectroscopy and machine learning methods. *Scientific Reports*. 2021 Jan 26;11(1):1-0. doi: 10.1038/s41598-021-81945-7.

30. Irshad H, Veillard A, Roux L, Racoceanu D. Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential. *IEEE reviews in biomedical engineering*. 2013 Dec 20;7:97-114. doi: 10.1109/RBME.2013.2295804.

31. McCann MT, Ozolek JA, Castro CA, Parvin B, Kovacevic J. Automated histology analysis: Opportunities for signal processing. *IEEE Signal Processing Magazine*. 2014 Dec 4;32(1):78-87. doi: 10.1109/MSP.2014.2346443.

32. Veta M, Pluim JP, Van Diest PJ, Viergever MA. Breast cancer histopathology image analysis: A review. *IEEE transactions on biomedical engineering*. 2014 Jan 30;61(5):1400-11. doi: 10.1109/TBME.2014.2303852.

33. Chen CL, Chen CC, Yu WH, Chen SH, Chang YC, Hsu TI, et al. An annotation-free whole-slide training approach to pathological classification of lung cancer types using deep learning. Nature communications. 2021 Feb 19;12(1):1-3. doi: 10.1038/s41467-021-21467-y.

34. Zhou LQ, Wu XL, Huang SY, Wu GG, Ye HR, Wei Q, et al. Lymph node metastasis prediction from primary breast cancer US images using deep learning. *Radiology*. 2020 Jan;294(1):19-28. doi: 10.1148/radiol.2019190372.

35. Cireşan DC, Giusti A, Gambardella LM, Schmidhuber J. Mitosis detection in breast cancer histology images with deep neural networks. International conference on medical image computing and computer-assisted intervention *Med Image Comput Comput Assist Interv*.

2013;16(Pt 2):411-8. doi: 10.1007/978-3-642-40763-5_51.

36. Albawi S, Mohammed TA, Al-Zawi S. Understanding of a convolutional neural network. *international conference on engineering and technology (ICET)* 2017 Aug 21 (pp. 1-6). *IEEE*. doi: 10.1109/ICEngTechnol.2017.8308186.

37. Han Z, Wei B, Zheng Y, Yin Y, Li K, Li S. Breast cancer multi-classification from histopathological images with structured deep learning model. *Scientific reports*. 2017 Jun 23;7(1):1-0. doi: 10.1038/s41598-017-04075-z.

38. Salakhutdinov R, Hinton G. (2009). Deep boltzmann machines. *Proceedings of AISTATS* 2009 (pp. 448–455). PMLR.

39. Hinton G, Salakhutdinov R. An efficient learning procedure for deep Boltzmann machines. *Neural Computation*. 2012;24(8):1967-2006. doi: 10.1162/NECO_a_00311.

40. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science*. 2006 Jul 28;313(5786):504-7. doi: 10.1126/science.1127647.

41. Shaffie A, Soliman A, Ghazal M, Taher F, Dunlap N, Wang B, et al. A new framework for incorporating appearance and shape features of lung nodules for precise diagnosis of lung cancer. *IEEE International Conference on Image Processing (ICIP)* 2017 Sep 17 (pp. 1372-1376). IEEE. doi: 10.1109/ICIP.2017.8296506.

42. Kingma DP, Welling M. Auto-encoding variational bayes. *arXiv preprint arXiv*:1312.6114. 2013 Dec 20. doi: 10.48550/arXiv.1312.6114.

43. Wang YW, Chen CJ, Wang TC, Huang HC, Chen HM, Shih JY, et al. Multi-energy level fusion for nodal metastasis classification of primary lung tumor on dual energy CT using deep learning. *Computers in biology and medicine*. 2022 Feb 1;141:105185. doi: 10.1016/j.compbiomed.2021.105185.

44. Satoh Y, Imokawa T, Fujioka T, Mori M, Yamaga E, Takahashi K, et al. Deep learning for image classification in dedicated breast positron emission tomography (dbPET). *Annals of Nuclear Medicine*. 2022 Jan 27:1-0. doi: 10.1007/s12149-022-01719-7.

45. Erhan D, Courville A, Bengio Y, Vincent P. Why does unsupervised pre-training help deep learning?. InProceedings of the thirteenth international conference on artificial intelligence and statistics 2010 Mar 31 (pp. 201-208). *JMLR Workshop and Conference Proceedings*. doi: Not Available

46. Yu D, Deng L, Dahl G. Roles of pre-training and fine-tuning in context-dependent DBN-HMMs

for real-world speech recognition. *InProc. NIPS Workshop on Deep Learning and Unsupervised Feature Learning* 2010 Dec. sn. doi: Not Available

47. Lee C, Panda P, Srinivasan G, Roy K. Training deep spiking convolutional neural networks with stdp-based unsupervised pre-training followed by supervised fine-tuning. *Frontiers in neuroscience*. 2018 Aug 3;12:435. doi: 10.3389/fnins.2018.00435.

48. Abdollahi J, Amani F, Mohammadnia A, Amani P, Fattahzadeh-ardalani G. Using Stacking methods based Genetic Algorithm to predict the time between symptom onset and hospital arrival in stroke patients and its related factors. *JBE*. 2022;8(1):8-23.

49. Abdollahi J, Nouri-Moghaddam B. Feature selection for medical diagnosis: Evaluation for using a hybrid Stacked-Genetic approach in the diagnosis of heart disease. *arXiv preprint arXiv*:2103.08175. 2021 Mar 15. doi: 10.48550/arXiv.2103.08175.

50. Abdollahi J, Nouri-Moghaddam B, Ghazanfari M. Deep Neural Network Based Ensemble learning Algorithms for the healthcare system (diagnosis of chronic diseases). *arXiv preprint arXiv*:2103.08182. 2021 Mar 15. doi: 10.48550/arXiv.2103.08182.

How to Cite This Article