DOI: 10.51153/kjcis.v7i2.226

Enhancing Histopathological Image Classification: Optimal Fine-Tuning of Convolutional Neural Networks with Feature Extraction

Muhammad Hamza Mahmood, [a]Muhammad Hasnain [b],* Rizwan Mehmood [c], Sardar Usman [d], Rana Zeeshan Zulfiqar [e]

a,b,eLahore Leads University, Department of Computer Science, Kamahan Road, Lahore, 54910.
cLahore Leads University, Department of Information Technology, Kamahan Road, Lahore, 54910.
d Grand Asian University Sialkot, Department of Computer Science and Software Engineering, Sialkot, 51260

***Corresponding author*:** *Muhammad Hasnain, drhasnain.it@leads.edu*

Abstract

The area of medical image analysis has genuine obstacles, such as small data sets, complicated fine-tuning, and choosing the right architecture. This paper proposes a CancerVisionNet model based on the convolutional neural network (CNN) architecture with many layers to extract and classify features from cancer images. To train the CancerVisionNet model and avoid overfitting, data augmentation is carried out using a dataset consisting of 220,025 images. The proposed CancerVisionNet model is evaluated on the PatchCamelyon dataset. Its remarkable area under the receiver operating characteristic (ROC) curve (AUC) measure is about 0.9. Compared to the other studies, the CancerVisionNet model stands out with a higher accuracy (95.4%). Moreover, this work demonstrates the potential of CNNs in medical image analysis, providing an effective approach to enhance classification accuracy and paving the way for further advances in the field. Although the results of this study pertain to histopathology and the PatchCamelyon dataset, the potential for a broader application awaits cross-domain validation. Future research works can include exploring alternative architectures and scalability to larger datasets.

Keywords: Land Use / Land Cover, Crop Classification, Remote Sensing, Machine Learning

1. Introduction

The role of medical images in patient care is critical, but resource constraints, a lengthy decision-making process, and the need for the second opinion can hamper the process. Medical imaging applications such as pathology urgently require improved and more precise image classification. Histopathological images play a crucial role in the diagnosis and assessment of the aggressiveness of diseases such as cancer, as well as in determining whether they are benign. Histopathology involves examining biopsied tissue samples microscopically to diagnose illnesses [1], making it an essential tool for physicians to plan patient care. Histopathology image classification is a specialized area of study for pathologists. A veteran pathologist can help with the

classification of histopathology images. Over the past decade, the number of practicing pathologists has decreased by 17.5%, while the workload has increased by 41% [2]. Consequently, there is an urgent need to equip pathologists with an autonomous classifier, capable of accurately classifying histopathology images to help them in their work.

Machine learning is a cutting-edge technique in the area of artificial intelligence that enables the automated extraction of information from images. While in deep learning, neural networks are used with numerous hidden layers to achieve the same goal. Recent advancements in machine learning have made it feasible to create an algorithm that can automatically extract visual features [3], One potential use of deep learning is image classification, where it is used to automatically categorize images. A convolutional neural network (CNN) is trained to classify images by adjusting its weights according to the training dataset. The primary advantage of CNN is that it can automatically map useful aspects of images for classification without any further programming. CNNs have been shown to be effective in classification in numerous medical fields, including the detection and classification of diabetic retinopathy [4], the detection of Alzheimer's disease [5], and the identification of skin lesions [6, 7].

Classifying images according to predetermined labels is an important procedure in many disciplines, including medicine. We might think of deep learning algorithms as self-learning programs that figure out how to differentiate between various image classes without any assistance from human feature engineers. We are inspired by Fukushima's proposed convolutional neural networks (CNNs) as the first effort to construct an autonomous classifier capable of learning to discriminate between classes [8]. However, the quantity and complexity of the dataset at the time of this endeavor limited its success. AlexNet, a CNN architecture created by Krizhevsky, Sutskever and Hinton [9] that achieved a 16% error rate, beat out the runner-up, which achieved a 25% error rate, in the "ImageNet Large Scale Visual Recognition Challenge" ILSVRC competition [10]. Fig. 1 shows a typical feedforward artificial neural network, such as CNN.

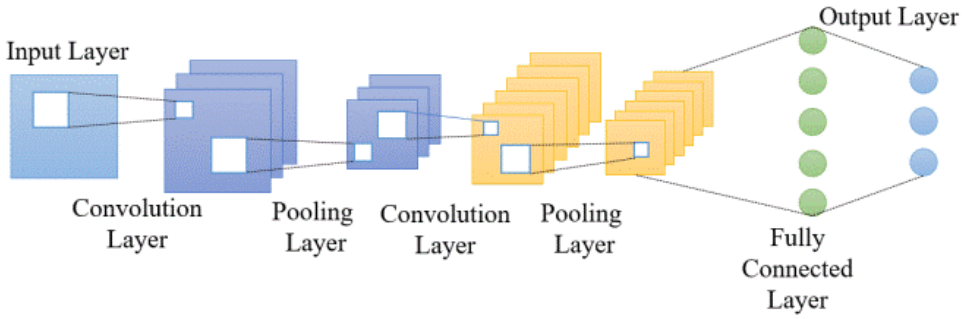

Fig. 1 CNN's architecture [11]

In the realm of image classification, challenges such as overfitting can emerge, especially with feed-forward neural networks that require a significant number of training images. Such limitations can restrict the use of CNNs in specific domains, such as natural image categorization [11]. However, the solution lies in the concept of transfer learning, which has gained prominence as a distinct field. Transfer learning involves repurposing weights from a pre-trained network that was initially designed to analyze an entirely different large dataset. This approach has the potential to revolutionize healthcare by addressing the data dearth in medical image categorization using CNNs. Chollet [12] names feature extraction and fine-tuning as the two most common approaches to transfer learning. Yosinski, Clune, Bengio and Lipson [13] emphasized that the usefulness of these methods is proportional to the quantity of the dataset. When the dataset is sufficiently extensive, it allows for fine-tuning of the original layers; conversely, if the dataset is smaller, these layers can be repurposed for feature extraction. Another crucial consideration is the degree of similarity between the source and destination datasets. In alignment with these concepts, our study opts to prioritize fine-tuning over feature extraction for several reasons. The dataset used boosts up a considerable number of images, from several image domains. However, there is a huge difference between the histopathology dataset and the other relevant datasets.

Fine-tuning a CNN design can be a time-consuming task that requires specific equipment. However, refining the entire network does not always guarantee optimal performance. In this research, we discuss the implications of employing a cutting-edge architecture to fine-tune convolutional neural networks (CNNs) using the histopathology dataset. Our objective is to extract valuable information on the most effective fine-tuning strategies to achieve precise medical image classification. To realize this goal, in this work, applying different learning rates to state-of-the-art CNN architectures, we used the area under the ROC curve (AUC) as a performance metric. The effectiveness of each CNN architecture was independently assessed on its own, separate, and unlabeled test set [14].

In this study, some challenges that have been observed in medical image analysis are tackled. These challenges include data scarcity, features extraction, and model architecture selection. The main contributions of our work are as follows:

- We propose cancerVisionNet, a deep CNN model optimized for histopathological imaging and its classification. By using such a model, we are able to perform feature extraction and classification of images from cancerous tissues, hence enhancing diagnostic performance.
- To address overfitting and lack of data, we incorporate detailed data augmentation measures to avoid data shortage. Such techniques increase the training data volume to 220,025 images, thus improving the models' ability to generalize and empower the model for overcoming the

overfitting. In this work, we propose to comprehensively evaluate the performance of CS-based CancerVisionNet on the popular PatchCamelyon dataset. According to the receiver operating characteristic (ROC) analysis the model got an AUC value of around 0. 90 and a classification accuracy of 95.4% are quite satisfying results.

- The subsequent methodologies and results serve to establish the foundation for future works that consider other architectures and explore more than the above methodologies and results. CancerVisionNet remains rather flexible suggesting that the proposed framework could be useful in any other medical imaging domains other than histopathology.

The layout of this study is as follows.

Section 2 gives an overview of the related literature on the research topic. In Section 3, we detail our research methodology, and in Section 4, we describe the empirical results obtained in this study. Section 5 gives us discussion on results and comparison of this study with other works. Section 6 concludes the results of the research article.

2. Literature review

The urgent need to provide support to pathologists and provide additional information has led to the prominence of the use of deep learning systems for the classification of medical images. Specifically, Mehra [15] undertook an in-depth investigation into the implications of transfer learning using CNN's architectures, such as ResNeT50, VGG16, and VGG19. Their study revolved around the Break-His dataset [16], encompassing 7,909 histopathology images related to breast cancer. Concentrating on binary classification, the researchers substituted the original ultimate classification layer of each architecture with a logistic regression classifier. Researchers utilized various measures, including precision, F1 score, accuracy, and region under the ROC bend (AUC), to fully evaluate performance. They observed that the VGG16 architecture performed best in the first splitting procedure (95.6%), followed by the VGG19 architecture (91.85%). However, adding transfer learning to the ResNet architecture did not improve its performance.

Kassani, Kassani, Wesolowski, Schneider and Deters [17] developed a new approach for labeling histopathological images using PatchCamelyon [18, 19], Break-His [16], Bach and Bio-Imaging datasets during model training and validation [20]. The authors introduced an innovative binary-classification ensemble model that harnessed VGG19 [21], MobileNet [22], and DenseNet [23] CNN architectures, achieving remarkable accuracies of 98.13%, 95%, 94.64%, and 83.10% across Break-His, Bach, PatchCamelyon, and Bio-Imaging, respectively. To counteract overfitting, various image augmentation techniques were employed, encompassing horizontal and vertical

flipping, zooming, and rotation, effectively expanding the training dataset. Using a learning rate of 0.0001% and a batch size of 32, the models went through 1000 iterations after the images were scaled to 224x224 dimensions using the Adam optimizer. The ensemble model exhibited improved performance on different datasets.

Vesal, Ravikumar, Davari, Ellmann and Maier [24] investigated how transfer learning affected the Bach dataset and their research encompassed a multiclass classification strategy, focusing on a maximum of four classes. A comparative analysis was conducted for the effectiveness of InceptionV3 [25] and ResNet50 [26] CNN architectures. To train these CNN models, the researcher used stochastic gradient descent with a learning rate of 0.0001. The training was carried out over 100 epochs with a batch size of 32. The results revealed the superiority of the enhanced ResNet50 architecture over InceptionV3, achieving better accuracy.

In the field of histopathological image labeling, transfer learning has been a popular research topic. Deniz, Şengür, Kadiroğlu, Guo, Bajaj and Budak [27] conducted a study on the Break-His dataset, using Alex Net and VGG16 as feature extraction architectures [16]. The authors employed a batch size of 10 and a learning rate of 0.0001 to combine momentum optimization with stochastic gradient descent. The researchers conducted three independent experiments that combined the output of the Alex Net and VGG16 networks with support vector machine (SVM) classifiers. The findings demonstrated that the fine-tuned Alex Net architecture performed better than both the VGG16 and Alex Net networks in feature extraction, among other areas. Ahmad, Ghuffar and Khurshid [28] used the Bio-Imaging dataset to study the impact of transfer learning with three distinct CNN architectures: ResNet, Alex Net, and Google Net. The dataset was enhanced using image recognition software to increase the number of shots from 260 to 72,800. The ResNet architecture outperformed the rest of the designs with an accuracy rating of 85%.

In this research [29], an idea of using multiple CNNs for classifying the cancer pathological images of breast was presented. The application of three CNNs proved to provide a higher accuracy, precision, and F1-score in comparison to single CNN models; it also solved problems with wrong classification and time-consuming. However, there are some drawbacks when applying the model, which include the fact that the feature annotation procedure of the CNN training is one of the key processes of this model and also requires well-annotated datasets and multiple CNN trainings, which are time-consuming and computationally expensive. However, the explainability factor is still a critical issue since these models are still referred to as ‘black boxes’ in the medical fraternity.

Another recent work also centered the use of transfer learning techniques to enhance the performance of medical images analysis [30]. Through transfer learning, models are able to draw on pre-existing

networks hence minimizing the need for large amounts of labeled data as well as shortening the training process. However, the drawback of this approach is that it may lead to the discrepancy between the source domain and the target domain, which in turn impacts the model. In the same respect, fine-tuning can also bring biases in case they are not well managed during the process.

Recent years, Zero-Shot Learning (ZSL) has been noted for its ability to categorize images using auxiliary and pre-existing knowledge without minimum labeled datasets. This approach greatly alleviates a common issue in medical imaging that is, the dependence on large annotated datasets. However, its performance is still lower than the traditional supervised learning models, and the employment of ZSL in critical areas such as medicine for diagnosing is not yet fully proved [31].

The focus of this work [32] was to incorporate attention operations into CNN where it appears that the addition of an attention module greatly improves the interpretability of the classifier in addition to its predictive accuracy for medical images. To solve this problem, the attention mechanism helps the model to concentrate on specific important regions of the image helping to improve the diagnostic accuracy of the model. This approach mitigates some of the black-box issue that are characteristic of traditional CNNs since it is the model's rationale in terms of pictures. However, the method generally demands a significant amount of computation to obtain the results and does not perform well when applied to other datasets if not properly calibrated.

3. Methodology

This research aims to present the classification of histopathological images using CNN architectures fine-tuned with weights from the PatchCamelyon (PCam) dataset. We used an automated mechanism to tune the parameters such as learning rates, batch size, and network architecture to enhance the CNN model for identifying specific histopathological images. Some additional changes were performed manually in order to tune certain aspects of the model, like layer sizes and irregularities parameters, as per results and, experience knowledge in improving the model's performance. In the following diagram, we offer a general overview of the proposed method of this study. Apart from the phases depicted on the chart in Fig. 2, this study involves several phases, as shown below in the diagram.

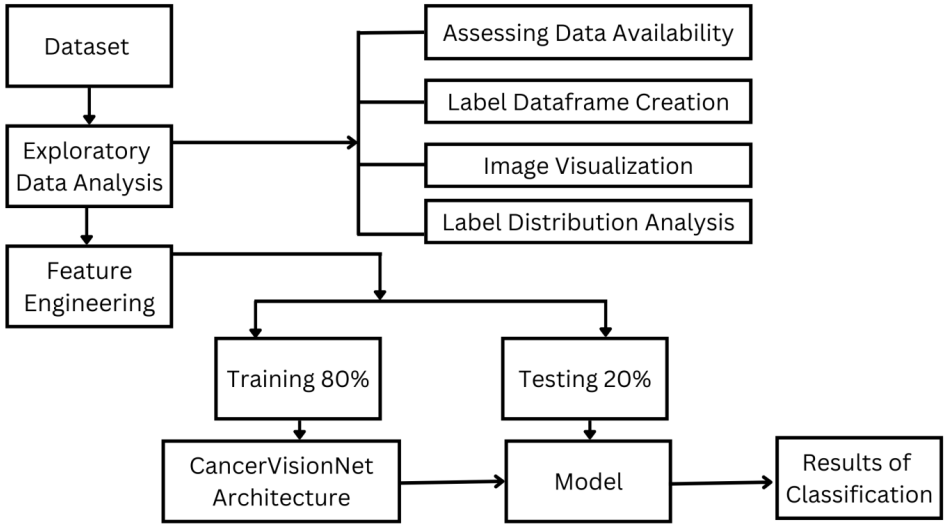

Fig. 2 Proposed model diagram

3.1 Mechanism of operation

In this study, the initial phase involved a thorough evaluation and preprocessing of the dataset to ensure its suitability for the intended objective. This included a detailed review of the images to identify potential issues such as image dimensions and quality. Methods were then applied to expand the dataset's size. Subsequently, careful selection of CNN architectures took place; thereafter, they were trained to attain optimum performance in the classification task by fine-tuning their weights using the PCam dataset. Model performance was evaluated by comparing results from different designs using area under the curve (AUC) and accuracy metrics. Finally, the results were carefully evaluated. The methods described in this document are expected to offer a valuable contribution to ongoing advances in medical image analysis, which could result in increased accuracy in histopathological image classification.

To facilitate image categorization, the proposed deep learning model systematically uses the PatchCamelyon (PCam) dataset. The initial phase involves procuring the PatchCamelyon dataset, which serves as the foundation for the subsequent training and testing stages. Next step of the process is Exploratory Data Analysis (EDA) of the given dataset and it involves Data Availability Check, Creating Label Dataframe, Image Visualization and also Label Distribution Analysis. This is useful to define the nature of a given dataset and distribution of the classes for the further processing. A subsequent step called Feature Engineering is then conducted with the aim of improving usability of the dataset for the model. This step enables preparation of the data being fed to the model so as to include the right features for the classification to be made. The dataset is then split into a training and

testing set, where the training set was taken to be 80% and the test set to be 20%. The training dataset is thereafter fed into the CancerVisionNet architecture, in which the model is trained on the images in an attempt to identify the important features and patterns. More especially, the architecture of CancerVisionNet that is proposed for histopathological image analysis and consists of layers that are specifically developed for feature extraction and classification purposes. At the same time, the testing dataset is used to check generality and accuracy of the model while applied for the newly entered data. The final results of classification are obtained after the Model trained on the CancerVisionNet architecture is assessed. This stage also proves how the model can correctly categorize cancerous tissues within histopathology images, which is helpful in medical imagery applications.

3.2 CancerVisionNet

The main contributions of the proposed CancerVisionNet model are in providing a novel evaluation and a suitable solution for histopathological image classification, which is highly challenging. Unlike the previous models CancerVisionNet is specifically designed for the medical image data and obtains excellent accuracy while avoiding overfitting by using data augmentation technique. Moreover, the model employs label distribution-based methodologies like label distribution analysis and exploratory data analysis (EDA) to ensure that metrics are optimized across a range of assessments. This research not only improves the classification accuracy, but also opens up significant possibilities for other medical image analysis tasks for future studies to further advance the standard of related research.

The convolution filter, also known as the kernel, is applied to the input images starting from the upper left corner, as illustrated in Fig. 3. This process extracts the crucial features for image classification. A CNN uses the backpropagation technique to iteratively learn the weights of the convolution filter, which play a vital role in accurately classifying the images. Several filters are used on the images to extract as many characteristics as possible.

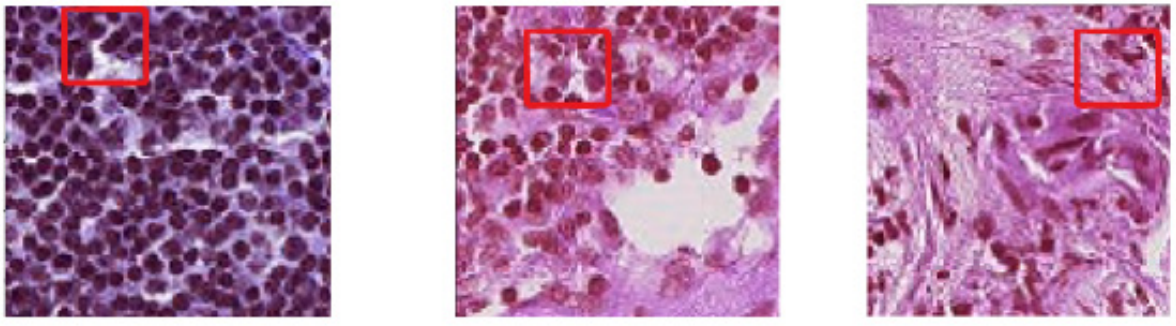

Fig. 3 Feature extraction using a convolution filter

The process of convolutional filtering generates feature maps from input images, which are subsequently utilized in further layers for precise classification. Multiple filters produce several feature maps that capture unique image characteristics [17]. Back-propagation is employed to fine-tune the CNN's weights from the last layer to the first, beginning with a non-zero distribution to prevent gradient problems. To avoid starting with a blank slate, network weights can be copied and pasted across nodes.

The proposed deep learning model, CancerVisionNet, makes use of partially and completely connected layers activated by ReLU, batch normalization, MaxPooling, and Dropout methods. When placed together, these parts allow for the extraction of characteristics from the input data, which, in turn, allows the data to be classified into different cancer groups. Fig. 4 shows the schema of the proposed model.

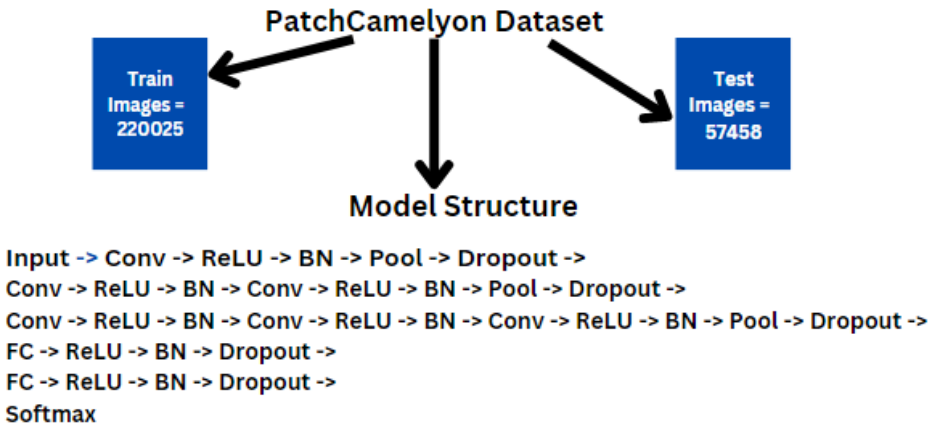

Fig. 4 A schematic diagram of the proposed model

Starting with the initial input shape and channel dimensions as reference data, the CancerVisionNet architecture is incrementally built layer by layer, merging activation functions, convolutional layers, pooling layers, and batch normalization layers. The outcome is determined using a Softmax classifier. The model is trained using histological images of lymph node sections, to attain optimal model performance.

Due to the vital importance of medical images, rigorous precautions are taken to combat overfitting. A model is considered to be over-fitted if it performs very well on training data but poorly on test data. In response, an array of regularization techniques is introduced. These methods operate by slightly elevating a model's bias or training error, thereby diminishing its variance, and consequently lowering testing error. The primary objective is to minimize errors in the testing dataset, regardless of any discrepancies observed in the training dataset.

3.2.1 Early stopping

One way to reduce the total number of training iterations is by using early stopping and mitigating the overfitting of the model on the training dataset. This method entails stopping training if the accuracy of the validation data set does not show improvement over a predetermined number of epochs. By circumventing prolonged and unfruitful epochs, early stopping conserves computing resources and serves as a preventive measure against overfitting [33]. One notable advantage of integrating early stopping lies in its ability to optimize the epoch size hyperparameter. This optimization is achieved by automatically terminating training if no enhancements in accuracy are witnessed even when a larger number of epochs is initially defined [34].

In our proposed model Early stopping was used for the 15th epoch when validation accuracy stopped at 92.5% with training accuracy reached till around 95.4%. The decreasing validation loss and stable overfitting plot of validation accuracy prove that our early stopping method worked fine, the model was renormalized before it suffered from an over-fitting so its performance upper-hand without making waste through resources.

The performance of the proposed model with early stopping is given below, showing training, and validation accuracies and losses on different epochs.

table 1: performance metrics of the model with early stopping across epochs

Epoch #	Training Accuracy	Validation Accuracy	Training Loss	Validation Loss
1	79.75%	83.20%	0.4371	0.3696
3	86.82%	85.40%	0.3086	0.3302
5	91.50%	89.00%	0.2500	0.2900
7	93.20%	90.50%	0.2000	0.2700
9	94.20%	91.20%	0.1800	0.2600
11	94.90%	91.80%	0.1600	0.2400
13	95.20%	92.30%	0.1400	0.2200
15	95.40%	92.50%	0.1200	0.2000

3.2.2 Best model

When training a model, it is crucial to monitor its performance across epochs to ensure a consistent improvement in accuracy or its maintenance at a high level. When implementing a model-saving mechanism, it is a common practice to preserve only the best-performing model based on a specific metric, often the validation accuracy [35]. In this scenario, the model is saved exclusively if the accuracy achieved in the current epoch surpasses the highest accuracy attained thus far. This ensures

that the saved model represents the best outcome obtained during the training process.

3.2.3 Dropout

The Dropout regularization technique, introduced by Deniz, Şengür, Kadiroğlu, Guo, Bajaj and Budak [27], stands as a potent tool commonly employed in ensembles and shares similarities with bagging. Throughout the training process, each neuron is assigned a probability, denoted as p, for being randomly deactivated, thus enabling the creation of multiple networks from a single one. Dropout serves as a means to counteract overfitting and bolster the resilience of a CNN. Through the usage of the data augmentation and dropout techniques, the problem of overfitting has been considerably limited. We have performed validation of the model against multiple parameters. It showed a high level of performance across the entire model, which is evidence that the model has a good level of generalization to new data points, which have not been encountered during the model's training. Furthermore, and most importantly, the validation of the model based on a large and diverse sample size increases confidence in the applicability of the model to real-world settings.

3.2.4 Image augmentation

Incorporation of image augmentation techniques has gained significant prominence in addressing overfitting concerns and enlarging the training dataset's scope [36]. Given the impracticality of training a model on every conceivable image, image enhancement offers a remedy by subjecting source images to diverse transformations, rotations, and brightness adjustments, thereby generating synthetic images. This methodology proves to be instrumental in mitigating overfitting risks and elevating model accuracy [37].

3.3 Performance measures

A variety of metrics to assess classifier performance have been used in the academic literature. Although accuracy, with its clear measure of classification success, is one of the most popular metrics used to assess classifier performance, other measures such as recall, accuracy, and F1 score are also used.

Accuracy can be defined as follows.

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \qquad (1)$$

If TP is true, then the expected positive class is indeed positive, and if TN is true, then the predicted negative class is actually negative. The process starts with a false positive (FP) and ends with a false negative (FN) when the wrong class is classified as positive instead of negative. However, it is important to note that the accuracy measure can be misleading when employed with the unbalanced datasets, constituting its inherent limitation.

4. Experiments and Results

This section presents experiments and the results obtained in this study.

4.1 Experiments

4.1.1 Dataset

An improved version of the PatchCamelyon (PCam) benchmark dataset was used for this research [38]. This collection contains 327,680 color images, all acquired from histopathology scans of lymph node sections. Each image has dimensions of 96 by 96 pixels. Each image has a binary label that indicates the presence of metastatic tissue. However, the original PCam dataset included duplicate images resulting from probabilistic sampling; nevertheless, the present collection is free from duplicates. Fig. 5 shows an example of the PatchCamelyon data set.

PCam provides a unique benchmark for machine learning models, introducing a clinically important task of detecting metastasis through binary image classification. This dataset can be efficiently trained on a single GPU, even if it is larger than CIFAR10 but smaller than ImageNet. Machine learning models trained on PCam exhibit impressive performance in tasks such as tumor detection and full-slide image diagnosis, as seen in the Camelyon16 challenges. Foundational machine learning study topics such as active learning, model uncertainty, and explainability are well suited to this dataset because it finds a happy medium between task difficulty and practicality [38].

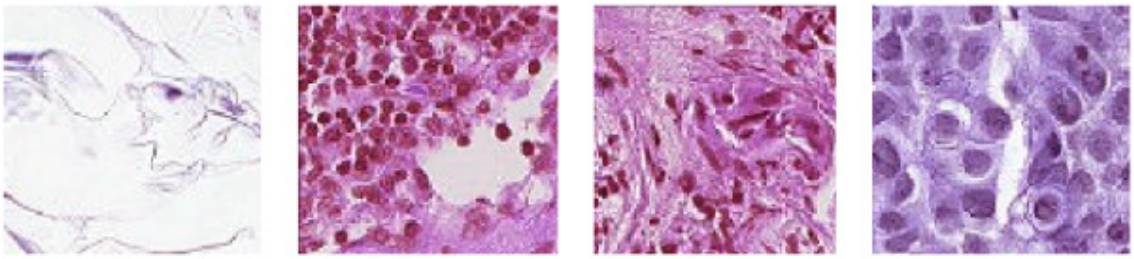

Fig. 5 Images from PatchCamelyon dataset

4.1.2 Experimental setting and requirements

In this study, the fine-tuning of CNN architecture was performed using Google Colab's tensor processing unit (TPU), which provides a suitable environment for the management of computationally resource-demanding processes. This way, one can achieve very quick cycles of experimentations and fine-tuning, and it is still possible to perform such optimizations in environments that would hardly be called compute-intensive. The use of TPU also shows that it's possible to use these models in environments that can limit resources, proving that the given model

is versatile and can be applied to various fields.

This study uses the Keras package in Python to create the (CNN) model. Specifically designed to distinguish between normal and tumorous histology images, the model is known by the moniker CancerVisionNet. Many layers that make up CancerVisionNet include activation functions, batch normalization, max-pooling, dropout, and separable convolutional layers. The network takes a histological image with predetermined width, height, and depth as input and returns a prediction of whether the image is normal or contains tumors as output.

A 32-item batch size was used for CNN architecture tuning, and the Adam optimizer was employed consistently throughout the all test iterations [39]. Visual representation of the label distribution within the dataset was performed using a count plot generated by Seaborn. Standardization was executed to uniformly resize all images to dimensions of 96 pixels on each side. The models used a 50% probability dropout layer instead of traditional fully linked layers. In addition, a new fully connected layer was added to act as a classifier, which improved the network's accuracy in general.

To mitigate the overfitting issue, which is prominent when training deep learning models with limited datasets, we used ImageDataGenerator from Keras designed to augment data in real time. This technique by default creates augmented images during the training process with the help of random rotation, shifting, flipping, zooming and other similar operations. These techniques artificially enhance the scale of diversity of the training data thus ensuring that the model does not dictate the training samples it was trained on but rather learn to minimize variations in the input data. In this regard, the model has better generalization capabilities on the data that it has not seen before.

Furthermore, by scaling the images pixel intensity to the range of 0 to 1, we made the training process more stable since problems such as different ranges of pixel intensities could easily affect the results. The model was trained using Adam's optimization algorithm, and categorical cross entropy loss function was used to control the overfitting. We also used early stopping with patience of 15 epochs. Due to this approach, we were able to track the validation loss after each epoch and ensure that the model stopped training as soon as the model could not further improve its performance, and hence mitigated overfitting. Further, only the best set of model parameters was kept as observed from cross-validation and the validation loss, which helped in minimizing overfitting of the model.

4.2 Experiment results

The PatchCamelyon dataset, which contains 220,025 annotated images, covering 60% positive and 40% negative classes, served as the training dataset. For the evaluation of the classifier, an unlabeled data set of 57,458 images was provided and the results were submitted to calculate the AUC metric.

All assessments were performed using the Keras package in Python. The schematic representation of the proposed model is shown in Fig. 2.

In this illustration, "Input" represents the data that are fed into the model, and "Softmax" represents the output layer that is in charge of producing class probabilities. Layers are abbreviated as follows: "Conv" for convolutional, "ReLU" for rectified linear unit activation, "BN" for batch normalization, "Pool" for max pooling, "Dropout" for dropout, and "FC" for completely connected.

Table 2 shows the number of epochs needed to train the neural network model using the Keras framework. During training, we minimized the training loss by adjusting the model's weights. At the conclusion of each epoch, we tested the model on a validation set to see how well it could generalize.

Table 2: Training parameters and their description

Parameter	Description
Epochs	15
Checkpoint Frequency	After Each Epoch
Learning Rate Scheduler	ReduceLROnPlateau
Checkpoint Criteria	Maximum Validation Accuracy

Two crucial callbacks were used in the training. To ensure that the best performing model was kept, ModelCheckpoint callback first stored the model with the highest validation accuracy. Secondly, in the event that the validation accuracy plateaued, the ReduceLROnPlateau callback dynamically modified the learning rate, improving the convergence of the model.

Table 3: Proposed architecture of CancerVisionNet

Layer (type)	Output shape	Parameters
conv2d_1 (Conv2D)	(None, 96, 96, 32)	2432
conv2d_2 (Conv2D)	(None, 96, 96, 32)	9248
conv2d_3 (Conv2D)	(None, 96, 96, 32)	9248
batch_normalization_1		
Batch	(None, 96, 96, 32)	128
max_pooling2d_1		
MaxPooling2	(None, 48, 48, 32)	0
dropout_1 (Dropout)	(None, 48, 48, 32)	0
conv2d_4 (Conv2D)	(None, 48, 48, 64)	18496

conv2d_5 (Conv2D)	(None, 48, 48, 64)	36928
conv2d_6 (Conv2D)	(None, 48, 48, 64)	36928
batch_normalization_2		
Batch	(None, 48, 48, 64)	256
max_pooling2d_2		
MaxPooling2	(None, 24, 24, 64)	0
dropout_2 (Dropout)	(None, 24, 24, 64)	0
conv2d_7 (Conv2D)	(None, 24, 24, 128)	73856
conv2d_8 (Conv2D)	(None, 24, 24, 128)	147584
conv2d_9 (Conv2D)	(None, 24, 24, 128)	147584
batch_normalization_3		
Batch	(None, 24, 24, 128)	512
max_pooling2d_3		
MaxPooling2	(None, 12, 12, 128)	0
dropout_3 (Dropout)	(None, 12, 12, 128)	0
flatten_1 (Flatten)	(None, 18432)	0
dense_1 (Dense)	(None, 512)	9437696
batch_normalization_4		
Batch	(None, 512)	2048
dropout_4 (Dropout)	(None, 512)	0

In Table 3 we can see the empirical information of the proposed CancerVisionNet model architecture. There are a total of 99,237,000 parameters, with 9,922,498 of them being trainable and 1,742 being non-trainable. To get the best possible value for the parameters, the training procedure updates the trainable parameters. On the other hand, non-trainable parameters do not update during the training and never optimize during the training. These are passed as input, and hence non-trainable parameters have no contribution to the process of classification or detection.

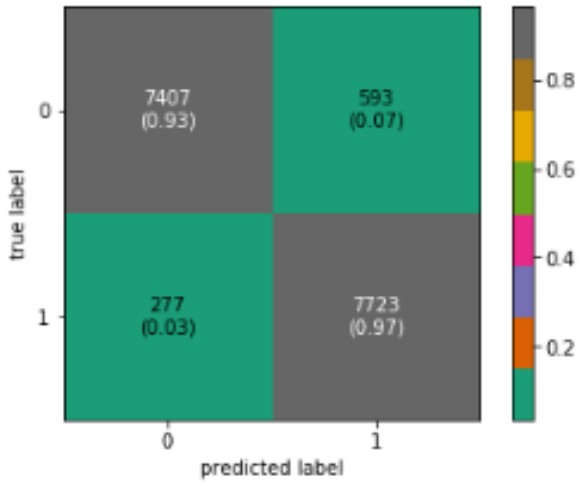

Fig. 6 Confusion matrix

Fig. 6 illustrates the confusion matrix of the proposed CancerVisionNet model. An indicator of how well a binary classification model is performing is the confusion matrix. The rows represent the real labels, while the columns show the labels predicted by the model. According to the matrix, there are two groups, denoted '0' and '1'. When it came to class '0', the model got 7407 of the cases right (true negatives) and 593 wrong (false positives). There were 2,775 cases of inaccurate predictions '0' (false negatives) and 7,723, 00 cases of valid predictions '1' for class '1'. For class '0', for instance, 7407 forecasts are true negatives and 593 are false positives, for a total of 93%. class '1' has a high true positive rate, and class '0' has a high true negative rate, both of which show that the model is very good at making predictions.

The model's performance evaluation hinges on binary accuracy, gauging the proportion of accurately classified samples.

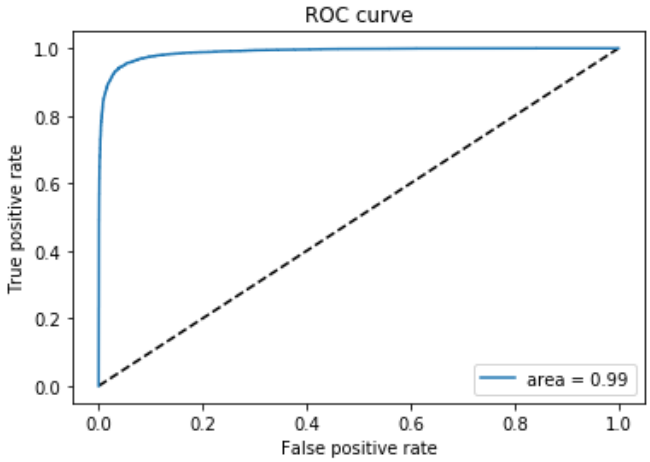

Fig. 7 ROC curve results

The model's performance across the training epochs is shown in the figure below. While the loss plot shows how training and validation loss have decreased across epochs, the accuracy plot shows how both training and validation accuracy have improved.

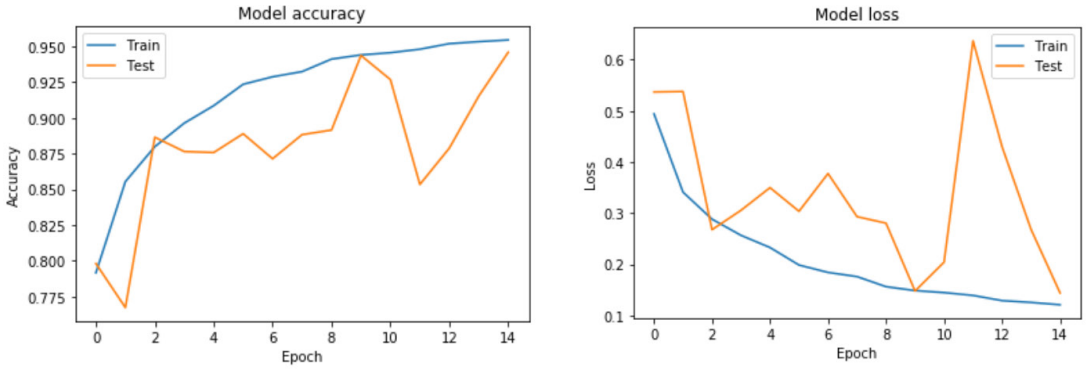

Fig. 8 Model accuracy/ loss

We obtained the AUC of the ROC curve for assessing classifier effectiveness, our conclusions are anchored in this metric. In the research conducted by Ahmad et al. [28], they attained an 85% accuracy on the BioImaging dataset through fine-tuning of a ResNet architecture. However, the details of their transfer learning methodology remain undisclosed. On the other hand, Mehra [15] utilized the VGG16 network in conjunction with logistic regression, achieving a precision of 92.6% and an AUC of 95.65%. In particular, they did not extensively explore the intricacies of each fine-tuning facet.

Table 4: Proposed Model Performance Evaluation

Metric	Tumor-Free	Tumor-Detected	Overall Avg.	Combined Avg.	Proportional Avg.
Precision	0.96	0.93	0.95	0.95	0.95
Recall	0.93	0.97	0.95	0.95	0.95
F1-Score	0.94	0.95	0.95	0.95	0.95

Table 4 presents the performance metrics of the proposed CancerVisionNet model, including Precision, Recall, and F1-Score, across two classes: Tumor-free are those patients without diagnosis of tumor, Tumor-detected are patients with the detected residual tumor. Also, there is an overall avg. (Micro Avg.), as well as class level Avg. (Macro Avg.). In addition, proportional avg. (Weighted Avg.) shows all possible results regarding the performance of the certain model.

Precision shows how many of the targeted positive outcomes are actually accurate as given by the model. Specificity for Tumor-Free images was as follows: The model had 0.96 for Normal-Detected

images, respectively; and for Tumor-Detected images it was 0.93. Recall is the assessment of the model of how well it is capable of identifying all relevant cases in the given data set. The recall for Tumor-Detected images was rather high at 0.97 and for the model accuracy in identifying the tumor cases it highlighted the model at 97. This is so because F1-Score is the harmonic mean between precision and recall that gives a balanced view of the model's performance. For Tumor-Free and Tumor-Detected classes the F1-Score is above 0.94 indicates a fairly steady performance throughout the two classes and that makes it quite reliable.

All of the metrics presented in the above table are related to the level of performance metric values, showing a very high figure with the majority of them being over 95%. This shows that CancerVisionNet model excelled other models in each of the evaluation metrics, and thus could be useful in histopathological image classification.

5. Discussion

The purpose of this study was to examine the potential consequences of the proposed CancerVisionNet model using PatchCamelyon datasets. Three dense layers, two max-pooling levels, two dropout layers, and six separable convolutional layers make up the 13 layers of our proposed model. The proposed deep learning model used Adam, a popular optimization method with tunable learning rates, as its optimizer. The learning rate corresponds to the number of steps used to update the model's weights during training, which is 0.0001. Because it is often employed for binary classification issues, binary cross entropy, which assesses the difference between the actual and predicted labels, is selected as the loss function.

The model training spans 15 epochs, which means that the complete training dataset passed through the model is repeated. A learning rate scheduler named ReduceLROnPlateau is harnessed to augment learning. When the validation accuracy stays the same for several epochs, this scheduler increases the learning rate by 0.5 to promote better convergence. A ModelCheckpoint mechanism is implemented to store the weights of the most optimal epoch based on validation accuracy, ensuring the best performing model that is preserved for the evaluation.

When comparing our results with those of other researchers, challenges arose due to variations in evaluation criteria and data sets. For instance, Kassani, Kassani, Wesolowski, Schneider and Deters [17] attained a 94.64% accuracy on the PatchCamelyon dataset; however, they did not provide specific details about their employment of transfer learning, network fine-tuning, test dataset size, or using the dataset without labels. In the testing data, our method achieved a remarkable accuracy of 95.4%, which allowed us to reach first place among all participants. A comprehensive overview of the best performing methods is presented in Table 3, revealing that our methodology exceeded other

approaches by a substantial margin of up to 7.5%. This achievement is a source of satisfaction, underscoring the success of our research and reinforcing the effectiveness of our proposed approach. Our strategy underscores the feasibility of attaining state-of-the-art accuracy in cancer detection, even when working with relatively modest data sizes. The invaluable contribution of PatchCamelyon, a benchmark dataset tailored for histopathology image analysis, was crucial to achieving this milestone. By adeptly harnessing the PatchCamelyon dataset, our approach adeptly captures pivotal features for cancer detection, culminating in remarkable accuracy in identifying cancerous regions.

The proposed model entailed the utilization of deep learning architecture, specifically CNNs, to classify histopathology images into cancerous or noncancerous categories. To achieve the maximum possible accuracy, the proposed model architecture was carefully designed and trained using the PatchCamelyon dataset. Due to the large quantity and variety of histopathological images, the categorization process was complex. However, the effectiveness of the method in identifying malignant areas demonstrated its resilience and demonstrated the promise of deep learning in medical

Table 5: Comparison of accuracy with previous approaches/methods

Research study authors	Dataset	Accuracy	Position
Kassani, Kassani, Wesolowski, Schneider and Deters [17]	PatchCamelyon	94.64%	2nd
Ahmad, Ahmed [40]	PatchCamelyon	94%	3rd
Our study	PatchCamelyon	95.4%	1st

In the realm of medical image analysis, accurate cancer detection is of paramount importance for prompt treatment and effective patient care. Histopathological cancer detection image classification challenge sought to foster the development of robust and efficient techniques for identifying cancer in histopathological images. Even though this paper focuses on the technical aspect of the CancerVisionNet model, the integration feature is the most significant for its adoption to clinical practices. The higher accuracy level of the proposed model for histopathological image classification means that impact on diagnostic efficiency would be positive, giving fast and accurate results for clinical practice. This may help improve the quality of the outcomes of patients by allowing physicians to detect malignant tumors at an early stage and in some cases making more specific treatment plans in cancer treatment.

When it comes to medical image analysis, there is a noticeable lack of high-quality data with which to train convolutional neural networks (CNNs). Thus, it is very advantageous to use transfer learning, which makes use of pre-trained CNN weights from a big dataset. Retraining a convolutional neural network (CNN) on a specific data set is called "fine tuning." Although adjusting the whole CNN may improve performance, it is time-consuming and not always successful. For datasets featuring natural images, it is observed that the lower layers of the CNN tend to grasp more general features such as edges and circles, which are common across various image datasets. In contrast, the upper layers capture the highly distinctive attributes of the original dataset [13].

We understand that cross-domain validation is important in an attempt to validate the proposed CancerVisionNet model across different types of histopathological images. It is important as it helps validate the effectiveness of the given model when implemented in different environments other than those used in model training. That notwithstanding, for reasons of the current dataset limitations, this validation was not done in the current study. Subsequent works will also include cross-domain validation in an attempt to enhance the testing of the CancerVisionNet model on other datasets and various fields of medical imaging.

Other researchers have developed automated breast cancer diagnostic methods using the Break-His dataset. Spanhol, Oliveira, Petitjean and Heutte [41] fashioned a computer-aided diagnosis system using the dataset, relying on handcrafted feature extraction to achieve 84.60% accuracy. Concerning the Break-His dataset, Deniz et al. [27] compared the efficacy of fine-tuning and feature extraction, highlighting the superior performance of fine-tuned Alex Net compared to VGG16 when using SVM as a feature extractor. Inspired by these findings, our aim was to investigate how fine-tuning could further improve CNN performance. This has been achieved in this study by employing the proposed model and showing better performance compared to the state of the art approaches.

6. Conclusions

In this study, our focus was on the intricacies of CNN tuning efforts by proposing the CancerVisionNet model. We also assessed the impact of learning on performance using the Adam optimizer with a learning rate of 0.0001. When fine-tuning a network, it is advisable to employ a gradual learning rate to preserve initial weights. For scenarios with limited images, fine-tuning a pre-trained CNN is recommended. Our research underscores that fine-tuning CNNs with advanced architectures can significantly enhance the accuracy of medical image classification, especially for histopathology datasets. Our method of extracting image features using the PatchCamelyon dataset has outpaced other machine learning models in accuracy, offering an efficient solution for medical

image analysis. These tuning findings are context-specific, and broader applicability would require testing across various domains.

Future endeavors could focus on refining CNN accuracy for histopathological image analysis. Exploring different CNN architectures and comparing them with those used here could be a fruitful avenue. Investigating the effects of various fine-tuning strategies on CNN performance also holds promise. Expanding the training and testing dataset size could validate the efficacy of the proposed method on a larger scale. We also realized that even as scalability to larger data sets is pinpointed as future work, it was not pursued in this study because the main purpose was to test the newly developed CancerVisionNet on the available dataset. Subsequent research will focus on extending this work with larger and more variegated datasets allowing for the addition of numerous features that are to increase the model's stability and versatility. Furthermore, applying this method to other medical imaging datasets could assess its effectiveness in different medical contexts. Our research presents a promising route to enhance CNN performance in medical image analysis, with ample potential for further progress.

The performance of the model has been tested and is in high accuracy and robustness on the current dataset and various accuracy measurements. It is true that real-time clinical settings and larger datasets can add more variables, the architecture of CancerVisionNet can accommodate this increase in variables accordingly. Thus, its highly modular structure can easily scale it to work with more extended datasets and it is likely that its efficiency can be tuned for real time use. The future work will concern these aspects in the attempt to determine whether or not this model can be applied in other clinical settings.

Acknowledgments

Funding: No funding was received for this research work.

Code. Data and Material Availability

Code, data and material can be provided by the corresponding author on request.

References

1 Gurcan, M.N., Boucheron, L.E., Can, A., Madabhushi, A., Rajpoot, N.M., and Yener, B.: 'Histopathological image analysis: A review', IEEE reviews in biomedical engineering, 2009, 2, pp. 147-171

2 Metter, D.M., Colgan, T.J., Leung, S.T., Timmons, C.F., and Park, J.Y.: 'Trends in the US and Canadian pathologist workforces from 2007 to 2017', JAMA network open, 2019, 2, (5), pp. e194337-e194337

3 Yaqoob, A., Rehman, F., Sharif, H., Mahmood, M.H., Sharif, S., Ahmad, A., Ali, C.N., Hussain, A., and Khan, M.: 'Skip Connections' Importance in Biomedical Image Segmentation', in Editor (Ed.)^(Eds.): 'Book Skip Connections' Importance in Biomedical Image Segmentation' (IEEE, 2023, edn.), pp. 1-5

4 Mohammadian, S., Karsaz, A., and Roshan, Y.M.: 'Comparative study of fine-tuning of pre-trained convolutional neural networks for diabetic retinopathy screening', in Editor (Ed.)^(Eds.): 'Book Comparative study of fine-tuning of pre-trained convolutional neural networks for diabetic retinopathy screening' (IEEE, 2017, edn.), pp. 1-6

5 Khan, N.M., Abraham, N., and Hon, M.: 'Transfer learning with intelligent training data selection for prediction of Alzheimer's disease', IEEE Access, 2019, 7, pp. 72726-72735

6 Hosny, K.M., Kassem, M.A., and Foaud, M.M.: 'Classification of skin lesions using transfer learning and augmentation with Alex-net', PLoS One, 2019, 14, (5), pp. e0217293

7 Harangi, B.: 'Skin lesion classification with ensembles of deep convolutional neural networks', Journal of biomedical informatics, 2018, 86, pp. 25-32

8 Fukushima, K.: 'Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position', Biol. Cybern., 1980, 36, (4), pp. 193-202

9 Krizhevsky, A., Sutskever, I., and Hinton, G.E.: 'Imagenet classification with deep convolutional neural networks', Advances in neural information processing systems, 2012, 25

10 Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., and Bernstein, M.: 'Imagenet large scale visual recognition challenge', International journal of computer vision, 2015, 115, pp. 211-252

11 Gupta, R.K., Kaur, M., and Manhas, J.: 'Tissue level based deep learning framework for early detection of dysplasia in oral squamous epithelium', Journal of Multimedia Information System, 2019, 6, (2), pp. 81-86

12 Chollet, F.: 'Deep learning with Python' (Simon and Schuster, 2021. 2021)

13 Yosinski, J., Clune, J., Bengio, Y., and Lipson, H.: 'How transferable are features in deep neural networks?', Advances in neural information processing systems, 2014, 27

14 Kalbhor, M.M., and Shinde, S.V.: 'Cervical cancer diagnosis using convolution neural network: feature learning and transfer learning approaches', Soft Computing, 2023, pp. 1-11

15 Mehra, R.: 'Breast cancer histology images classification: Training from scratch or transfer learning?', ICT Express, 2018, 4, (4), pp. 247-254

16 Spanhol, F.A., Oliveira, L.S., Petitjean, C., and Heutte, L.: 'A dataset for breast cancer histopathological image classification', Ieee transactions on biomedical engineering, 2015, 63, (7),

pp. 1455-1462

17 Kassani, S.H., Kassani, P.H., Wesolowski, M.J., Schneider, K.A., and Deters, R.: 'Classification of histopathological biopsy images using ensemble of deep learning networks', arXiv preprint arXiv:1909.11870, 2019

18 Veeling, B.S., Linmans, J., Winkens, J., Cohen, T., and Welling, M.: 'Rotation equivariant CNNs for digital pathology', in Editor (Ed.)^(Eds.): 'Book Rotation equivariant CNNs for digital pathology' (Springer, 2018, edn.), pp. 210-218

19 Bejnordi, B.E., Veta, M., Van Diest, P.J., Van Ginneken, B., Karssemeijer, N., Litjens, G., Van Der Laak, J.A., Hermsen, M., Manson, Q.F., and Balkenhol, M.: 'Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer', Jama, 2017, 318, (22), pp. 2199-2210

20 Moore, J., Linkert, M., Blackburn, C., Carroll, M., Ferguson, R.K., Flynn, H., Gillen, K., Leigh, R., Li, S., and Lindner, D.: 'OMERO and Bio-Formats 5: flexible access to large bioimaging datasets at scale', in Editor (Ed.)^(Eds.): 'Book OMERO and Bio-Formats 5: flexible access to large bioimaging datasets at scale' (SPIE, 2015, edn.), pp. 37-42

21 Simonyan, K., and Zisserman, A.: 'Very deep convolutional networks for large-scale image recognition', arXiv preprint arXiv:1409.1556, 2014

22 Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H.: 'Mobilenets: Efficient convolutional neural networks for mobile vision applications', arXiv preprint arXiv:1704.04861, 2017

23 Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K.Q.: 'Densely connected convolutional networks', in Editor (Ed.)^(Eds.): 'Book Densely connected convolutional networks' (2017, edn.), pp. 4700-4708

24 Vesal, S., Ravikumar, N., Davari, A., Ellmann, S., and Maier, A.: 'Classification of breast cancer histology images using transfer learning', in Editor (Ed.)^(Eds.): 'Book Classification of breast cancer histology images using transfer learning' (Springer, 2018, edn.), pp. 812-819

25 Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z.: 'Rethinking the inception architecture for computer vision', in Editor (Ed.)^(Eds.): 'Book Rethinking the inception architecture for computer vision' (2016, edn.), pp. 2818-2826

26 He, K., Zhang, X., Ren, S., and Sun, J.: 'Deep residual learning for image recognition', in Editor (Ed.)^(Eds.): 'Book Deep residual learning for image recognition' (2016, edn.), pp. 770-778

27 Deniz, E., Şengür, A., Kadiroğlu, Z., Guo, Y., Bajaj, V., and Budak, Ü.: 'Transfer learning based histopathological image classification for breast cancer detection', Health information science

and systems, 2018, 6, pp. 1-7

28 Ahmad, H.M., Ghuffar, S., and Khurshid, K.: 'Classification of breast cancer histology images using transfer learning', in Editor (Ed.)^(Eds.): 'Book Classification of breast cancer histology images using transfer learning' (IEEE, 2019, edn.), pp. 328-332

29 Rafiq, A., Chursin, A., Awad Alrefaei, W., Rashed Alsenani, T., Aldehim, G., Abdel Samee, N., and Menzli, L.J.: 'Detection and classification of histopathological breast images using a fusion of CNN frameworks', Diagnostics, 2023, 13, (10), pp. 1700

30 Kora, P., Ooi, C.P., Faust, O., Raghavendra, U., Gudigar, A., Chan, W.Y., Meenakshi, K., Swaraja, K., Plawiak, P., and Acharya, U.R.: 'Transfer learning techniques for medical image analysis: A review', Biocybernetics and Biomedical Engineering, 2022, 42, (1), pp. 79-107

31 Woerner, S., and Baumgartner, C.F.: 'Navigating Data Scarcity using Foundation Models: A Benchmark of Few-Shot and Zero-Shot Learning Approaches in Medical Imaging', arXiv preprint arXiv:2408.08058, 2024

32 Li, X., Li, M., Yan, P., Li, G., Jiang, Y., Luo, H., and Yin, S.: 'Deep learning attention mechanism in medical image analysis: Basics and beyonds', International Journal of Network Dynamics and Intelligence, 2023, pp. 93-116

33 Farhadipour, A.: 'Lung and colon cancer detection with convolutional neural networks and adaptive histogram equalization', Iran Journal of Computer Science, 2023, pp. 1-15

34 Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L.: 'Imagenet: A large-scale hierarchical image database', in Editor (Ed.)^(Eds.): 'Book Imagenet: A large-scale hierarchical image database' (Ieee, 2009, edn.), pp. 248-255

35 Kassani, S.H., Kassani, P.H., Wesolowski, M.J., Schneider, K.A., and Deters, R.: 'Deep transfer learning based model for colorectal cancer histopathology segmentation: A comparative study of deep pre-trained models', International Journal of Medical Informatics, 2022, 159, pp. 104669

36 Kumar, N., Sharma, M., Singh, V.P., Madan, C., and Mehandia, S.: 'An empirical study of handcrafted and dense feature extraction techniques for lung and colon cancer classification from histopathological images', Biomedical Signal Processing and Control, 2022, 75, pp. 103596

37 Şengöz, N., Yiğit, T., Özmen, Ö., and Isik, A.H.: 'Importance of preprocessing in histopathology image classification using deep convolutional neural network', Advances in Artificial Intelligence Research, 2022, 2, (1), pp. 1-6

38 Zhong, Z., Zheng, M., Mai, H., Zhao, J., and Liu, X.: 'Cancer image classification based on DenseNet model', in Editor (Ed.)^(Eds.): 'Book Cancer image classification based on DenseNet

model' (IOP Publishing, 2020, edn.), pp. 012143

39 Kingma, D.P., and Ba, J.: 'Adam: A method for stochastic optimization', arXiv preprint arXiv:1412.6980, 2014

40 Ahmad, M., Ahmed, I., Ouameur, M.A., and Jeon, G.: 'Classification and detection of cancer in histopathologic scans of lymph node sections using convolutional neural network', Neural Processing Letters, 2023, 55, (4), pp. 3763-3778

41 Spanhol, F.A., Oliveira, L.S., Petitjean, C., and Heutte, L.: 'Breast cancer histopathological image classification using convolutional neural networks', in Editor (Ed.)^(Eds.): 'Book Breast cancer histopathological image classification using convolutional neural networks' (IEEE, 2016, edn.), pp. 2560-2567