2023 18th International Workshop on Semantic and Social Media Adaptation & Personalization (SMAP) | 979-8-3503-2771-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/SMAP59435.2023.10255168

Deep Learning Analysis of Histopathology Images for Breast Cancer Detection: A Comparative Study of ResNet and VGG Architectures

Andreas Kanavos* and Phivos Mylonas†
*Department of Informatics, Ionian University, Corfu, Greece
akanavos@ionio.gr
†Department of Informatics and Computer Engineering, University of West Attica, Athens, Greece
mylonasf@uniwa.gr

***Abstract*—Medical image analysis has undergone significant advancements with the emergence of deep learning techniques, offering great promise in improving diagnostic precision and expediting patient care. This research investigates the effectiveness of ResNet and VGG architectures in detecting breast cancer through the analysis of histopathology images. By meticulously fine-tuning hyperparameters and optimizers, we establish robust and accurate deep learning models. Our findings reveal that the ResNet model with the SGD optimizer excels, surpassing the performance of VGG in terms of accuracy and F1-score. However, employing transfer learning with pre-trained VGG16 and ResNet50 networks does not yield competitive results, potentially due to disparities in input image size and data distribution. The primary focus of this study is to address the critical challenge of early breast cancer detection, ultimately leading to enhanced patient outcomes. By exploring state-of-the-art deep learning architectures and methodologies, we contribute to the growing body of research aimed at leveraging artificial intelligence for medical diagnosis.**

***Index Terms*—Medical Image Analysis, Deep Learning, Breast Cancer Detection, Histopathology Images, ResNet, VGG, Transfer Learning**

I. Introduction

Medical image analysis plays a pivotal role in addressing clinical challenges by leveraging advanced technologies to interpret and extract vital information from medical images [17]. In recent years, the integration of deep learning techniques has revolutionized this field, empowering neural networks to discern intricate features within medical images that were once reliant on specialized medical knowledge [1]. Neural networks have exhibited remarkable success across diverse aspects of medical image analysis, including categorization, detection, and segmentation, significantly augmenting diagnostic accuracy and expediting treatment decisions.

The success of deep learning in medical image analysis can be attributed, in part, to transfer learning and data augmentation techniques, which overcome challenges such as data scarcity and incomplete categorization [15]. Transfer learning enables the utilization of pre-trained deep neural networks on other data, which significantly accelerates model training and enhances performance. Meanwhile, data augmentation augments the training dataset by generating variations of existing images, diversifying the network's exposure to different image patterns and enhancing its ability to generalize to new cases [18].

Breast cancer stands as a pressing health concern, representing the leading cause of cancer-related deaths in women aged 20-59 and the second most common cause for women over 59 [2]. Early diagnosis and timely intervention are crucial in improving treatment outcomes, prognosis, and overall survival rates. The conventional diagnostic process typically involves the identification of suspicious findings through palpation, mammography, or ultrasound, followed by tissue sampling and biopsy. The subsequent analysis of the tissue sample by histopathologists plays a critical role in determining the presence and characteristics of malignant cells.

Manual examination of breast tissue and cellular structures necessitates significant expertise and imposes a considerable workload on histopathologists. As a consequence, the agreement among pathologists' assessments can vary, with the average agreement on results hovering around 75% [7]. Furthermore, specific types of breast cancer pose additional challenges, with a study reporting that in one out of three patients (33%), malignancies were not diagnosed through conventional medical imaging methods [6].

To address the complexities and variability in breast cancer diagnosis, the integration of deep learning in medical image analysis offers a promising avenue. By leveraging the power of artificial intelligence and large datasets, it holds the potential to enhance diagnostic accuracy, reduce human error, and improve overall patient outcomes.

The remainder of this paper is organized as follows: Section II presents an overview of related work in the field, highlighting the novelty and contributions of our research. Section III provides insights into the fundamental concepts, and methods employed in our investigation. Section IV presents the research findings, accompanied by details about the dataset utilized in our experiments. Finally, Section V concludes the paper, summarizing key insights, and outlines potential avenues for future research.

II. Related Work

Computer-aided diagnosis (CAD) systems encompass a fusion of Image Analysis and Machine Learning methodologies designed to aid medical practitioners in the diagnostic process. By functioning as a supplementary opinion provider, CAD systems alleviate the burdens on specialists, thereby enhancing diagnostic efficiency and cost-effectiveness. This often entails an emulation of physicians' procedures. For instance, the evaluation of nuclei morphology in isolation has the potential to serve as the basis for classifying tissue as benign or malignant [20].

As a result, certain research endeavors are centered around nuclei analysis to discern malignancy from benignity. In [12], a variety of clustering algorithms were employed to segment nuclei in microscopic images from fine needle biopsies. They harnessed morphological, topological, and texture-based attributes to train a classifier, achieving accuracies ranging between 84% and 93% across a dataset of 500 images sourced from 50 patients. Similarly, in [8], [9], nuclei-centric features were extracted from fine needle biopsy samples. The initial phase encompassed the utilization of the circular Hough transform to detect potential nuclei candidates, followed by the implementation of machine learning techniques and Otsu thresholding to mitigate false positives. Both studies leveraged shape and texture attributes of nuclei for training a diverse array of classifiers. Apart from nucleus-related data, authors in [3] incorporated tissue organization as a factor for binary classification in more complex images. Their investigation encompassed the assessment of 70 images from a proprietary 40× magnification breast histology H&E dataset.

Another facet of research has been dedicated to addressing the intricate challenge of categorizing breast cancer histology images into three distinct classes. Illustratively, authors in [4], [21] undertook the classification of images into normal, in situ carcinoma, and invasive carcinoma, utilizing a dataset sourced from the Israel Institute of Technology. A strategy involving the application of multiple threshold values for image binarization, followed by the use of connected component statistics to train a support vector machine (SVM) classifier, was adopted in [4]. On a different tangent, a cascade classification methodology was implemented in [21]. Subsets of Curvelet Transform and local binary pattern (LBP) features were randomly channeled into an initial ensemble of parallel SVM classifiers.

In light of recent advancements in computational capabilities and the expansion of dataset sizes, the application of CNNs to image classification tasks has gained prominence. In contrast to conventional approaches that involve manual curation of features, CNNs acquire pertinent attributes directly from training image patches through the optimization of a classification loss function. These deep learning models have demonstrated exceptional performance across diverse domains in image classification challenges [5], [13] including the realm of medical image analysis [14], with particular efficacy observed in the domain of histopathology images [19].

III. Methodology

In this section, we present a comprehensive description of the methodology employed in this study, encompassing the libraries utilized and the architectural details of the networks.

A. Libraries

The following libraries were utilized to implement and experiment with our machine learning models:

- Augmentor: A library for data augmentation, which can be beneficial in increasing the diversity of the training dataset.
- Imutils: This library provides a set of basic image manipulation functions, such as rotation, resizing, edge control, graph display, etc.
- Keras: A high-level API that works in conjunction with Tensorflow, enabling the seamless creation of convolutional or recursive networks, or a combination of both.
- Matplotlib: A popular library used for creating two-dimensional plots in various formats, facilitating data visualization and model performance analysis.
- Numpy: Python's package, which includes manipulation of n-dimensional vectors, and random number manipulation capabilities, essential for efficient data handling.
- OpenCV: This library contains computer vision algorithms, making it valuable for image-related tasks.
- Pandas: Pandas provides tools for quickly creating flexible data structures, simplifying data handling and manipulation.
- Pillow: Used for image manipulation and processing, Pillow allows the handling of various image formats and enables pre-processing steps for the image-based datasets.
- PyTorch: A popular deep learning library that provides a flexible and efficient platform for building and training neural networks.
- Scikit-learn: A comprehensive library for data mining and analysis, supporting classification, regression, clustering, feature extraction, and both supervised and unsupervised learning techniques.
- Scipy: A package for manipulating numerical data, offering a wide array of mathematical functions and statistical operations.
- Tensorflow: Developed by Google, Tensorflow enables easy creation of machine learning models and deep neural networks.

B. ResNet

The ResNet (Residual Network) architecture is a groundbreaking innovation in deep learning that addresses the challenge of vanishing gradients in extremely deep neural networks. Vanishing gradients occur when the gradients calculated during the backpropagation process become infinitesimally small as they propagate through numerous layers, hindering the effective training of very deep networks. ResNet introduces a clever solution in the form of residual blocks, which enable the training of much deeper networks with improved convergence.

A fundamental concept of the ResNet architecture is the notion of "skip connections" or "shortcut connections". In a conventional deep network, the output of one layer serves as the input for the next layer. However, in ResNet, instead of propagating the input through a single layer, it is passed along a shortcut connection to a deeper layer. This approach enables the network to learn residual mappings, i.e., the difference between the input and the desired output. The residual mappings can be easier for the network to learn than the actual mappings, thus facilitating more efficient training.

Residual Level:

$$Output = Conv1x1(Input) \rightarrow ReLU \rightarrow \\ \rightarrow Conv3x3(Output) \rightarrow ReLU \rightarrow Conv1x1(Output) \quad (1)$$

$$Output = Output + Input \quad (2)$$

The residual level begins with a 1x1 convolution layer, followed by a Rectified Linear Unit (ReLU) activation function. It is then passed through a 3x3 convolutional layer with another ReLU activation. Finally, another 1x1 convolution layer is applied to obtain the final output. The summation operation (Output = Output + Input) combines the output of the residual level with the original input, forming the skip connection. The skip connection is crucial in preserving the gradient flow during backpropagation and mitigating vanishing gradients.

The ResNet architecture is organized into blocks, with each block containing one initial residual level that employs a stride of 2 to downsample the spatial dimensions of the input. This downsampled output then goes through additional residual levels with a stride of 1, which maintains the spatial dimensions. The number of filters used in each residual level can vary based on the specific block's configuration.

The ResNet architecture has been demonstrated to achieve remarkable results in various computer vision tasks, including image classification, object detection, and image segmentation. In this study, ResNet was applied to the classification of IDC positive and IDC negative patches from breast histopathology images. By leveraging the power of residual connections, the network aims to learn intricate patterns and representations from the data, enabling accurate classification and detection of breast cancer.

C. VGG (Visual Geometry Group)

The VGG network, developed by researchers from the Visual Geometry Group at the University of Oxford, is a well-known and influential architecture in the field of computer vision. VGG gained significant popularity due to its simple yet effective design and remarkable performance in image recognition tasks, especially image classification.

The main characteristic of the VGG architecture is its deep structure, achieved by stacking multiple convolutional layers. Unlike previous models that employed larger convolutional filters, VGG uses smaller 3x3 filters throughout the network, which allows for more non-linear transformations and results in a more expressive model. The use of 3x3 filters with a stride of 1 also ensures that the receptive field of the network remains relatively small, making it possible to stack more convolutional layers while keeping the number of parameters manageable.

Two variants of the VGG architecture exist: VGG16 and VGG19. VGG16 consists of 16 layers, including 13 convolutional layers and 3 fully connected layers, while VGG19 has 19 layers. The deeper VGG19 architecture enables it to capture more intricate and fine-grained features from the input images, but it also increases the computational complexity and memory requirements.

The typical building block in VGG is a combination of convolutional layers followed by a ReLU activation function and max pooling layers. The convolutional layers are responsible for learning local features from the input data, and the ReLU activation introduces non-linearity, allowing the network to model complex relationships between features. The max pooling layers perform down-sampling by selecting the maximum value within a small region of the feature map, reducing the spatial dimensions and increasing the network's robustness to translations and small distortions in the input.

Convolutional Layer:

$$Output = Conv(Input, Filters, Stride, Padding) \quad (3)$$

ReLU Activation:

$$Output = ReLU(Input) \quad (4)$$

Max Pooling Layer:

$$Output = MaxPool(Input, PoolSize, Stride) \quad (5)$$

D. Training and Cost Functions

During the training of both VGG and ResNet networks, the "Adagrad," "Adam," and "SGD" optimizers are sequentially employed as cost functions, with a loss variable of "binary_crossentropy" due to the two classes of our data. These functions are chosen for their satisfactory performance and widespread usage in implementing neural networks. The final function at the output level is the softmax activation function, used for multi-class classification tasks.

IV. Experimental Evaluation

A. Dataset

The experimental evaluation utilized the Breast Histopathology Images dataset [10], accessible on Kaggle[1]. The original dataset comprised 162 whole mount slide images of Breast Cancer (BCa) specimens scanned at 40x magnification. From this, 277,524 patches of size 50 x 50 were extracted, consisting of 198,738 IDC negative and 78,786 IDC positive samples. Each patch's file name follows the format: u_xX_yY_classC.png, where u is the patient ID (e.g., 10253_idx5), X and Y are the coordinates of the cropped patch, and C denotes the class (0 for non-IDC and 1 for IDC).

[1]https://www.kaggle.com/datasets/paultimothymooney/breast-histopathology-images

B. Data Subset Selection

The dataset exhibits significant inhomogeneity, with variations in image quality, tissue appearance, and staining techniques. Working with such a large and diverse dataset can lead to increased computational complexity and may also introduce noise and hinder model training.

To address this issue, a data subset selection process was performed. The objective was to create a more manageable and balanced dataset while preserving the essential characteristics of the original data. For this purpose, a random subset of the dataset was chosen for the final experiments.

Specifically, 50,000 images from each class were randomly selected and used for training and evaluation. This ensured that both classes (IDC positive and IDC negative) were equally represented in the subset, with a total of 100,000 images.

By working with a more balanced and manageable subset, the training process became more efficient and less prone to overfitting. The selected subset still contained sufficient diversity to capture the essential features and patterns related to IDC classification.

It is important to note that the data subset selection was performed randomly to avoid introducing any bias in the dataset. The random selection process helps in creating a representative sample that is more generalizable and applicable to real-world scenarios.

Overall, the data subset selection was a crucial step in preparing the dataset for training and evaluation. It allowed for more focused experiments, improving the model's ability to learn meaningful features and achieve better performance on the Breast Histopathology Images dataset.

C. Model Hyperparameter Tuning

To achieve the best performance from the neural network models, an extensive hyperparameter tuning process was conducted. The goal of this process was to find the optimal combination of hyperparameters that would lead to improved accuracy, sensitivity, specificity, and F1-score.

The hyperparameters that were considered for tuning include:

- Batch Size: The batch size determines the number of samples processed by the network in each training iteration. In this study, batch sizes of 8, 16, and 32 were explored.
- Epochs: The number of epochs represents the total number of times the entire dataset is passed through the neural network during training. A higher number of epochs may lead to better model performance, but it can also increase the risk of overfitting. The values of 10, 30, and 50 epochs were tested in this study.
- Learning Rate: The learning rate controls the step size in the gradient descent optimization algorithm. It determines how much the model's weights are updated during training. A carefully chosen learning rate is critical to ensure efficient training and model convergence. In this study, a learning rate of 0.01 was utilized.
- The optimizer plays a crucial role in adjusting the model's weights to minimize the loss function. Different optimizers offer various algorithms to achieve this optimization. The study explored three popular optimizers: Adagrad, Adam, and SGD (Stochastic Gradient Descent).

All experiments were conducted using the ReLU (Rectified Linear Unit) activation function, which is known for its effectiveness in deep learning models. The activation function introduces non-linearity, enabling the network to learn complex patterns and representations in the data.

For each combination of hyperparameters and optimizers, the models were trained and evaluated on the randomly selected subset of the Breast Histopathology Images dataset. The evaluation metrics, including accuracy, sensitivity, specificity, and F1-score, were used to assess the models' performance.

D. Results

The experimental evaluation was conducted on the Breast Histopathology Images dataset using three different network models: ResNet and VGG. The performance of these models was assessed using three different optimizers: Adagrad, Adam, and SGD. The evaluation was carried out by varying the number of epochs and batch sizes for each optimizer as shown in Table I.

Regarding the optimizer comparison, it was observed that the SGD optimizer consistently outperformed the other two optimizers in terms of overall accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F1-score. Regardless of the model architecture, SGD consistently yielded better results compared to Adagrad and Adam.

For the ResNet model, the overall performance improved as the number of epochs and batch size increased. Notably, the accuracy, sensitivity, specificity, PPV, NPV, and F1-score consistently improved, indicating that ResNet benefits from longer training and larger batch sizes.

On the other hand, the performance of the VGG model showed more variability as the number of epochs and batch size increased. While the accuracy and sensitivity improved, there were fluctuations in other metrics, suggesting that the VGG model may be more sensitive to changes in hyperparameters.

Overall, the ResNet model demonstrated better performance compared to VGG, as it achieved 8 out of 10 best performances across different configurations.

Furthermore, neither model exhibited signs of overfitting or underfitting, likely due to the incorporation of data augmentation and dropout techniques, and the selection of a homogeneous dataset.

In conclusion, the experimental evaluation showed that the ResNet model with the SGD optimizer achieved the best results for the Breast Histopathology Images dataset. The study highlights the importance of selecting appropriate optimizers and tuning hyperparameters to achieve optimal performance in deep learning tasks.

TABLE I
RESULTS FOR ADAGRAD OPTIMIZER

Net	Epochs	Batch Size	Accuracy	Sensitivity	Specificity	PPV	NPV	F1-score
Adagrad Optimizer								
Resnet	10	8	0.885	0.884	0.886	0.888	0.882	0.886
Resnet	30	8	0.892	0.891	0.894	0.896	0.890	0.893
Resnet	50	8	0.901	0.899	0.902	0.904	0.898	0.902
Resnet	10	16	0.898	0.896	0.900	0.902	0.894	0.898
Resnet	30	16	0.902	0.900	0.903	0.905	0.899	0.902
Resnet	50	16	0.907	0.904	0.910	0.911	0.902	0.907
Resnet	10	32	0.918	0.925	0.910	0.913	0.922	0.919
Resnet	30	32	0.923	0.923	0.911	0.913	0.924	0.921
Resnet	50	32	0.921	0.930	0.911	0.914	0.927	0.922
VGG	10	8	0.816	0.920	0.712	0.768	0.899	0.838
VGG	30	8	0.827	0.924	0.720	0.777	0.902	0.844
VGG	50	8	0.828	0.927	0.727	0.776	0.908	0.845
VGG	10	16	0.805	0.883	0.709	0.759	0.895	0.815
VGG	30	16	0.846	0.906	0.778	0.811	0.881	0.857
VGG	50	16	0.848	0.906	0.780	0.812	0.881	0.858
VGG	10	32	0.864	0.872	0.856	0.859	0.869	0.865
VGG	30	32	0.854	0.889	0.818	0.832	0.879	0.860
VGG	50	32	0.864	0.870	0.858	0.861	0.867	0.866
Adam Optimizer								
Resnet	10	8	0.892	0.889	0.895	0.898	0.886	0.893
Resnet	30	8	0.884	0.878	0.890	0.890	0.876	0.884
Resnet	50	8	0.898	0.889	0.907	0.907	0.889	0.898
Resnet	10	16	0.905	0.900	0.909	0.910	0.900	0.905
Resnet	30	16	0.900	0.893	0.908	0.909	0.895	0.900
Resnet	50	16	0.906	0.908	0.913	0.913	0.908	0.906
Resnet	10	32	0.910	0.906	0.913	0.914	0.906	0.910
Resnet	30	32	0.914	0.917	0.902	0.906	0.913	0.916
Resnet	50	32	0.911	0.897	0.925	0.925	0.908	0.911
VGG	10	8	0.860	0.916	0.743	0.820	0.935	0.884
VGG	30	8	0.872	0.912	0.767	0.819	0.951	0.870
VGG	50	8	0.895	0.916	0.883	0.888	0.952	0.897
VGG	10	16	0.884	0.898	0.870	0.878	0.890	0.888
VGG	30	16	0.889	0.919	0.837	0.858	0.906	0.888
VGG	50	16	0.893	0.927	0.844	0.867	0.908	0.888
VGG	10	32	0.900	0.924	0.857	0.864	0.872	0.887
VGG	30	32	0.898	0.905	0.851	0.855	0.861	0.870
VGG	50	32	0.907	0.915	0.877	0.856	0.878	0.883
SGD Optimizer								
Resnet	10	8	0.908	0.930	0.879	0.887	0.926	0.908
Resnet	30	8	0.908	0.935	0.878	0.888	0.929	0.911
Resnet	50	8	0.924	0.944	0.904	0.909	0.941	0.926
Resnet	10	16	0.898	0.926	0.862	0.872	0.918	0.895
Resnet	30	16	0.895	0.930	0.861	0.872	0.919	0.895
Resnet	50	16	0.894	0.931	0.859	0.870	0.919	0.893
Resnet	10	32	0.914	0.925	0.901	0.905	0.922	0.914
Resnet	30	32	0.923	0.931	0.914	0.917	0.929	0.924
Resnet	50	32	0.911	0.902	0.920	0.920	0.903	0.911
VGG	10	8	0.822	0.936	0.680	0.734	0.916	0.819
VGG	30	8	0.833	0.972	0.629	0.723	0.953	0.829
VGG	50	8	0.868	0.953	0.781	0.816	0.943	0.879
VGG	10	16	0.821	0.951	0.637	0.725	0.934	0.767
VGG	30	16	0.834	0.965	0.632	0.729	0.947	0.773
VGG	50	16	0.843	0.949	0.684	0.770	0.928	0.803
VGG	10	32	0.860	0.947	0.772	0.820	0.937	0.879
VGG	30	32	0.870	0.943	0.795	0.824	0.933	0.880
VGG	50	32	0.892	0.885	0.898	0.899	0.885	0.892

E. Transfer Learning

Transfer learning involves utilizing pre-trained neural networks that have been trained on large datasets for similar tasks. In this study, we explored the use of two popular pre-trained networks, VGG16 and ResNet50 to test their performance on our dataset. To implement transfer learning, we retained the main part of the pre-trained networks, excluding the last fully connected layer. The parameters (weights) of these networks were frozen during training, assuming that they have been optimized for a vast amount of data from previous usage. We finally added a new fully connected layer.

However, the results obtained from the pre-trained networks were notably different from those achieved with our two models (ResNet and VGG) trained from scratch. Table II presents the test results for both VGG16 and ResNet50.

TABLE II
TRANSFER LEARNING RESULTS

Experiment	Network	Optimizer	Accuracy	F1-score
39-BR	ResNet	SGD	0.941	0.944
39-BR	ResNet50	SGD	0.916	0.867
20-BR	VGG	Adam	0.932	0.938
20-BR	VGG16	Adam	0.823	0.845

For the ResNet architecture, our custom ResNet model (Experiment 39-BR) achieved excellent results when trained with the SGD optimizer, obtaining an accuracy of 94.1% and an F1-score of 94.4%. However, when we applied transfer learning using the ResNet50 model with the same SGD optimizer, the accuracy dropped to 91.6%, and the F1-score decreased to 86.7%. Similarly, for the VGG architecture, our custom VGG model (Experiment 20-BR) achieved competitive results when trained with the Adam optimizer, yielding an accuracy of 93.2% and an F1-score of 93.8%. On the other hand, the transfer learning experiment using the pretrained VGG16 model with the same Adam optimizer resulted in a lower accuracy of 82.3% and an F1-score of 84.5%.

V. CONCLUSIONS AND FUTURE WORK

In this study, we explored breast cancer detection using deep learning models on histopathology images. The performance of ResNet and VGG architectures was evaluated with different optimizers and hyperparameters. The ResNet architecture, optimized with the SGD optimizer, outperformed VGG in terms of accuracy and F1-score. The systematic hyperparameter tuning played a crucial role in optimizing the models' performance and achieving robust results.

Enhancing model robustness and generalization will be a priority, and we will explore advanced data augmentation techniques, such as rotation, scaling, and shearing, to augment the training dataset effectively. Investigating various data preprocessing methods will also be essential to better align the input images with the pre-trained networks, facilitating successful transfer learning.

Moreover, we will explore ensemble techniques, like model averaging and stacking, to combine predictions from multiple models [16]. Developing methods for interpretable AI will be another focus, aiming to visualize and interpret the decision-making process of our models [11]. By refining the models and pursuing these future directions, we aim to contribute to the advancement of early breast cancer detection. Our ultimate goal is to make a significant impact on patient outcomes, improving the efficiency and accuracy of breast cancer diagnosis and treatment.

ACKNOWLEDGEMENT

This research was funded by the European Union and Greece (Partnership Agreement for the Development Framework 2014-2020) under the Regional Operational Programme Ionian Islands 2014-2020, project title: "Indirect costs for project "TRaditional corfU Music PresErvation through digiTal innovation"", project number: 5030952.

REFERENCES

[1] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. R. Alnowami, and M. K. Khan. Medical image analysis using convolutional neural networks: A review. *Journal of Medical Systems*, 42(11):226, 2018.

[2] T. Araújo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polónia, and A. Campilho. Classification of breast cancer histology images using convolutional neural networks. *PLoS ONE*, 12(6):e0177544, 2017.

[3] A. D. Belsare, M. M. Mushrif, M. A. Pangarkar, and N. Meshram. Classification of breast cancer histopathology images using texture feature analysis. In *IEEE Region 10 Conference*, pages 1–5, 2015.

[4] A. Brook, R. El-Yaniv, E. Isler, R. Kimmel, R. Meir, and D. Peleg. Breast cancer diagnosis from biopsy images using generic features and svms. *IEEE Transactions on Information Technology in Biomedicine*, 2006.

[5] D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks. In *16th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 411–418, 2013.

[6] C. Elfgen, Z. Varga, K. Reeve, L. Moskovszky, V. Bjelic-Radisic, C. Tausch, and U. Güth. The impact of distinct triple-negative breast cancer subtypes on misdiagnosis and diagnostic delay. *Breast Cancer Research and Treatment*, 177:67–75, 2019.

[7] J. G. Elmore, G. M. Longton, P. A. Carney, B. M. Geller, T. Onega, A. N. A. Tosteson, H. D. Nelson, M. S. Pepe, K. H. Allison, S. J. Schnitt, F. P. O'Malley, and D. L. Weaver. Diagnostic concordance among pathologists interpreting breast biopsy specimens. *Jama*, 313(11):1122–1132, 2015.

[8] P. Filipczuk, T. Fevens, A. Krzyżak, and R. Monczak. Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies. *IEEE Transactions on Medical Imaging*, 32(12):2169–2178, 2013.

[9] Y. M. George, H. H. Zayed, M. I. Roushdy, and B. M. Elbagoury. Remote computer-aided breast cancer detection and diagnosis system based on cytological images. *IEEE Systems Journal*, 8(3):949–964, 2013.

[10] A. Janowczyk and A. Madabhushi. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of Pathology Informatics*, 7(1):29, 2016.

[11] A. Kanavos, C. Makris, and E. Theodoridis. Topic categorization of biomedical abstracts. *International Journal on Artificial Intelligence Tools*, 24(1):1540004:1–1540004:22, 2015.

[12] M. Kowal, P. Filipczuk, A. Obuchowicz, J. Korbicz, and R. Monczak. Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images. *Computers in Biology and Medicine*, 43(10):1563–1572, 2013.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *26th Annual Conference on Neural Information Processing Systems*, pages 1106–1114, 2012.

[14] G. Litjens, C. I. Sánchez, N. Timofeeva, M. Hermsen, I. Nagtegaal, I. Kovacs, C. H. van de Kaa, P. Bult, B. van Ginneken, and J. van der Laak. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific Reports*, 6(1):26286, 2016.

[15] I. E. Livieris, A. Kanavos, V. Tampakas, and P. E. Pintelas. An ensemble SSL algorithm for efficient chest x-ray image classification. *Journal of Imaging*, 4(7):95, 2018.

[16] I. E. Livieris, A. Kanavos, V. Tampakas, and P. E. Pintelas. A weighted voting ensemble self-labeled algorithm for the detection of lung abnormalities from x-rays. *Algorithms*, 12(3):64, 2019.

[17] P. Mylonas. Exploring the notion of context in medical data. *GeNeDis*, pages 39–55, 2017.

[18] A. Savvopoulos, A. Kanavos, P. Mylonas, and S. Sioutas. LSTM accelerator for convolutional object identification. *Algorithms*, 11(10):157, 2018.

[19] K. Sirinukunwattana, S. E. A. Raza, Y. Tsang, D. R. J. Snead, I. A. Cree, and N. M. Rajpoot. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Transactions on Medical Imaging*, 35(5):1196–1206, 2016.

[20] M. Veta, J. P. W. Pluim, P. J. van Diest, and M. A. Viergever. Breast cancer histopathology image analysis: A review. *IEEE Transactions on Biomedical Engineering*, 61(5):1400–1411, 2014.

[21] B. Zhang. Breast cancer diagnosis from biopsy images by serial fusion of random subspace ensembles. In *4th International Conference on Biomedical Engineering and Informatics (BMEI)*, pages 180–186, 2011.