American Journal of Interdisciplinary Research and Development
ISSN Online: 2771-8948
Website: www.ajird.journalspark.org
Volume 06, July, 2022

EFFICIENT MACHINE LEARNING ALGORITHM FOR CANCER DETECTION USING BIOMEDICAL IMAGE

Elaf Nabeel Mahdi Albu Jyash
Altınbas Unıversıty, Turkey
Email: 20320203@ogr.altinbas.edu.tr

Asst. Prof. Dr. Mesut Cevik
Altınbas Unıversıty, Turkey
Email: mesut.cevik@altinbas.edu.tr

ABSTRACT

Cancer is an epidemic that has high mortality and incidence rate worldwide. Breast cancer and oral cavity cancer is the most prevailing type of cancer in females and males, respectively. Biopsy is the only method to determine the presence of cancer with confidence which includes various processing steps such as grading, staging and visual inspection of histological slides under the microscope. The manual analysis of histopathology slides is a labor-intensive task and influenced by various factors like fatigue, attention, and expertise of pathologist. However, recent developments in soft computing techniques allow us to build an automated computer assisted diagnostic system for cancer detection which further assists pathologists in providing a reliable and consistent results on cancer diagnosis. In this context, many efforts are dedicated to feature extraction step in conventional machine learning. Deep learning is the latest advancement in this direction and opens up a new horizon in the field of machine learning. Automatic representation of the data is the key asset of the deep learning technique but require intense training and a comprehensive well-annotated dataset for their good performance. As a result of the lengthy and expensive process of collecting data from patients, obtaining large annotated datasets in health informatics is a difficult undertaking. As part of a deep learning methodology, this research proposes a CNN architecture for efficient classification. The present work proposes a new model which is applicable for both, binary as well as multi-classification of breast cancer tissue images. As a result, the suggested model does not depend on picture magnification. The beauty of this model is that the pooling layer is only present in the last convolutional layer, which aids in preventing excessive information loss. It is also possible to enhance the dataset size by using a technique called data augmentation. It's clear from a comparison of the suggested model's classification performance with the existing one that the new one is superior. To determine the influence of training in extracting significant characteristics from the images, feature maps are visually analyzed. Generalization, scalability and robustness to the image of any magnification are the powerful assets of the developed model

Keywords: Cconvolutional Neural Network, Breast Cancer Detection, Medical Image

American Journal of Interdisciplinary Research and Development
ISSN Online: 2771-8948
Website: www.ajird.journalspark.org
Volume 06, July, 2022

INTRODUCTION

As a collection of illnesses, cancer is characterized by unregulated growth and alteration in the body's cells. Approximately 1.5 million women a year are diagnosed with breast cancer, according to the WHO [1]. Five hundred and seventy thousand women died from breast cancer in 2015, or 15 percent of all cancer deaths. [1]. Women's mortality rates rose from breast cancer to lung and bronchus cancer from 1987 onwards according to the American Cancer Society (ACS). Cancer-related mortality for men and women combined were ranked first and fourth, respectively, by the American Cancer Society. Cancerous cells in the lobules or the ducts give rise to tumors. Swollen or red breasts as well as anomalies such as bloody discharge and shrinking of nipples are all signs of breast cancer [2].

[2] The majority of cancers are invasive. as well as invasive (malignant) breast cancer (malignant) Mammography and MRI, ultrasound, positron emission mammography (PEM), thermography, breast-specific gamma imaging, and tissue sampling are some of the technologies used to identify breast cancer.

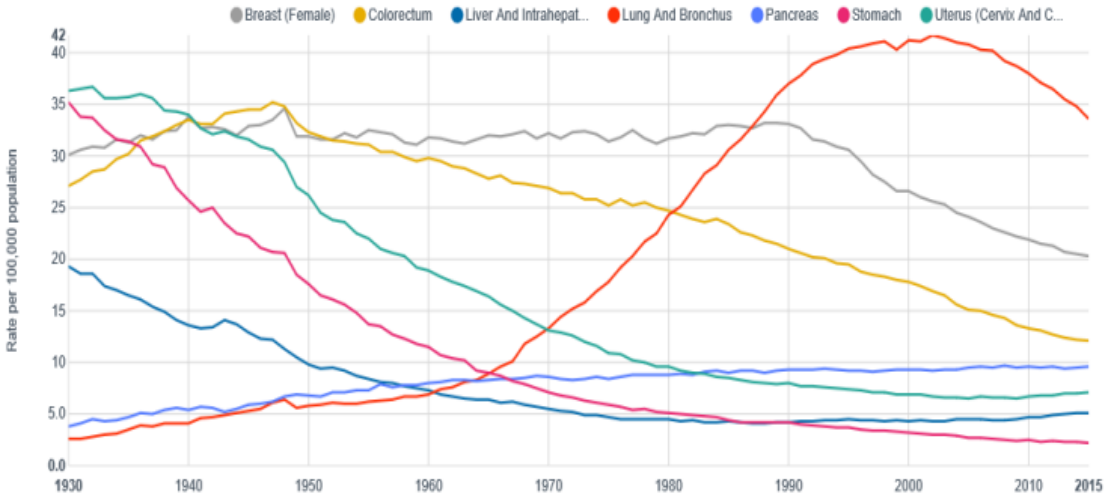

Figure 1:Trends in Death Rates (Female), 1930-2015 [2]

[4] Another option is to do a surgical incision or biopsies depending on the size and location of lump, the patient's preferences, and available resources. It's apparent that pictures play a crucial role in breast cancer detection after all these imaging screening techniques. These imaging results are being evaluated for their dispersion of various breast cancer patterns. Analyzing breast tissue is a key step in the diagnosis of breast cancer. As a way to provide a trustworthy diagnosis, histopathology slides are used to classify cancers. To categorize histological pictures in several cancer types, including benign, malignant and puzzling, manually is quite difficult. As a result of confounding tissue patterns and similar clinical manifestation, classes often imitate and share common traits [5]. When it comes to identifying a class, it might be very challenging to do so correctly. There is also a global shortage of experienced and competent pathologists [6] due to the increase in new cancer cases. Understaffed pathologists who manually analyze biopsy slides are more likely than not to make a mistake. An

automated system that not only supports pathologists but also decreases their workload while improving the accuracy of diagnosis is needed. Multi-classification systems, on the other hand, have serious challenges.

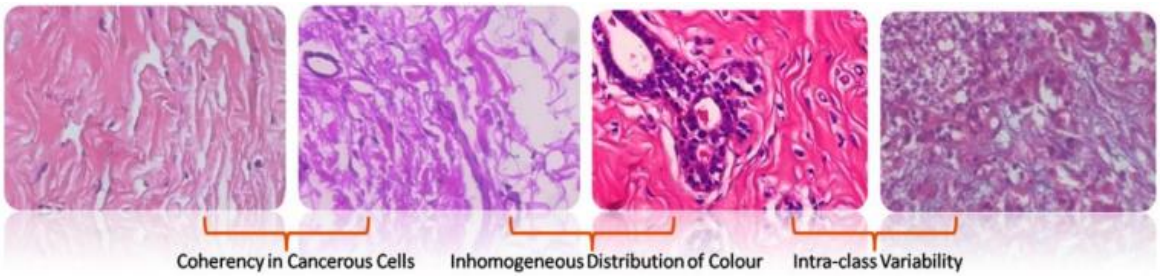

Figure 2:Mucinous carcinoma class of breast cancer histopathological images acquired at a magnification of 200X from BreaKHis dataset**.**

In the construction of an efficient automated system, handcrafted feature engineering is the key challenge. Feature engineering by hand relies heavily on pre-processing procedures, which need a high level of domain expertise to extract task-specific characteristics from images [5-6]. Many prior research has used histopathological images with texture-based features to distinguish between classes. Multiple feature descriptors such as SIFT, SURF, ORB, HOG, GRLM and wavelet-based descriptors are already in use by the research community to achieve this goal [7,8]. It's time-consuming and requires a lot of computing power, but it's worth it. To get the most significant features from segmented nuclei and glands in histological pictures, segmentation is a prerequisite, and it is only achievable if an area of interest is effectively segmented from the image. Pre-success processing's determines how well a system performs overall. Because of intra-class diversity, color distribution inhomogeneity, and malignant cells' strong cohesiveness, examining histological images is difficult. Eight kinds of breast cancer histological pictures from the BreakHis dataset are shown in Figure 2.

As the image's magnification increases, the image's graininess gets more pronounced. Images became noisier as the magnification increased. For differential diagnosis, it is difficult to extract distinct features from a digital image of biopsy slides taken at different magnification levels because of the complex background. As a result of the varying magnification levels, the background will be more uniform. For each magnification factor, an independent experiment is undertaken. Because of their inability to adapt to new magnification factors and prior knowledge of image magnification, systems developed by Spanhol et al. and Keskin et al. have major performance limitations. Since magnification levels vary, the automated system should be able to adapt to images with new magnification levels.

LITERATURE REVIEWS

In-depth knowledge of advancements and identification of gaps in the field of research is the only key to make new valuable discoveries, especially in designing automated systems. The current state of knowledge and gaps on a research topic can only be identified through an extensive review of the literature. Considering the feature extraction techniques, the literature can be divided into three main categories, namely: 1) conventional machine learning, 2) deep learning and 3) transfer learning-based works. The works in which the features are extracted manually from the images using some pre-defined feature descriptors are included in the first category of reviewed literature. The relevant studies in which features are extracted automatically from the images in a hierarchical manner are considered for the second category. In addition, the studies in which discriminating features are obtained from the pretrained networks are assigned to the third category of surveyed literature. To determine the new developments and enhancements in the field of automation for image data classifications, a survey on the above mentioned three techniques are described in brief.

An automated classification system consists of four stages: feature extraction, training, validation and testing. As with any classification system, the feature extraction stage is crucial since the recovered features have a substantial impact on the system's performance. Some studies use 'texture' of the images to classify objects, using various descriptors like gray level co-occurrence matrix (GLCM) [20], local binary pattern (LBP) [21], completed LBP (22), graph run-length matrix (GRLM) [23], histogram of gradient (HOG) [24], local phase quantizations (LPQs) [25-26], scale-invariant feature transform (SIFT) and speeded up robust features (SURF) [27-28]. SIFT descriptor is used specifically for real-time application but it imposes a large computational burden on the system. However, SURF features are computationally inexpensive and surpassed SIFT feature in terms of speed and accuracy as well [29-30].

Dheeba et al. [31] developed a computer-aided diagnosis (CAD) system to diagnose BC based on mammogram imaging modality, where the textural information is used as input. In this work, laws texture energy measures (LTEM) are computed from the region of interest (ROI) by applying five one-dimensional (1-D) convolutional kernel of length five. The images are two-dimensional structures, so require two-dimensional (2-D) convolutional kernel for its processing which are generated by convolving a 1-D vertical kernel with a 1-D horizontal kernel. The LTEM for each pixel in ROI (15x15 pixels) is identified by extracting a set of features and then a non-linear filter is applied to the feature set. In this process, 25 TEM are obtained for each pixel and normalized by zero-mean to produce a feature map of the input image. Eventually, the extracted feature map is fed to wavelet neural network for the classification which provides an accuracy of 93.671% on the dataset of 1064 mammogram images collected from 54 patients.

Keskin et al. [32], computed various morphological attributes by utilizing a waveletbased covariance descriptor to classify 14 classes of human carcinoma cell lines images, where 7 classes belong to breast cancer and rest of 7 classes to liver cancer. In terms of features, the authors found dual-tree complex wavelet transform (DT-CWT) coefficients as an optimal choice because edges are the most relevant feature in the carcinoma cell line images. DT-CWT has a great potential to describe the edges at multiple orientations, so it provides more discriminative features from the images for the efficient analysis of cancer cell morphology. In this work, segmentation is performed before the extraction of features from the images because the images in the dataset consist of a large amount of background pixel. Expectation-Maximization (EM) algorithm is applied after modelling the image as a mixture of two Gaussians and obtained noisy patterns. Morphological operation 'closing' and median filtering is applied to reduce the noise. The square windows are randomly selected from the image and a covariance matrix is computed for each window. An SVM classifier with radial basis function (RBF) is then trained over the computed matrices which achieved 98% accuracy on the dataset of 840 images.

Wan et al. [33] proposed two new variant of LBP feature for classification of breast tissues from optical coherency microscopy (OCM) images. These variants are based on two different organizations of neighborhood blocks around a pixel, known as spoke LBP and ring LBP. The major drawback of traditional LBP lies in their inability in extracting information from the pixels because it utilizes the intensity difference only between the center pixel and single pixels from the neighbour. But in both spoke and ring LBP, average intensity of pixels is computed for a considered shape of blocks in the neighborhood of the central pixel and provides more information related to the texture of an image. Further, a multi-scale blocked LBP is also constructed by combining the spoke and ring LBP with different radius parameters and claimed an accuracy of 92.4% using a shallow neural network classifier on a dataset of 4310 OCM images with 5 classes.

There is also a model called the one class kernel principal component analysis (KPCA) introduced by Zang et al. To make the decision, the trained KPCA models were combined with images from other classes and the same operation was repeated. On histopathology pictures of BC, the KPCA approach had a 92 percent accuracy rate, which marked a new turning point in the research.

Öztürk et al. [35] provided a comparative analysis in order to determine the most successful feature extraction algorithm for cell detection in histopathological images. The authors considered six algorithms: HOG, LBP, SIFT, FAST, CANNY and maximally stable extremal region (MSER). Some pre-processing steps are also applied to the raw images before applying the feature extraction algorithm. The RGB (Red Green Blue) color space is converted into HSV (Hue Saturation Value) color space and wiener filtering is performed on the V component of the image to reduce the noise. The performance of all the considered algorithms are examined and evaluated on 12

histopathological images. The results reveal that MSER is the best feature extraction algorithm, followed by HOG due to better cell detection ability of these algorithms. However, the perception of closely associated cells as a single cell is a major limitation of MSER.

Based on the literature, it is concluded that no single algorithm suits to the classification of all types of histopathological imaging data. The selection of the right algorithm greatly depends on the type of problem. An efficient algorithm can be selected by testing multiple algorithms under different settings of the hyper-parameters. The major challenges observed in conventional machine learning algorithms are complexity in data models, the problem of overfitting, cost of computational techniques, parameter tuning and the scaling of data to the number of classes, which opened up a new window to overcome these issues with an alternative mean. To address these challenges, a modified and new strategy is required for classification of histopathological data which become possible through deep learning. The implementation of strategies like batch normalization, dropout, greedy layer-wise learning, regularization, different pooling methods, transfer learning, layer-wise fine-tuning assists in boosting the performance of models with a significant margin. To further enhance the performance of these networks, strategies of data and model parallelism are also need to be developed which can be possible with the hybridization of algorithms to obtain better results in term of training speed. Besides, filter size, stride value, activation function, optimization technique, the number of the convolutional and fully connected layer are other main factors which require special attention during the execution of an algorithm to enhance the performance of networks.

PROBLEM STATEMENT

Despite recent advances in molecular biology with the advent of genetic markers, the histopathological analysis always remains a ‘gold standard’ in the diagnosis of BC [13-15]. In the process of histopathological analysis, pathologists visually inspect the histology slide under a microscope to determine a particular pattern that they learned from the previous slides of the disease. Hence, a visual inspection of histology slides is based on pattern recognition approach and subjective opinion of the pathologists. Analysis of histology slide is a time-consuming task and requires great care in the interpretation of the disease [10, 16]. New cases of cancer are projected to rise globally. Due to matter the fact, pathology services are continuously pressurized to provide quality analyses. There is a severe shortage of pathologists, which makes it difficult to quickly analyze histology slides, putting a heavy pressure on the few pathologists that are still accessible. In sub-Saharan Africa and China, there is one pathologist for every 100,000 and 130,000 persons, respectively [17]. In both India and the United States, the situation is identical (US). In India, one pathologist is available for every 65,000 people, but in the United States, there are 5.7 pathologists for every 100,000 people

[18]. To address the challenge of a potential workforce crisis, there is a need to enrich the critical services that histopathology offers to cancer care.

In recent time, digitization of tissue samples has become possible due to digital pathology and computational algorithms which includes machine learning and image processing techniques for analyzing the image [19-20]. The computational algorithms in digital pathology are employed to detect fine details that not easily be determined by the human eye. New machine learning techniques have also been evolved and derived new innovations in healthcare by assisting in developing computer-aided detection and diagnosis systems. High-performance computing system and low-cost memory devices are the fundamental factors behind the evolution of soft-computing techniques [21]. Improvement in image processing techniques and machine learning algorithms also provide potential to develop histopathological image-based computer-aided detection and diagnosis systems that can add a new value in pathology services by making it more objective, productive and coherent in diagnosis. These systems can also be used as a second opinion in pathology in order to complement the decision of the pathologists. In near future help in maintaining the data concerning the variety of cancer types to diagnose the sufferer form healthy one.

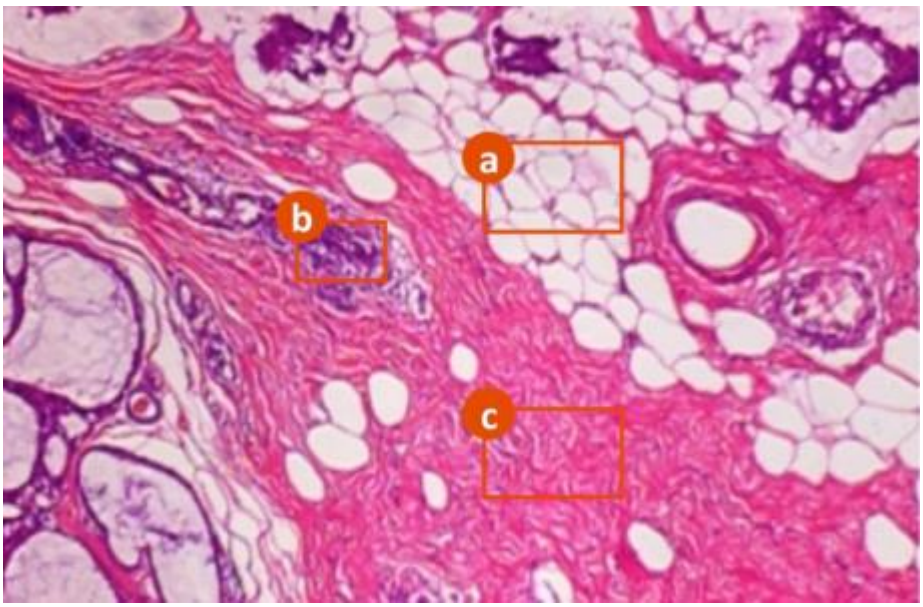

Figure 3: Representation of different tissues: (a) fat tissue, (b) cell nuclei, and (c) collagenrich stroma in an example of Mucinous Carcinoma at 100X.

PROPOSED SYSTEM

The prime aim of this research work is to assist pathologists by proposing a proficient and accessible automated diagnosis system for BC. The system must capable to discriminate BC as benign or malignant in conjunction with the classification of breast lesions in sub-ordinate classes of benign and malignant to determine the location of

American Journal of Interdisciplinary Research and Development
ISSN Online: 2771-8948
Website: www.ajird.journalspark.org
Volume 06, July, 2022

BC. Additionally, the proposed system must be robust to the magnification level of the digital histopathological images for a reliable diagnosis.

A magnification independent classification model for the identification of BC histopathology pictures was created to give an effective, reliable and resilient automated system for binary as well as multi-classification. As a result of the hierarchical representation of learning semantic and discriminating features from low-level to high-level, CNN does not require handcrafted feature engineering. In the present work, we have determined the potential of proposed magnification independent CNN in binary and multi-classification of BC using histopathological image.

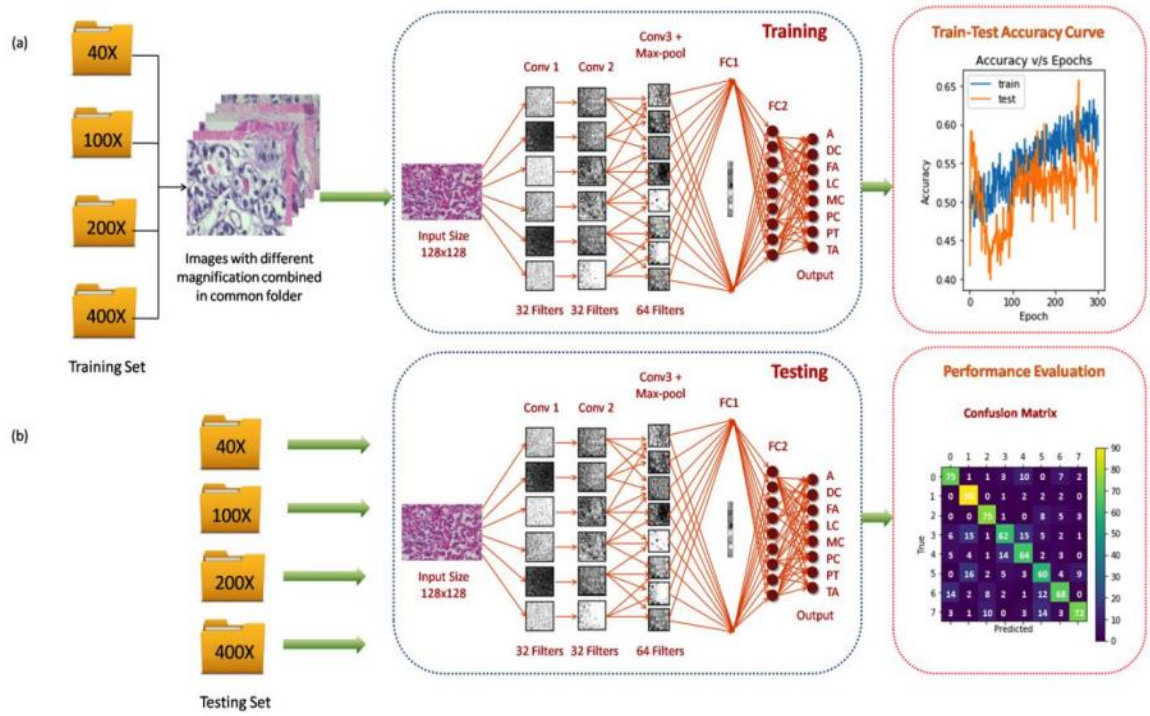

Figure 4: Schematic representation of (a) training and (b) testing protocol in conjunction with performance evaluation metrics used in magnification independent multi-classification.

METHODOLOGY

A new CNN architecture based on a predetermined training strategy has been suggested. Image samples for each class are equally distributed throughout a stratified training set. In order to get an equal distribution of images across all magnification settings, balancing was performed prior to training. In order to make all classes equal, they are randomly down sampled. Classes with the fewest picture samples are left alone. Down sampling is applied to each magnification level's classes in order to reduce their size. In order to create a magnification independent binary as well as multi-classification system, all the photos of the dataset are placed in a same folder

without considering their magnification level. However, independent tests have been conducted to assess the model's capacity to classify data based on magnification. Figure 4 shows a diagrammatic representation of the proposed technique. A 90 percent training set and a 10 percent testing set have been created from the dataset. There is a 75/25 split between a training and validation set. Both balanced training and validation sets have been augmented with 900, 1800, and 2700 rotations of each image [9].

Table 1: Description of CNN Architecture

Layer	Type	Activation	Regularization	No. of Filters	Filter Size	Pooling Size/Stride	Zero padding
Conv1	Convolutional	ReLU	-	32	3x3	-	-
Conv2	Convolutional	ReLU	-	32	3x3	-	2x2
Conv3	Convolutional	ReLU	-	64	3x3	-	2x2
Max-Pool	Pooling	-	-	-	-	2x2/2	-
FC1	Fully-Connected	ReLU	Dropout	128 nodes	1x1	-	-
FC2	Fully-Connected	ReLU	Dropout	128 nodes	1x1	-	-

CNN TOPOLOGY

In CNN, visual representations are identified directly from images [160]. Tensorflow framework was used to implement the proposed CNN. 224x224 input picture is mapped to 56x56 feature maps in the first two convolutional layers (Table 1). A 3x3 kernel with 32 filters is used in both convolutional layers. We add another convolutional layer to the network (Conv3). This layer is comprised of 32 filters of size 3x3. This is followed by one layer of pooling and flattening in the architecture. Maximum pooling is done by utilizing a kernel with a stride value of 2 and a kernel size 2x2. For the second and third convolutional layers, no padding is applied because a convolutional operation always results in a smaller image. A flattened layer of 50,176 features is followed by two completely connected levels (FC1 and FC2). The entire network has 128 completely connected nodes in each layer. A softmax layer is added in the end to acquire the desired outcome. ReLU activation is applied to the entire convolutional and fully connected layer for nonlinearity in the network.

Dataset

Broken Histopathological Images is the largest dataset of histopathological images of breast cancer. It was developed in partnership with Parana's Prognostics and Diagnostic (P&D) Laboratory. At four different magnification levels, the collection contains 7909 histological pictures of breast cancer from 82 patients, totaling 7909

photos in total (40X, 100X, 200X, and 400X). Thereafter, paraffin was used to encapsulate the sampled tissue. Paraffin is microtome into sections of 3-5 m thickness, which are subsequently mounted on glass slides [11]. H&E-stained segment to emphasize the tissue's structure of interest the nuclei and the cytoplasm [4]. The samples in the dataset are divided into two major classes: i) benign, and ii) malignant. The benign class consists of 2480 image samples while malignant class is composed of 5429 samples. Each class is further partitioned into sub-classes. Adenosis, fibro-adenoma, phyllods tumor, and tubular adenoma are the four sub-classes of the 'benign' class. On the other hand, the "malignant" class includes: Carcinoma of the ducts, Lobular Carcinoma, Mucinous Carcinoma, and Papillary Carcinoma. An equal number of pictures from each class is utilized in a stratified training set to train the network. With the amount of photos related with each magnification factor, Figure 6 shows a thorough distribution for BreakHis' images.

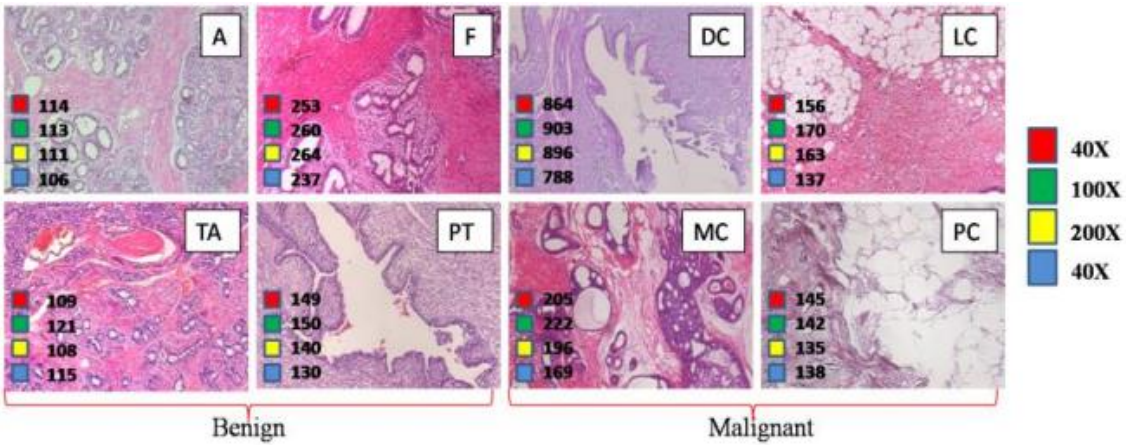

Figure 5: An illustration of data organization in the BreakHis Dataset

TRAINING PROTOCOL

As the first phase of a multi-classification system, training is vital since a model can only function efficiently if it is well-trained. This can be done by measuring the difference between intended and predicted production. This mistake is measured using a variety of loss functions, the most popular of which is the cross-entropy:

$$Cross\ Entropy\ Loss\ (CEL) = -\sum_{c=1}^{N} y_c\ log(P_c) \qquad (1)$$

where N is the number of classes, where yc indicates proper classification, and pc represents predicted probability observation of class c. As a positive, non-zero function, CEL becomes null if it predicts exactly the same output as desired. In order to minimize this loss function, it is necessary to utilize the backpropagation algorithm. When this is done, the Adam optimizer is used to optimize the loss function. A single scalar value must be returned from the CEL function in order to regulate optimization. It is computed and averaged for all photos in this context. This averaged CEL is then used in the categorization procedure.

Many trial-and-error experiments are used to identify the network parameters used in training. According to the results of the investigation, a learning rate of 0.0001 is a fair decision because a lower learning rate slowed down the convergence process. The converse is also true: a faster learning rate leads to a failure of convergence. With a standard deviation of 0.05, weights are randomly initialized in this network. A 32-bit mini-batch is recommended to prevent a computer crash and memory scarcity. It is only once the validation set has been used to its fullest precision that the training process comes to a close.

PERFORMANCE METRICS

The robustness of the generated breast cancer classification model is determined by evaluating the image recognition rate for each magnification independently. It's possible to define the image recognition rate as the ratio between successfully classified cancer photos and the total number of images in the test set [12], denoted as [13]:

$$Image\ Recognition\ Rate = \frac{correctly\ classified\ images\ of\ cancer}{total\ number\ of\ images\ in\ test\ set} \quad (2)$$

If you want to do multi-classification, then the training set for each experiment consists of a total of 2,400 pictures (400 images per class). There are 800 pieces in each group. Binary-classification requires 3,000 photos (1500 each class), 800 images and 800 images accordingly for the training set, validation set and testing set. Based on the confusion matrix and activation levels, model performance is evaluated in order to offer a visual analysis for their subsequent computation and display, respectively.

RESULT AND DISCUSSION

Using the suggested model to classify BC histopathology photos, the findings are shown in this section. There's also a performance evaluation for each trial to determine whether or not the suggested model can classify histopathology pictures regardless of magnification.

BINARY CLASSIFICATION

Second, the suggested model is compared to existing state-of-the-art techniques. Second, a system's additional performance is examined in the third section.

COMPARISON WITH PREVIOUS WORK

An experiment was conducted to evaluate the performance of the suggested model for magnification-independent classification. A data augmentation technique, which rotates images about their center at 900, 1800, and 270° [11-12], is also used to expand the dataset size. Table 2 shows the binary classification accuracy for breast cancer tissue pictures at the image level:

American Journal of Interdisciplinary Research and Development
ISSN Online: 2771-8948
Website: www.ajird.journalspark.org
Volume 06, July, 2022

Table 2: Image Recognition Rate at Image Level (Magnification Independent)

Recognition Rate at	Layers				
	40x	100x	200x	400x	Average
Image Level	90.4±1.5	86.3±3.3	83.1±2.2	81.3±3.5	85.3

For example, in [19], four methodologies were used to produce image patches, which were then used for training: When all of the photos from BreakHis' database were mixed up, the classification performance of the model was improved, as can be shown in Table 4. As the training data grows in size, the model's performance improves. Pooling is only applied at the last convolutional layer, hence it prevents information loss at a low level. To build high-level features, convolutional neural networks merge low-level elements. When it comes to texturing, low-level qualities are crucial.

Table 3: Image Recognition Rate at Image Level [19]

Recognition Rate at	Strategy	Layers			
		40x	100x	200x	400x
Image Level [19]	1	79.9±1.5	80.8±1.5	**84.0**±1.5	80.7±1.5
	2	80.6±1.5	81.0±1.5	82.7±1.5	**80.8**±1.5
	3	81.8±1.5	82.3±1.5	82.4±1.5	80.3±1.5
	4	**89.6**±1.5	**85.0**±1.5	82.8±1.5	80.2±1.5

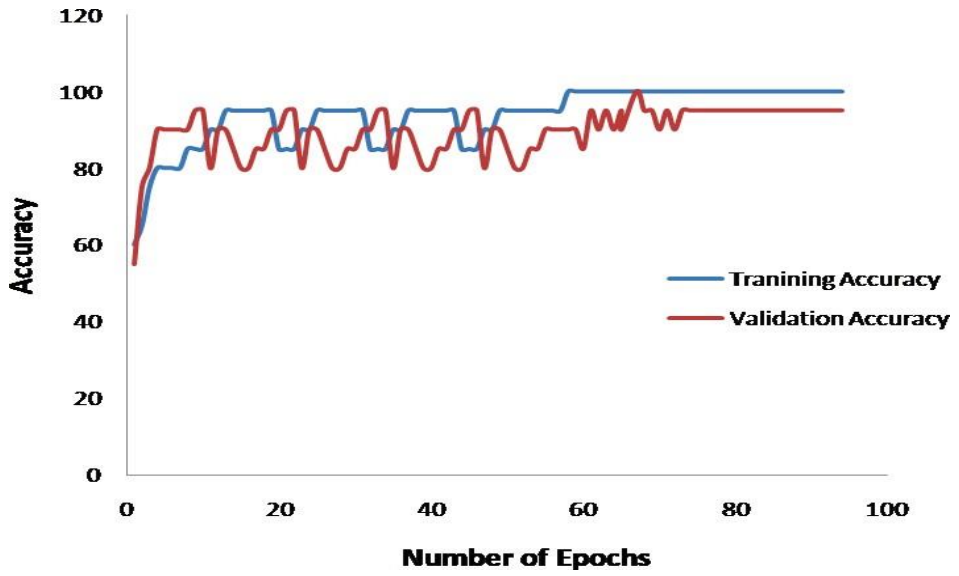

Table 4: Comparison of Model Performance between [19] and the proposed approach based on Average Image Recognition Rate

Recognition Rate at	[19]	Proposed Approach
Image Level	84.4	85.3

Figure 5: Training and validation accuracy curves at the time of training

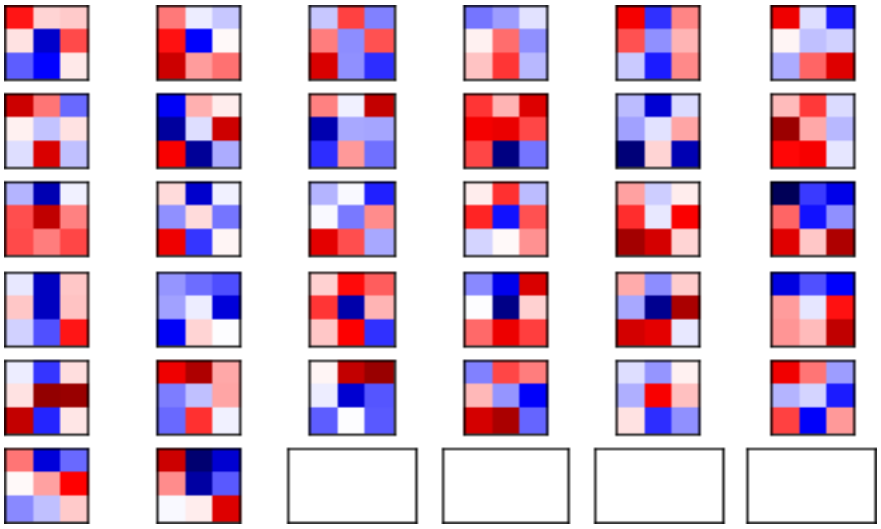

Figure 6: 32 Filters from the first convolution layer of the proposed model with kernel size (3x3)

SYSTEM PERFORMANCE ANALYSIS

To enhance understanding, a collection of performance curves obtained during the model's training is provided. FIGURE 6: Training Accuracy and Validation Accuracy of a Model. Figure 7 shows the first convolution layer's 32-filter kernel. At a lower representation level, these filters' kernels extract characteristics like edge and corner, then combine them at a higher representation level to learn complex features. According to Figure 8, the test set comprises of 40x magnification photographs of breast cancer tissue. As a result of similar textures or appearances, benign and cancerous breast tissue are often mistaken for one another. Some of the photos in Figure 9 have been misclassified. In order to avoid confusion between the two classes, a detailed description of texture is essential.

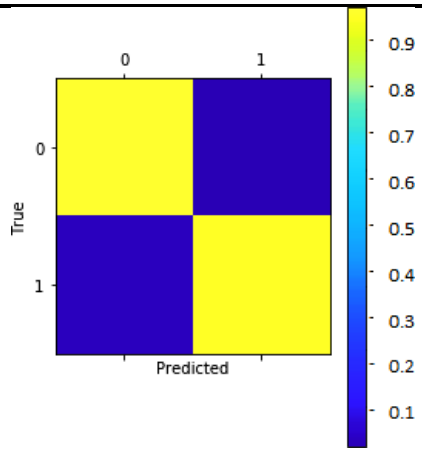

Figure 7: Confusion Matrix for the Test Set with Magnification Factor (40x) 0: Benign and 1: Malignant

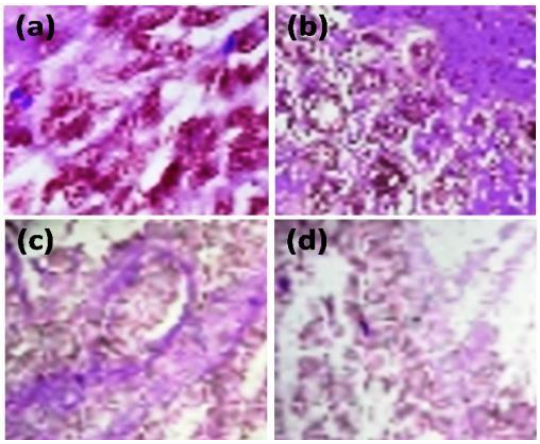

Figure 8: Misclassified Images of Breast Cancer Tissue, (a) & (b) are Benign but Classified as Malignant, and (c) & (d) are Malignant but Classified as Benign

A broad framework based on CNN is offered to extract characteristics from breast cancer histology photographs. The proposed framework is more resilient and faster than existing classification methods since it only requires one training phase. With the magnification-independent model, the image can be resized to a new magnification. In order to achieve this, the training set contains images with new magnifications that are used in the training process. By comparison with the previously trained data, it is therefore relatively easy to generalize the suggested framework to new images with varied magnifications. According to BreakHis's dataset, the proposed CNN performed well at all magnification levels in binary classification. of 90.4% (40X), 86.3% (100X), 83.1% (200X), 81.3% (400 X), and proved as state-of-the-art technique by outperforming the other pre-existing works.

CONCLUSION

To extract features from breast cancer histology pictures, a general framework based on CNN is given. One training step makes the proposed framework more robust and

faster, while existing classification systems use numerous training steps to distinguish photos of specific magnification. An additional benefit of the magnification independent approach is its ability to scale the image. In order to achieve this, the training set contains images with new magnifications that are used in the training process. By comparison with the previously trained data, it is therefore relatively easy to generalize the suggested framework to new images with varied magnifications.

REFERENCES

1. F. A. Spanhol, et al., "A dataset for breast cancer histopathological image classification," IEEE Transactions on Biomedical Engineering, vol. 63, pp. 1455-1462, 2016.
2. F. A. Spanhol, et al., "Breast cancer histopathological image classification using convolutional neural networks," in International Joint Conference on Neural Networks, 2016, pp. 2560-2567.
3. S. Sharma, et al., "Detection and Analysis of Lung Cancer Using Radiomic Approach," in Smart Computational Strategies: Theoretical and Practical Aspects, 2019, pp. 13-24.
4. Z. Han, et al., "Breast cancer multi-classification from histopathological images with structured deep learning model," Scientific Reports, vol. 7, pp. 1-10, 2017.
5. D. Shen, et al., "Deep learning in medical image analysis," Annual Review of Biomedical Engineering, vol. 19, pp. 221-248, 2017.
6. S. Doyle, et al., "Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer," BMC Bioinformatics, vol. 13, p. 1-15, 2012.
7. A. Belsare, et al., "Classification of breast cancer histopathology images using texture feature analysis," in TENCON IEEE Region 10 Conference, 2015, pp. 1-5.
8. T. J. Alhindi, et al., "Comparing LBP, HOG and deep features for classification of histopathology images," in International Joint Conference on Neural Networks, 2018, pp. 1-7.
9. S. Sharma and R. Mehra, "Automatic Magnification Independent Classification of Breast Cancer Tissue in Histological Images Using Deep Convolutional Neural Network," in International Conference on Advanced Informatics for Computing Research, 2018, pp. 772-781.
10. F. A. Spanhol, et al., "A dataset for breast cancer histopathological image classification," IEEE Transactions on Biomedical Engineering, vol. 63, pp. 1455-1462, 2016.
11. F. A. Spanhol, et al., "Breast cancer histopathological image classification using convolutional neural networks," in International Joint Conference on Neural Networks, 2016, pp. 2560-2567.

12. F. A. Spanhol, et al., "Deep features for breast cancer histopathological image classification," in IEEE International Conference on Systems, Man, and Cybernetics, 2017, pp. 1868-1873
13. M. A. Al-Fatah, "Importance of histopathological evaluation of appendectomy specimens," Al-Azhar Assiut Medical Journal, vol. 15, no. 2, pp. 97-103, 2017.
14. [K. Sirinukunwattana, et al., "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," IEEE Transaction on Medical Imaging, vol. 35, pp. 1196-1206, 2016
15. M. L. Wilson, et al., "Access to pathology and laboratory medicine services: a crucial gap," The Lancet, vol. 391, no. 10133, pp. 1927-1938, 2018.
16. S. J. Robboy, et al., "Pathologist workforce in the United States: I. Development of a predictive model to examine factors influencing supply," Archives of Pathology and Laboratory Medicine, vol. 137, no. 12, pp. 1723-1732, 2013.
17. N. Stathonikos, et al., "Going fully digital: Perspective of a Dutch academic pathology lab," Journal of Pathology Informatics, vol. 4, no. 15, pp. 1-25, 2013.
18. V. Sze, et al., "Efficient processing of deep neural networks: A tutorial and survey," Proceedings of the IEEE, vol. 105, no. 12, pp. 2295-2329, 2017
19. [F. A. Spanhol, et al., "Breast cancer histopathological image classification using convolutional neural networks," in International Joint Conference on Neural Networks, 2016, pp. 2560-2567.
20. R. M. Haralick and K. Shanmugam, "Textural features for image classification," IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-3, no. 6, pp. 610-621, 1973.
21. S. Liao, et al., "Dominant local binary patterns for texture classification," IEEE Transactions on Image Processing, vol. 18, pp. 1107-1118, 2009.
22. Z. Guo, et al., "A completed modeling of local binary pattern operator for texture classification," IEEE Transactions on Image Processing, vol. 19, pp. 1657-1663, 2010.
23. A. Belsare, et al., "Classification of breast cancer histopathology images using texture feature analysis," in TENCON IEEE Region 10 Conference, 2015, pp. 1-5.
24. T. J. Alhindi, et al., "Comparing LBP, HOG and deep features for classification of histopathology images," in International Joint Conference on Neural Networks, 2018, pp. 1-7.
25. V. Ojansivu and J. Heikkilä, "Blur insensitive texture classification using local phase quantization," in International Conference on Image and Signal Processing, 2008, pp. 236-243.
26. A. Bosch, et al., "Image classification using random forests and ferns," in IEEE International Conference on Computer Vision, 2007, pp. 1-8.
27. A. Bosch, et al., "Scene classification via pLSA," in European conference on Computer Vision, 2006, pp. 517-530.

28. L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, pp. 524-531.
29. H. Bay, et al., "Surf: Speeded up robust features," in European Conference on Computer Vision, 2006, pp. 404-417.
30. E. Rublee, et al., "ORB: An efficient alternative to SIFT or SURF," in IEEE International Conference on Computer Vision, 2011, pp. 1-8.
31. J. Dheeba, et al., "Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach," Journal of biomedical informatics, vol. 49, pp. 45-52, 2014;
32. F. Keskin, et al., "Image classification of human carcinoma cells using complex wavelet-based covariance descriptors," PloS One, vol. 8, pp. 1-10, 2013.
33. S. Wan, et al., "Spoke-LBP and ring-LBP: New texture features for tissue classification," in IEEE International Symposium on Biomedical Imaging, 2015, pp. 195-199.
34. Y. Zhang, et al., "One-class kernel subspace ensemble for medical image classification," EURASIP Journal on Advances in Signal Processing, vol. 2014, pp. 1- 13, 2014.
35. Ş. Öztürk and A. Bayram, "Comparison of HOG, MSER, SIFT, FAST, LBP and CANNY features for cell detection in histopathological images," HELIX, vol. 8, pp. 3321-3325, 2018.