Research on Biomedical Engineering
https://doi.org/10.1007/s42600-024-00375-2

ORIGINAL ARTICLE

A robust transfer learning approach for colorectal cancer identification based on histopathology images

Toto Haryanto[1] · Helmi Al Farel[2] · Heru Suhartanto[2] · Kusmardi Kusmardi[3] · Marina Yusoff[4,5] · Jasni Mohamad Zain[4,5] · Ari Wibisono[2]

Received: 1 August 2023 / Accepted: 30 July 2024

Abstract

Objective Diagnosis of cancer at the benign stage is crucial. Recently, pathologists have been using computer-aided diagnostics with machine learning to diagnose patients from medical images. The limitation of the medical image dataset is a challenge to obtain a robust model for cancer identification. This research was conducted to address this issue. The weight of the pre-trained model can be learned more efficiently with a limited amount of data. However, the training process is conducted by customizing the freeze rate so that feature extraction is preserved besides tuning some hyperparameters.

Methods Transfer learning is a technique that can handle data limitation issues in the medical field. Transfer learning will reuse pre-trained model for different or specific task. Choosing an optimum architecture and hyperparameters in machine learning is very important to improve model performance. In our experiment, we carried out a hyperparameter optimization of various deep learning architectures that classifies images containing healthy and cancer tissue.

Result The research concludes that CNN with architecture DenseNet121, freeze rate 75%, zero hidden layers on the classifier, learning rate 0.001, and optimizer RMSProp have the best performance with 98% accuracy and 19.5 s training time using NVIDIA A100 GPU accelerator. Testing with the real dataset for future direction will be an achievement for the model's success.

Conclusion This research has successfully optimized the Densnet121 deep learning architecture by tuning parameters. With a harmonic means value of 0.98, DenseNet121 outperforms the others. Compared to Camelyon, ImageNet still becomes the baseline of the transfer learning dataset because it has a rich amount of data.

Keywords Colorectal cancer · Diagnostic accuracy · Deep learning · Histopathology image · Transfer learning

✉ Toto Haryanto
totoharyanto@apps.ipb.ac.id

Helmi Al Farel
helmi.alfarel@ui.ac.id

Heru Suhartanto
heru@cs.ui.ac.id

Kusmardi Kusmardi
kusmardi.ms@ui.ac.id

Marina Yusoff
marina998@uitm.edu.my

Jasni Mohamad Zain
jasni67@uitm.edu.my

Ari Wibisono
ari.w@cs.ui.ac.id

[1] Department of Computer Science, IPB University, Bogor, Indonesia 16680

[2] Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia 16424

[3] Department of Pathology Anatomy, Faculty of Medicine, University of Indonesia, Depok, Indonesia 10430

[4] Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Kompleks Al-Khawarizmi, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor Darul Ehsan, Malaysia

[5] College of Computing, Informatic and Media, Kompleks Al-Khawarizmi, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor Darul Ehsan, Malaysia

Published online: 07 September 2024

Introduction

According to research by the World Health Organization, cancer is one of the leading causes of death in the world (WHO IARC 2018). Based on the data collected, it is estimated that there will be 19.3 million new cancer cases and 10 million deaths due to cancer in 2020 (Sung et al. 2021). Cancer is a condition in which cells grow uncontrollably and spread to other organs. The uncontrolled multiplication of cells can form tumors. Tumors can be benign or malignant. Malignant tumors can spread to other tissues through the lymph node system to form new tumors. The most important thing in cancer treatment is detecting it and treating it at a benign stage.

The test methods used to detect cancer, especially colorectal cancer, such as colonoscopy and flexible sigmoidoscopy (Lin et al. 2021). A pathologist performs this test method by analyzing colorectal tissue in situ using a tool. With the development of computer technology, pathologists are using tools such as computed tomography to obtain detailed tissue in the form of medical images that can be further analyzed. Currently, there are many studies that analyze medical images using a computer called computer-aided diagnosis (CAD) (Halalli and Makandar 2018). Computer-aided diagnosis (CAD) with medical images is increasingly used to assist pathologists in the process of detecting, diagnosing, and prognosing disease. Most of the use of CAD is to analyze cytology images because cytology images can be obtained by minimally invasive surgery (Gurcan et al. 2009). However, histopathological images provide a more comprehensive picture of the disease and its effects on tissues, because the image collection process preserves tissue structure. Thus, histopathological image analysis is the gold standard in diagnosing diseases such as cancer (Histopathology is ripe for automation. 2017).

Machine learning (ML) is a technology that has been used to help pathologists detect various types of cancer. For example, Zotin et al. (2019) extracted various types of structures from X-ray images and used a neural network to detect lung cancer. (Haryanto et al. 2017) used a convolutional neural network (CNN) to identify colon cancer on histopathological gland images. Machine learning is a branch of artificial intelligence that allows computer systems to learn from examples, data, and experiences. The challenges and opportunities of machine learning for digital pathology image analysis are comprehensively explained (Madabhushi and Lee 2016). Machine learning can be used to solve quite complex problems, such as creating a model that can accurately detect cancer in medical images. Machine learning requires a lot of data to solve complex problems, whereas medical images are generally available in limited quantities. To overcome this issue, Haryanto et al. (2021) propose the conditional sliding window (CSW) technique to reproduce data by taking sub-samples from the images in the dataset. The data augmentation method is also one way of reproducing data by giving a transformation to an image, for example, a geometric transformation, such as rotation, or a color transformation, such as adding brightness. In the article by Haryanto et al., they built a colorectal cancer detection system by training a Convolutional neural network using datasets processed using Conditional Sliding Windows (CSW) and augmentation. The technique proposed the resulting model with an accuracy of 81%.

Colorectal cancer, a type of cancer that has a high mortality rate and prevalence in the world, is still an interesting research area to study using digital imaging and deep learning approaches. Combining the ranking algorithm and CNN has been proposed for detecting colon cancer location and classification with performance above 90% in terms of accuracy, recall, and precision (Karthikeyan et al. 2024) (Yao et al. 2024). Deep learning helps radiologists detect colorectal cancer with contrast-enhanced CT images. The construction prognostic model for stadium III colorectal cancer was developed using VGG19 transfer learning on Whole Slide Images (WSI) (Zhou et al. 2024). This research achieved 94.4% accuracy for nine types of tissue. Colorectal cancer cell detection has been successfully carried out using YOLO and a variation of Residual Network (ResNet) as a hybrid architecture producing 98.70% accuracy (Haq et al. 2024).

Currently, deep learning is not only a technique for solving classification problems but also can perform feature extraction through a convolution mechanism. In terms of classification, RNN and CNN are proven to produce better performance compared to several other classifiers (Ershadi and Seifi 2022). DeepSVM is a novel proposed method as a hybrid framework between deep learning and support vector machine (SVM). Various deep learning architectures are applied as a technique to extract features, which are then combined with SVM to classify histological images of colorectal cancer (Khazaee Fadafen and Rezaee 2023).

Transfer learning is one technique that is currently being researched to address the issue of limited amounts of data. Transfer learning is a method in which the knowledge gained during the learning process to solve a problem is used to solve different problems. Transfer learning techniques have been successfully applied to create models that can detect breast cancer, cervical cancer (Li et al. 2019), skin cancer (Kalaiarasan et al. 2022), and lung cancer (Humayun et al. 2022) (Ren et al. 2022) that exceed the performance of models trained from scratch. The performance of deep learning can be improved by optimizing the hyperparameter values. There are some research related

to the overview of transfer learning, such as Weiss et al. (2016), Panigrahi et al. (2021), and a comprehensive literature review (Zhuang et al. 2021). Hyperparameters in transfer learning are variables that control the learning and training process of the model. There are several hyperparameters that, if they have the right values, can improve performance both in terms of accuracy and the required training time. Each hyperparameter has a large selection of values, so obtaining a set of hyperparameter values that produce the model with the best performance is necessary. Our research motivation is how to find the best hyperparameter among various pre-trained models. This research will look for combinations of hyperparameter values that produce the best performance in terms of accuracy and training time with transfer learning techniques and analyze the effect of these hyperparameters. The contributions of our research are as follows:

- This research proposed a robust model based on the transfer learning approach for colorectal cancer classification, DenseNet121.
- From 945 model candidates, DenseNet121 with a 75% freeze rate, 0.001 learning rate, and RMSProp is the best combination of the model.

Related works

This section describes research related to histopathology image-based cancer identification using a deep transfer learning approach. Several comparative studies of the transfer learning approach used to identify colorectal cancer are presented (Hosseinzadeh Kassani et al. 2022). More than 15 methods are reviewed in this article. From these methods, it is revealed that InceptionResNetV2 is superior to the others. Deep learning has open problems that allow researchers to carry out optimizations to improve the performance of the proposed model. Galactic Swarm Optimization (GSO) is proposed as a hyperparameter tuning to improve model performance (Escorcia-Gutierrez et al. 2022). The screening and analysis process for metastatic colorectal cancer tissue is necessary to prevent spreadable (Khan et al. 2023). The transfer learning capability to develop a model on the MNIST colorectal histology data set has been successfully implemented. This transfer learning has been successfully used for both feature extraction and classification processes (Reis and Turk 2023). Several related studies are presented in a summary table to show the potential of our study (Table 1).

According to Table 1, the use of transfer learning to identify colorectal cancer is quite promising for future research. Several open problems in transfer learning have the opportunity to obtain a robust model.

Material and methods

This section describes highlights of the process used to classify colorectal cancer. The first part describes datasets and augmentation, and then transfer learning models and hyperparameter tuning to produce a robust model. The experimental setup section also describes the environment during training for both software and hardware. In this section, we also explain the formulation used to assess the performance of the model.

Design of robust transfer learning for colorectal cancer identification

The development of the transfer learning model requires a combination of several hyperparameters and a base model as an initialization. In our study, there were 945 candidate models after combining all of experiment scenarios which will be explained in detail in Table 5. In general, the flow of developing the transfer learning model is shown in Fig. 1.

Dataset collection

The dataset used in this study was provided by Warwick University in the 2015 Gland Segmentation Challenge Contest (GlaS) (Sirinukunwattana et al. 2015, 2016). The Warwick dataset consists of 165 histopathological images taken from 16 colorectal tissues stained with hematoxylin and eosin (H&E). There are 74 images annotated with benign cancer (Benign) and 91 images annotated with malignant cancer (Malignant). In the GlaS'2015 event, the dataset was divided into three parts: Training, Test Part A, and Test Part B. This division will be explained further later. Most of the images have a size of 755×522 pixels, but some have different sizes. Each colorectal tissue used comes from another patient and the tissue is processed at different occasions. As a result, this dataset has a high level of variation both from the aspect of staining distribution and network architecture. The details of the benign and malignant dataset for colorectal histopathology images are shown in Table 2.

Usually, when taking histopathological images, lymph nodes (lymph glands) are taken from the patient's body and placed on a glass slide for analysis by a pathologist (Bejnordi et al. 2017). The tissue on the glass slide is scanned using a high-speed whole slide scanner. This process is called Whole Slide imaging and results obtained are called whole slide images (WSI). Digitized WSI provides many conveniences for research and education needs (Pantanowitz et al. 2012). Shifting from conventional glass to WSI can support cancer analysis via virtual microscopy (Idlahcen et al. 2020). Prior to scanning, the glass slide will be stained with

Table 1 Related study on colorectal cancer identification using deep learning

Study	Dataset	Technique	Performance	Contribution
Haq et al. (2024)	Colorectal cancer Procell Life Science & Technology Co., Ltd. (Wuhan, China)	YOLOv2 ResNet-18 ResNet-50	Accuracy: 98.70% (ResNet-18) 96.66% (ResNet-50)	Integrating YOLOv2, ResNet-50, and ResNet-18 to detect colorectal cell cancer detection and counting
Nemlander et al. (2023)	Colorectal cancer Real dataset from patients	SGB Model	Sensitivity: 73.3% Specificity: 83.5% ROC curve: 83.2%	Developing a predictive model for identifying non-metastatic CRC (NMCRC)
Khan et al. (2023)	Colorectal cancer LnSegment PatchCamelyon PatchCRC	Xception Vision Transformer	Adenocarcinoma cases sensitivity: 0.99 Specificity: 0.96 Mucinous, signet-ring cell sensitivity: 0.872 Specificity: 0.936	This research proposed complete deep learning (segmentation using UNet and classification by Xcepetion and ViT
Reis and Turk (2023)	Colorectal histology MNIST dataset	**DenseNet169** + SVM **DenseNet169** + GRU	Random weight accuracy: 94.29% Transfer learning accuracy: 95%	Propose hybrid classification (deep learning for feature extraction and classical machine for classification)
Hosseinzadeh Kassani et al. (2022)	Colorectal cancer DigestPath dataset	**DenseNet** LinkNet	Similarity index: 82.74% Accuracy: 87.07% f1-score: 82.79%	Combining DenseNet and LinkNet to develop the best performance model
Escorcia-Gutierrez et al. (2022)	Colorectal cancer Not available/not mentioned	GSODTL-C3M	Accuracy: 99.39% Sensitivity: 99.47% Specificity: 99.47%	Optimizing model called Galactic Swarm Optimization with Deep Transfer Learning Driven Colorectal Cancer Classification
Humayun et al. (2022)	Lung cancer	Transfer Learning VGG-16, VGG-19, and Xception	Accuracy VGG-16: 98.83% VGG-19: 98.05% Xception: 97.40%	This study proposed a low number parameter of model
Ren et al. (2022)	Lung cancer SPECT/CT from Hong Kong Queen Mary Hospital and Henan Cancer Hospital	**Transfer learning**	Average dice similarity coefficient (DSC): 0.8112 ± 0.0484	A transfer learning framework was developed to utilize the feature extraction
Haryanto et al. (2021)	Colorectal cancer Warwick dataset	Conditional Sliding Windows (CSW) CNN-7–5-7	Sensitivity: 0.820 Specificity: 0.815	Contribute in designing the method Conditional Sliding Window (CSW) to generate dataset and classify images via the proposed CNN architecture
Li et al. (2020)	Breast cancer BreakHisdataset	**DenseNet** VGG ResNet	Best accuracy (DenseNet)	The study proposed a framework that can produce improved accuracy
Kudva et al. (2020)	Cervical cancer	**Hybrid Transfer Learning**	Accuracy: 93.83% Sensitivity: 91.46% Specificity: 89.16%	Hybrid transfer learning with a simple number of parameters compared to existing architectures
Qaiser et al. (2019)	Colorectal cancer Warwick UHCW-dataset	Persistent Homology Profile (PHP) + CNN	Recall: 0.922 Precision: 0.926 f1_score: 0.924	The combination of PHP and CNN produces a fairly competitive feature extraction technique

Table 1 (continued)

Study	Dataset	Technique	Performance	Contribution
Raza et al. (2019)	Colorectal cancer Warwick dataset	Micro-Net	Dice: 0.857	Micro-Net is a proposed deep learning network architecture that has the ability to segment histopathological images of colon cancer
Nguyen et al. (2018)	PAP-smear 2D-Hela	Inception-v3 ResNet-152 Inception-ResNet-v2	Accuracy PAP-smear: 92.63% 2D-Hela: 92.57%	Transfer Learning and feature concatenation
Chen et al. (2017)	Colorectal cancer Warwick dataset	DCAN fully convolutional network (FCN)		This study proposes the role of contour information in histopathological images of colon cancer for the segmentation process, so it is known as Deep Contour Aware Network (DCAN)

hematoxylin and eosin (H&E). Hematoxylin and eosin give color to tissues so that cell structures such as cell nuclei are easier to see. As can be seen in Fig. 2, the results of the staining on a glass slide give the tissue a pink-purple color. There are variations in color and size for each WSI collection due to differences in lab conditions, scan tools, staining protocols, or tissue types (Inoue and Yagi 2020).

The availability of medical image datasets is a challenge. Therefore, using several techniques to generate datasets is an alternative. On the other hand, for more detailed analysis, the histopathological image must be enlarged to a magnification level of up to 400×. Large datasets are needed to represent the population, so the system can generalize. The Warwick dataset has a very limited number, so it is necessary to generate the data. The data generation process is achieved using the CSW technique to enrich the dataset as the first preprocessing step. This technique produces data by subsampling from the original data, so that the amount of data increases. Subsampling is done by cropping an area of the image and then cropping the next area. The size of the cropped area (window size) is set first. Haryanto et al. experimented with several different window sizes and different overlap schemes. Previous research found that a window size of 300×300 pixels with an overlap of 50 pixels produces the best performance. Overlap is considered to reduce the loss of information. This study will use the same scheme, namely window 26 size 300×300 with an overlap of 50 pixels. Table 3 shows the amount of data before and after the data generation process using CSW.

Data augmentation

The data augmentation process is carried out to overcome the limited data. Data augmentation is a technique for increasing the number and quality of datasets (Shorten and Khoshgoftaar 2019). Some examples of image data augmentation are geometric transformations, color transformations, random erasing, adversarial training, and neural style transfers. In this study, geometric transformation and color transformation methods are used to augment the data. The geometric transformation used follows previous research. At the same time, the color transformation used follows (Liu et al. 2017), who won the CAMELYON'16 Challenge, which is a competition for making a breast cancer detection system. Details regarding the augmentation techniques used and their value ranges are presented in Table 4. We visualize the result of augmentation techniques in Fig. 3.

Transfer learning approach

Transfer learning is the process of using learning that has already been done to solve a problem to solve a different problem. The learning outcomes (optimized parameters)

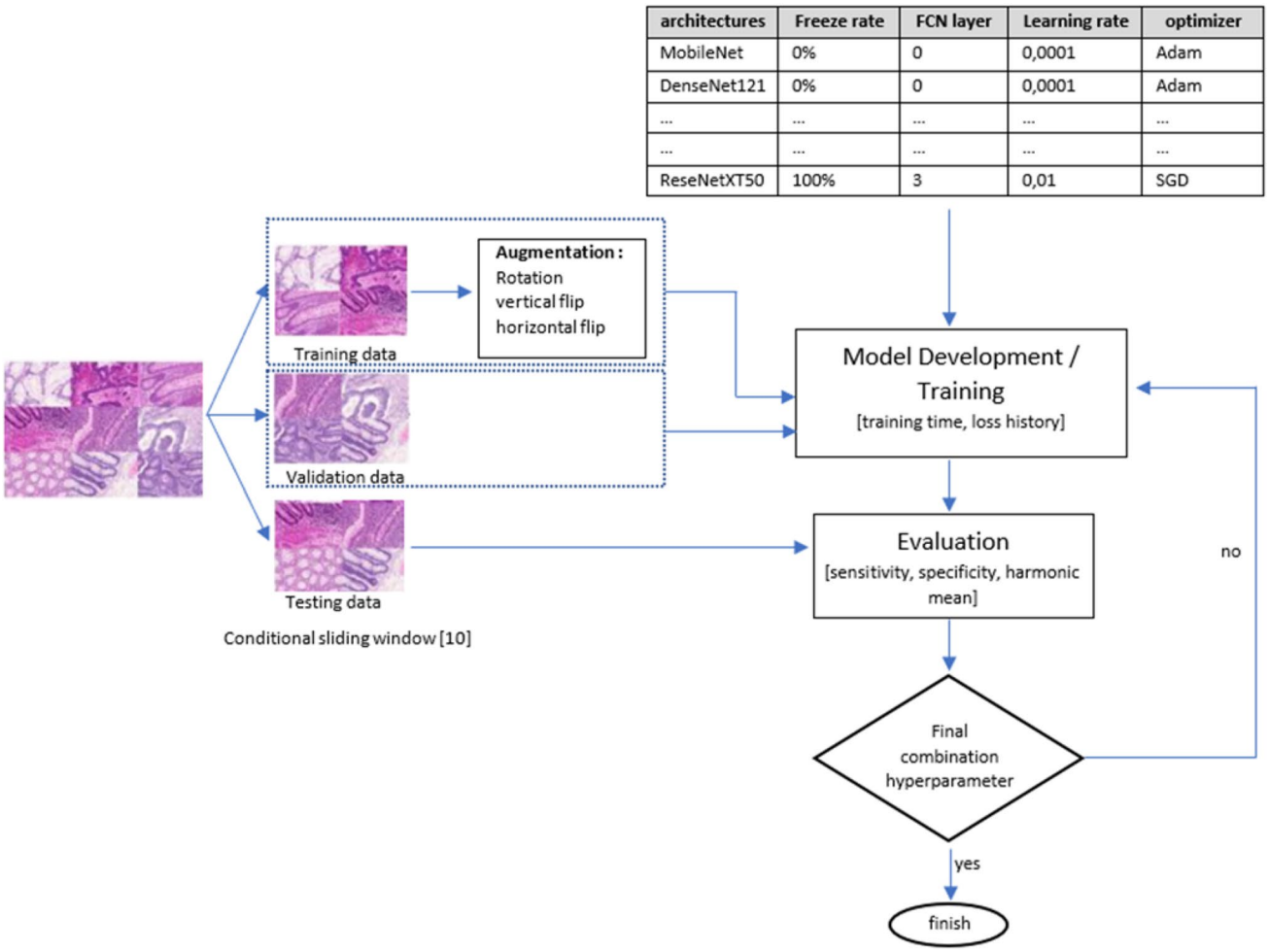
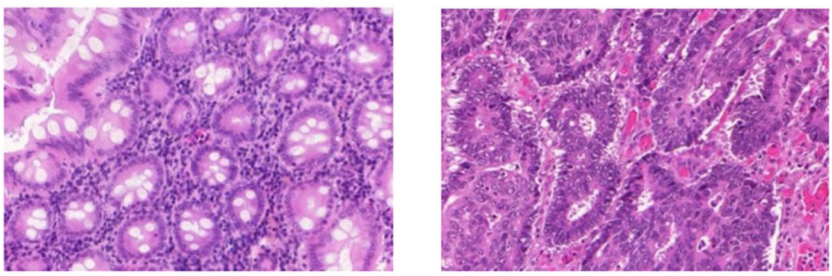


Fig. 1 Flowchart of robust model transfer learning model colorectal cancer identification

Table 2 Warwick dataset for colorectal histopathology

Class label	Number data		
	Training	Validation	Testing
Benign	37	33	4
Malignant	48	27	16

can be used as the initial parameters of another model that solves different problems, so the model does not need to start with randomly filled parameters. Then, this model will be retrained through a re-learning process (fine-tuning) to solve different problems. Several variables affect the ability of the model to learn. This variable is called a hyperparameter. Some examples of hyperparameters



(a) (b)

Fig. 2 **(a)** Benign tissues. **(b)** Malignant tissue

Table 3 Result of data generated by Conditional Sliding Windows (CSW)

Class label	Number data	
	Before CSW	After CSW
Benign	74	430
Malignant	91	532

Table 4 Detail applied augmentation for colorectal histopathology dataset

Augmentation techniques	Values range
Random horizontal flip	{0.1}
Random vertical flip	{0.1}
Random rotation	{−30–30}
Random adjust brightness	{−0.25–0.25}
Random adjust hue	{−0.4–0.4}
Random adjust saturation	{−0.25–0.25}
Random adjust contrast	{−0.75–0.75}

are architecture, learning rate, optimizer, regularization, batch size, and initial weight. Models that use their parameters as initial parameters for other models need to be trained on a large number of datasets. In another sense, the model must be a robust model. Transfer learning effectively solves problems that use images as a modality because images generally have similar low-level characteristics. Because of its shape, there are additional properties when CNN uses transfer learning. CNN consists of many layers, and in general, it can be divided into two parts, namely feature extractor and classifier. The feature extractor section transforms images into feature maps, and the classifier then uses these feature maps to produce the output labels. In transfer learning, a new model is created, and its parameters are initialized with parameters from the learned model to solve another problem. There is an option not to retrain some of the initial/low-level layers due to the nature of low-level similarity, meaning that the parameters in the low-level layers produce a good feature map for new problems and do not need to be retrained. This is called the freeze rate. An illustration of the transfer learning process to create a final model is presented in Fig. 4.

Hyperparameter optimization

Hyperparameters are very important in machine learning because they directly control the training process and have a significant effect on a model performance machine

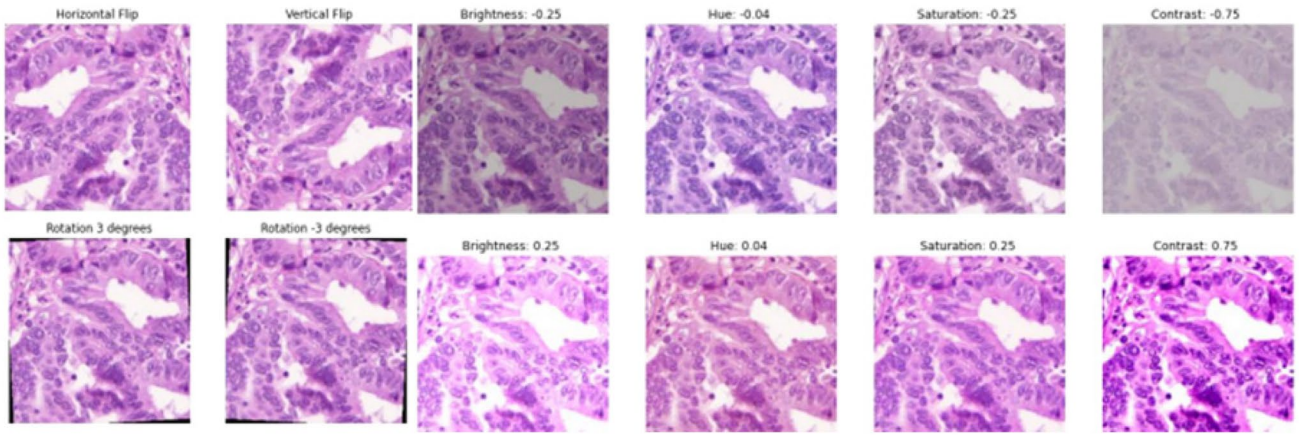


Fig. 3 Visualization of augmentation techniques

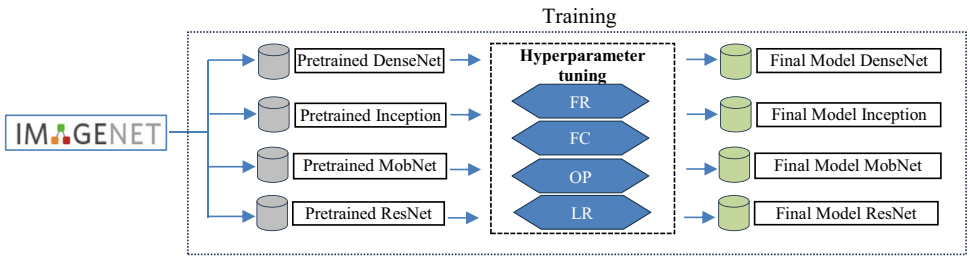


Fig. 4 Illustration of transfer learning approach

learning (Hutter et al. 2019). Hyperparameters are variables that must be determined before the training process begins. Finding the most optimal set of hyperparameter values is called hyperparameter optimization. In this research, hyperparameters are included in the process search: CNN architecture, freeze rate, number of layers in the classifier, optimizer, learning rate, and initial weight. There are two methods for hyperparameter optimization, grid search and random search. In this research, we apply grid search to obtain the best hyperparameters. Grid search is usually used for small search spaces of hyperparameters, but if we have so many hyperparameters, random search is a more suitable technique. The specification of architectures for transfer learning adopted in this research is shown in Table 5.

The hyperparameters optimization process is carried out using the grid search method. The program will build and train the model with every combination of the hyperparameter set to reach the conclusion of the hyperparameter combination with the best performance. Because there is an element of randomness in the learning process, this process is repeated five times so that the conclusions reached are correct. Table 5 displays the hyperparameters and their values to be used in the search process. From the table, it can be calculated that the total set of combinations investigated is 945 combinations model candidates.

The choice of freeze rate is related to the ability of the model in the feature learning layer to histopathology images. With several percentage values of freeze rate, we are able to gain insight into the selected basic architecture. The value of the freeze rate shows how much of the percentage of weight is used as non-trainable parameters. As an illustration, if we use a value of 75%, it means we only use 25% of the layer to learn to adapt to the new dataset. This approach is particularly useful when the new dataset is relatively small, as it prevents overfitting and makes the training process more efficient. For this reason, in this study, we used several combinations of values considering quartiles statistically.

A fully connected layer is the model's ability to train the results of the feature learning process. We use three optimizers that are widely applied to transitions between layers, namely: Adam, RMSProp, and SGD. Some values are used to consider how detailed the model is able to learn from the available dataset, as well.

Table 6 Hardware specification for training

Hardware	Specification
CPU	Dual AMD Rome 7742
RAM	32 GB
GPU	NVIDIA Tensor Core GPU A100 FP64 Tensor Core: 19.5 TFLOPS—VRAM 63 GB GPU Memory bandwidth: 1.935 GB/s

Experimental setup

For each combination, the model will be trained for a maximum of 100 epochs. In each epoch, there are two phases, namely the training and the validation phases. In the training phase, the model will predict the labels from the input set train, which will then be evaluated using the Cross-Entropy function. The loss value will be used to adjust the model parameters with the optimizer algorithm and learning rate. The duration of this phase will be recorded for each epoch; the loss value, accuracy, sensitivity, and specificity are also recorded. In the validation phase, the model will again be evaluated with the validation set as input. Loss values, accuracy, sensitivity, and specificity are also recorded in this phase. This training process also uses the Early Stopping mechanism.

The training process will stop five epochs after the model reaches both sensitivity and specificity evaluations above 95%. This was done because the pre-experiments showed quite a number of combinations that converged in a few epochs. So the authors decided to set the threshold for early stopping. After running the last epoch, the model will be filled again with the parameters in the epoch with the smallest loss. The training process uses a batch with size of 32. A batch size of 32 will accelerate the step toward

Table 5 Combination hyperparameters for deep learning architectures

Architectures	Freeze rate	Number of fully connected layer	Optimizer	Learning rate
DenseNet121 DenseNet161 Inception_v3 MobileNet_v2 ResNet18 ResNet50 ResNext50	0%, 25%, 50%, 75%, 100%	0, 1, 2	Adam RMSProp SGD	0.01 0.001 0.0001

convergence but is quite computationally heavy. Each model in this study was trained using the Nvidia DGX A100 system hardware, which is explained in Table 6. The library used to train the model in this study is Pytorch. Pytorch provides many pre-trained models to solve classification problems on ImageNet datasets. Meanwhile, the pre-trained model on CAMELYON has a ResNet architecture, and hyperparameter optimization is only done with this architecture.

Model evaluation

The evaluation phase begins by using a test set as the input model. The last layer of the model consists of two neurons that represent labels (benign, malignant). The Softmax function converts the model output into probabilities for each class. Therefore, the label predicted by the model can be concluded as the class with the most significant probability.

Because the type of problem in this study is binary classification, the determination of evaluation metrics (true positive (TP), false positive (FP), true negative (TN), and false negative (FN)) can be calculated easily. From these four assessment metrics, model accuracy performance can be measured with accuracy (1), precision, f1-score, sensitivity, and specificity metrics. ROC and AUC metrics can also be calculated using the results of calculating the probability of malignant class test data. The metric used as a reference for model accuracy performance is the harmonic mean defined (4). The value of the accuracy metric can be affected by the distribution of the data. The harmonic mean metric is used because the class distribution in the test set is unbalanced; the harmonic mean measures the accuracy of the model as if we were "blind" to the class distribution.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{1}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{2}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

$$\text{Harmonic mean} = \sqrt{\text{Sensitivity} \times \text{Specificity}} \tag{4}$$

Results

Experiment result and trade-off

To determine the best architecture, we should consider performance and training time. As the initial result in terms of performance, Table 7 shows us the model performance ascending by sensitivity because an acceptable rate for clinical diagnosis is very critical in medicine.

However, the best architectures consider how long we spend training for our model. Table 7 records information related to the performance of the model using harmonic mean and total training time as well. The best model was determined by grouping experimental results based on their accuracy and looking for the fastest hyperparameter combination at that level of accuracy. With this method, we can determine where the trade-off between accuracy and training time is the most optimal. It can be seen in Table 8 that the hyperparameter combination densenet121-75–0-0.001-RMSprop has an accuracy difference of only 0.5% with the hyperparameter combination with the best accuracy, namely inception_v3-50–0-0.01-Adam. The train time required by densenet121-75–0-0.001-RMSprop is only 19.5 s, which is 30 s faster than the time required by inception_v3-50–0-0.01-Adam. So it can be said that densenet121-75–0-0.001-RMSprop is one of the best hyperparameter combination candidates.

Table 7 Initial performance of architectures ascending by the highest sensitivity

No	Architectures–FreezeRate–LayerFCN–Learning rate–Optimizer	Harmonic mean	Sensitivity	Specificity
1	inception_v3-50–0-0.01-Adam	0.989	0.979	1.000
2	mobilenet_v2-50–1-0.0001-RMSprop	0.984	0.968	1.000
3	resnext50_32×4d-75–2-0.01-SGD	0.984	0.968	1.000
4	resnext50_32×4d-75–1-0.0001-Adam	0.984	0.968	1.000
5	densenet121-75–2-0.01-RMSprop	0.984	0.968	1.000
6	resnext50_32×4d-50–1-0.0001-SGD	0.984	0.968	1.000
7	densenet161-75–2-0.01-RMSprop	0.984	0.968	1.000
8	densenet121-75–0-0.001-RMSprop	0.984	0.968	1.000
9	mobilenet_v2-75–1-0.01-Adam	0.984	0.968	1.000
10	densenet161-75–1-0.01-Adam	0.984	0.968	1.000

Table 8 Performance of transfer learning with full training time

No	Architectures–FreezeRate–LayerFCN–Learning rate–Optimizer	Total training time	Harmonic mean
1	resnet18-75–0-0.0001-Adam	15.9 s	0.95
2	resnext50_32×4d-75–2-0.0001-RMSprop	17.2 s	0.96
3	densenet121-25–0-0.0001-RMSprop	18.5 s	0.97
4	inception_v3-50–0-0.01-SGD	18.6 s	0.95
5	resnext50_32×4d-75–2-0.0001-Adam	18.9 s	0.96
6	resnext50_32×4d-75–0-0.01-SGD	19.0 s	0.96
7	resnext50_32×4d-75–0-0.0001-RMSprop	19.1 s	0.97
8	**densenet121-75–0-0.001-RMSprop**	**19.5 s**	**0.98**
9	resnet18-50–1-0.0001-RMSprop	19.6 s	0.96
10	resnet18-0–1-0.01-SGD	19.7 s	0.95

The data in bold indicate the best combination of hyperparameters

Table 9 Combination hyperparameters

Hyperparameters	Values
Architecture	DenseNet121
Freeze rate	75%
#layer FCN	0
Learning rate	0.001
Optimizer	RMSProp
Initial weight	ImageNet

Table 10 Confusion matrix

Actual class	Predicted class	
	Benign	Malignant
Benign	24	0
Malignant	3	93

The combination of densenet121-75–0-0.001-RMSprop has a training time of 3 s slower than the fastest combination but with an accuracy difference of 3%. The increase in training time for 3 s is not significant compared to the 3% increase in accuracy. So, it can be concluded that the combination of densenet121-75–0-0.001-RMSprop is the best hyperparameter combination in accuracy and training time required (Table 9).

The model selected will be analyzed more deeply. The analysis was carried out by evaluating the model with metrics such as recall, precision, and f1-score. To evaluate the model, the image data after augmented are tested and reported via the confusion matrix presented in Table 10.

Experimentally, the DenseNet architecture with a total of 121 layers (DenseNet121) was the architecture chosen. By combining hyperparameters such as freeze rate configuration, number of fully connected layers, activation function, and learning rate, we get a robust model as the findings of this research.

Comparative analysis

Based on the results of the research conducted, we compared it with previous studies on several metrics parameters such as accuracy, sensitivity, specificity, recall, precision, and f1-score. All comparative analysis is limited to the existing methods using the Warwick dataset for colorectal cancer. The results are presented in Table 11. In the field of anatomical pathology, pathologists carry out tests by analyzing sensitivity, specificity, f1-score, and accuracy values, and then comparing them with previous studies. Table 11 presents the results of our proposed model among several models from other studies using similar dataset.

Table 11 Comparative analysis with existing techniques using the Warwick dataset (Sirinukunwattana et al. 2015, 2016)

Methods	Accuracy	Sensitivity	Specificity	Precision	f1-score
CNN 7–5-7 (Haryanto et al. 2021)	-	0.820	0.815	-	-
DCAN (Chen et al. 2017)	-	-	-	-	0.887
ResNet (Escorcia-Gutierrez et al. 2022)	0.800	0.870	0.830	-	-
PHP + CNN (Qaiser et al. 2019)	-	0,922	-	**0.936**	**0.924**
Micro-Net (Raza et al. 2019)	-	-	-	-	0.913
Ours (densenet121-75–0-0.001-RMSprop)	**0.975**	**0.968**	**1**	0.88	0.920

The data in bold indicates the highest value in term of metrics performance

Freeze rate analysis to training time

Based on the literature study conducted, not many studies have analyzed the effect of the freeze rate on model performance. CNN consists of many layers, and in general, it can be divided into two parts, namely feature extractor and classifier. The feature extractor section converts input images into feature maps, whereas the classifier uses these feature maps to produce label outputs. A freeze rate of 0% means that all layers in the feature extractor are retrained and a freeze rate of 100% means that no layers in the feature extractor are retrained; in other words, the feature extractor section uses the weights from the model trained with different datasets and only the classifier section retrains with the dataset. Warwick. A freeze rate of 25% means that 25% of the feature extractor layers from the very front are frozen (not retrained), and the remaining 75% are retrained. A 0% freeze rate will be used for performance changes to other freeze rate values. Table 12 shows the impact of the freeze rate to the training time for each transfer learning model.

Analysis impact of initial weight

In the transfer learning technique, the weights of the newly created model are filled with the weights of other models (but with the same architecture) that have been optimized to solve specific problems. The optimized weight is specific to solving that specific problem; in other words, other models trained with different datasets will have different optimized weights. So that we can determine that our model is filled with the weight (initial weight) of the model that was trained to solve different problems and datasets. In this study, the weights of the models trained with the ImageNet and Camelyon datasets will be used as initial weights. ImageNet was chosen because this dataset has a large number of classes and a very large amount of data. Camelyon was chosen because this dataset contains histopathological images of breast cancer, which are similar in characteristics to the dataset used in this study.

The discussion of initial weights is separated from other hyperparameters because only the resnet18 architecture is available with the initial weight Camelyon16. The effect of the initial weight is analyzed using the same method as the freeze rate analysis process, which compares the performance between models with the same configuration but different initial weights. Of the models that have the Camelyon initial weight, the hyperparameter combinations that produce the best models are shown in Table 13.

Comparative analysis of the initial training time and performance of Imagenet and camelyon

Table 14 shows the average change in model performance that has the same set of hyperparameter values but with different initial weight values at each freeze rate level. It was found that the model with the initial weight ImageNet

Table 13 The best parameter combination for Camelyon initial weight

Hyperparameter	Value
Architecture	ResNet18
Freeze rate	0%
Total layer FCN	1
Learning rate	0.0001
Optimizer	Adam
Initial weight	Camelyon

Table 12 Freeze rate analysis to training time

Model #trainable parameters	Metrics	Freeze rate				
		(0%) Baseline	25%	50%	75%	100%
MobileNetV2-2.226.434	Average total training time	2m0s	1m39s	1m18s	1m11s	2m38s
	difference with baseline	-	0m21s	0m42s	0m49s	0m37s
DenseNet121-6.955.906	Average total training time	2m13s	1m11s	1m7s	1m16s	3m6s
	difference with baseline	-	1m2s	1m6s	0m56s	0m53s
ResNet18-11.177.538	Average total training time	1m49s	1m34s	1m5s	0m57s	3m21s
	difference with baseline	-	0m15s	0m43s	0m51s	1m31s
ResNext50-22.984.002	Average total training time	2m20s	1m13s	1m3s	0m59s	2m40s
	difference with baseline	-	1m6s	1m16s	1m21s	0m20s
ResNet50-23.512.130	Average total training time	4m55s	2m45s	2m15s	1m53s	3m11s
	difference with baseline		2m10s	2m39s	3m2s	1m44s
Inception_V3-25.116.362	Average total training time	2m28s	1m41s	1m11s	1m14s	2m37s
	difference with baseline		0m47s	1m17s	1m13s	0m9s
DenseNet161-26.476.418	Average total training time	5m19s	2m17s	2m11s	1m47s	5m7s
	difference with baseline		3m2s	3m8s	3m32s	0m12s

In column baseline, the information "difference with baseline", all off value is "-"

Table 14 Comparative analysis of ImageNet and Camelyon dataset

Metrics	Initial weight	Freeze rate				
		0%	25%	50%	75%	100%
Total training time	Camelyon	2m0s	3m36s	2m56s	2m23s	3m42s
	ImageNet	**1m49s**	**1m34s**	**1m5s**	**0m57s**	**3m21s**
Accuracy	Camelyon	**88.0%**	92.2%	94.8%	**95.6%**	85.1%
	ImageNet	87.5%	**94.8%**	**94.9%**	95.1%	**91.5%**

Data in bold emphasis indicate the highest value in terma of freeze rate ratio from the best model used

achieves faster convergence than the model with the initial weight Camelyon at each freeze rate.

Discussion

Research related to histopathological image-based cancer is still an interesting study. Issues regarding limited medical data, performance, and training time are some of the points of the researchers. Starting from scratch with data limitation, the model has the potential for overfitting. Therefore, the transfer learning or pre-training approach was conducted to address this problem. The current transfer learning approach with ImageNet (Rusak et al. 2022) and Camelyon datasets (Litjens et al. 2018). Various transfer learning architectures that exist today are a challenge for researchers to get models that are robust and ready to be implemented. Histopathological images of colon cancer are relatively difficult to obtain because of the position of the tissue cells in the patient's body. This greatly affects the availability of datasets. Therefore, with transfer learning, this limited data availability is helped by the template model of the previous architecture by implementing transfer learning. However, the optimization process must be carried out to answer the open problems contained in creating specific deep learning models to solve our problems.

The performance of a model is sometimes faced with a trade-off between performance and training time. For this reason, a hyperparameter tuning or optimization process is needed. From the ten transfer learning architectures trained, the best combination was obtained as reported in Table 8. In terms of architecture, DenseNet121 was the best architectural option chosen with a harmonic mean value of 0.98 and a training time of 19.5 s per epoch. The 75% freeze rate is the optimal hyperparameter for re-training DenseNet121. This means that 25 percent of the parameters will be used with new data in the training process. In some research in terms of biomedical image classification, DenseNet121 obtained the best classifier for COVID-19 classification using CT images dataset (Hasan et al. 2021) and gained good performance for breast cancer identification via the BreakHIS dataset (Li et al. 2020). Compared to other architectures, DenseNet121 has smaller number parameters (Fig. 5). From the figure, can be illustrated that if DenseNet121 has 6.95.906 number parameters, only 1,738,977 parameters are updated with the new dataset. Meanwhile, 5,216,930 parameters are freezed and obtained from the pre-trained model.

Since ImageNet is a robust dataset (Rusak et al. 2022) for transfer learning to detect face spoofing attack (Rusia and Singh 2022), our model adopts this dataset's initial weight

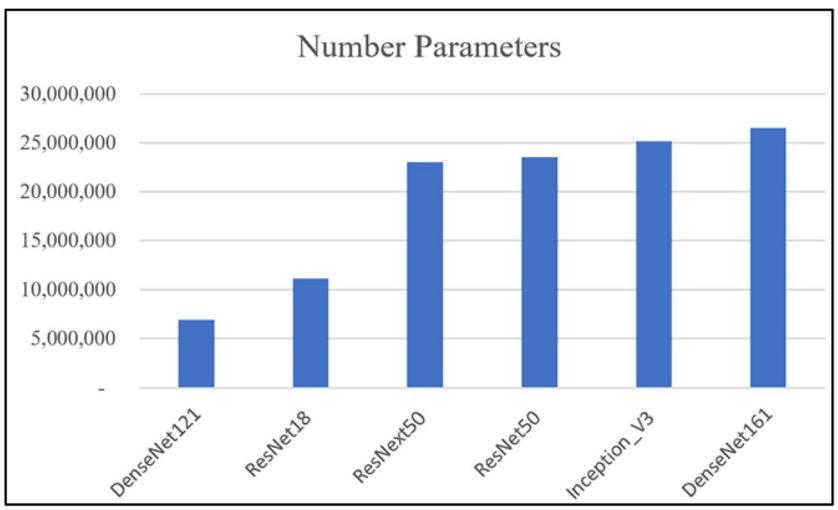


Fig. 5 Comparison number parameters of architecture used in this research

(Russakovsky et al. 2015). So, transfer learning with fellow cancer data is focused on the ImageNet, as a baseline initial weight. In fact, from our research, using the initial weight with ImageNet is still better even when compared to the Camelyon dataset which is actually based on the image of the cancer itself. This can be seen from the performance in terms of time (Table 13). On the other hand, the transfer learning model based on cancer histopathology images is still limited. This is certainly an opportunity as well as a challenge for researchers in the field of computer vision for cancer images.

The hyperparameters, CNN architecture, and freeze rate influence the performance of the image analysis solution. The CNN DenseNet121 architecture with a freeze rate of 75% exhibits the optimal hyperparameter value in this problem. Furthermore, the initial weight of ImageNet, the learning rate of 0.001, and the optimizer RMSProp all exert an influence. The ideal configuration of these hyperparameters yields a model with significant precision value and good performance in training time. It is interesting to note that the findings outperformed the previous performance models. This study investigated the impact of freezing rate and initial weight on the precision and duration of training. Our findings indicate that a freeze rate of about 75% yielded the model with the maximum accuracy and training duration. Examining the impact of the starting weight demonstrates that employing the weight of the model built using the ImageNet dataset offers a good performance model for the Camelyon dataset. The proposed approach, utilizing transfer learning and optimized hyperparameter settings, appears acceptable for classifying image-based data in histopathology images especially when data is imbalanced.

A limitation of the currently developed model is that we have not considered other structures of the histopathology image used outside the cell. The potential for bias that occurs is if the annotation process is carried out by people who do not have much experience with histology.

Conclusion

This research succeeded in building a CNN model with transfer learning techniques. This research found that the best hyperparameter value in this problem is the model with CNN DenseNet121 architecture, freeze rate 75%, initial weight ImageNet, learning rate 0.001, and optimizer RMSProp. The combination of these hyperparameter values produces a model with an accuracy of 0.98 with a training time of 19.5 s. These results surpass the performance of previous research models. This study analyzed the effect of freeze rate and initial weight on accuracy and training time as well. A freeze rate of 75% delivers the most accurate results and the best training time for the model. Analysis of the effect of the initial weight shows that using the weight of the model trained with the ImageNet dataset has better performance than the model trained with the Camelyon dataset Transfer learning has several advantages, including lower data requirements, increased performance, faster convergence, generalization to other domains, and facilitation of multi-task learning. These advantages make transfer learning an effective technique in machine learning research. Transfer learning architecture approach also has the potential to be used to detect cancer cell nuclei such as the YOLO Family. Another future direction, implementation, or testing of many real datasets becomes challenging. This research could be useful for other healthcare and medical image analysis, such as brain tumors, breast cancer, and skin diseases.

Funding This research was supported by the grant scheme UI-UiTM Bilateral Strategic Alliance (BISA) Matching Grant year 2021–2022.

Declarations

Conflict of interest The authors declare no competing interests.

References

Bejnordi BE, Veta M, Van Diest PJ, Van Ginneken B, Karssemeijer N, Litjens G, Van Der Laak JAWM, Hermsen M, Manson QF, Balkenhol M, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA - J Am Med Assoc. 2017;318(22):2199–210. https://doi.org/10.1001/jama.2017.14585.

Chen H, Qi X, Yu L, Dou Q, Qin J, Heng P. DCAN: Deep contour-aware networks for object instance segmentation from histology images. Med Image Anal. 2017;36:135–46. https://doi.org/10.1016/j.media.2016.11.004.

Editorial. Histopathology is ripe for automation. Nat Biomed Eng. 2017;1:925. https://doi.org/10.1038/s41551-017-0179-5.

Ershadi MM, Seifi A. Applications of dynamic feature selection and clustering methods to medical diagnosis. Appl Soft Comput. 2022;126:1–18. https://doi.org/10.1016/j.asoc.2022.109293.

Escorcia-Gutierrez J, Gamarra M, Ariza-Colpas PP, Roncallo GB, Leal N, Soto-Diaz R, Mansour RF. Galactic swarm optimization with deep transfer learning driven colorectal cancer classification for image guided intervention. Comput Electrical Eng. 2022;104:1–17. https://doi.org/10.1016/j.compeleceng.2022.108462.

Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B. histopathological image analysis: a review. IEEE Rev Biomed Eng. 2009;2:147–71.

Halalli B, Makandar A. Computer aided diagnosis - medical image analysis techniques. Intechopen. 2018;85–109. https://doi.org/10.5772/intechopen.69792.

Haq I, Mazhar T, Asif RN, Ghadi YY, Ullah N, Khan MA, Al-Rasheed A. YOLO and residual network for colorectal cancer cell detection and counting. Heliyon. 2024;10:2. https://doi.org/10.1016/j.heliyon.2024.e24403.

Haryanto T, Wasito I, Suhartanto H. Convolutional neural network (CNN) for gland images classification. In: Proceedings - International Conference On Information & Communication Technology

And System (ICTS) 2017. Surabaya; 2017. pp. 55–60. https://doi.org/10.1109/ICTS.2017.8265646.

Haryanto T, Suhartanto H, Arymurthy AM, Kusmardi K. Conditional sliding windows: an approach for handling data limitation in colorectal histopathology image classification. Inform Med Unlocked. 2021;23:1–12. https://doi.org/10.1016/j.imu.2021.100565.

Hasan N, Bao Y, Shawon A, Huang Y. DenseNet convolutional neural networks application for predicting COVID-19 using CT image. SN Comput Sci. 2021;2(5):1–11. https://doi.org/10.1007/s42979-021-00782-7.

Hosseinzadeh Kassani S, Hosseinzadeh Kassani P, Wesolowski MJ, Schneider KA, Deters R. Deep transfer learning based model for colorectal cancer histopathology segmentation: a comparative study of deep pre-trained models. Int J Med Inform. 2022;159:1–11. https://doi.org/10.1016/j.ijmedinf.2021.104669.

Humayun M, Sujatha R, Almuayqil SN, Jhanjhi NZ. A transfer learning approach with a convolutional neural network for the classification of lung carcinoma. Healthcare (switzerland). 2022;10(6):1–15. https://doi.org/10.3390/healthcare10061058.

Hutter F, Kotthoff L, Vanschoren J. Automated machine learning; methods, systems, challenges. Springer; 2019. https://link.springer.com/book/10.1007/978-3-030-05318-5.

Idlahcen F, Himmi MM, Mahmoudi A. 2020. CNN-based approach for cervical cancer classification in whole-slide histopathology images. Di dalam: *ICLR Workshop on AI for Overcoming Global Disparities in Cancer Care,*. hlm 1–5. https://arxiv.org/abs/2005.13924.

Inoue T, Yagi Y. Color standardization and optimization in whole slide imaging. Clin Diagn Pathol. 2020;4(1):1–22. https://doi.org/10.15761/cdp.1000139.

Kalaiarasan R, Madhan Kumar K, Sridhar S, Yuvarai M. Deep learning-based transfer learning for classification of skin cancer. Sensors (switzerland). 2022;21(8124):450–4. https://doi.org/10.1109/ICAAIC53929.2022.9792651.

Karthikeyan A, Jothilakshmi S, Suthir S. Colorectal cancer detection based on convolutional neural networks (CNN) and ranking algorithm. Meas Sensors. 2024;31:100976. https://doi.org/10.1016/j.measen.2023.100976.

Khan A, Brouwer N, Blank A, Müller F, Soldini D, Noske A, Gaus E, Brandt S, Nagtegaal I, Dawson H, et al. Computer-assisted diagnosis of lymph node metastases in colorectal cancers using transfer learning with an ensemble model. Mod Pathol. 2023;36(5):100118. https://doi.org/10.1016/j.modpat.2023.100118.

KhazaeeFadafen M, Rezaee K. Ensemble-based multi-tissue classification approach of colorectal cancer histology images using a novel hybrid deep learning framework. Sci Rep. 2023;13:1. https://doi.org/10.1038/s41598-023-35431-x.

Kudva V, Prasad K, Guruvare S. Hybrid transfer learning for classification of uterine cervix images for cervical cancer screening. J Digit Imaging. 2020;33:619–31. https://doi.org/10.1007/978-981-15-0626-0_25.

Li C, Xue D, Zhou X, Zhang J, Zhang H, Yao Y, Kong F, Zhang L, Sun H. Transfer learning based classification of cervical cancer immunohistochemistry images. In: Proceedings-International Symposium on Image Computing and Digital Medicine. 2019. pp. 102–6. https://doi.org/10.1145/3364836.3364857.

Li X, Shen X, Zhou Y, Wang X, Li TQ. Classification of breast cancer histopathological images using interleaved DenseNet with SENet (IDSNet). PLoS One. 2020;15:5. https://doi.org/10.1371/journal.pone.0232127.

Lin JS, Perdue LA, Henrikson NB, Bean SI, Blasi PR. Screening for colorectal cancer: updated evidence report and systematic review for the US preventive services task force. JAMA - J Am Med Assoc. 2021;325(19):1978–97. https://doi.org/10.1001/jama.2021.4417.

Litjens G, Bandi P, Bejnordi BE, Geessink O, Balkenhol M, Bult P, Halilovic A, Hermsen M, van de Loo R, Vogels R, et al. 1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset. Gigascience. 2018;7:6. https://doi.org/10.1093/gigascience/giy065.

Liu Y, Gadepalli K, Norouzi M, Dahl GE, Kohlberger T, Boyko A, Venugopalan S, Timofeev A, Nelson PQ, Corrado GS, *et al.* 2017. Detecting cancer metastases on gigapixel pathology images. http://arxiv.org/abs/1703.02442.

Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: challenges and opportunities. Med Image Anal. 2016;33:170–5. https://doi.org/10.1016/j.media.2016.06.037.

Nemlander E, Ewing M, Abedi E, Hasselström J, Sjövall A, Carlsson AC, Rosenblad A. A machine learning tool for identifying non-metastatic colorectal cancer in primary care. Eur J Cancer. 2023;182:100–6. https://doi.org/10.1016/j.ejca.2023.01.011.

Nguyen LD, Lin D, Lin Z, Cao J. Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation. In: Proceedings - IEEE International Symposium on Circuits and Systems. 2018. pp. 1–5. https://doi.org/10.1109/ISCAS.2018.8351550.

Panigrahi S, Nanda A, Swarnkar T. A survey on transfer learning. Smart Innovation, Syst Technol. 2021;194:781–9. https://doi.org/10.1007/978-981-15-5971-6_83.

Pantanowitz L, Szymas J, Yagi Y, Wilbur D. Whole slide imaging for educational purposes. J Pathol Inform. 2012;3(1):46. https://doi.org/10.4103/2153-3539.104908.

Qaiser T, Tsang YW, Taniyama D, Sakamoto N, Nakane K, Epstein D, Rajpoot N. Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features. Med Image Anal. 2019;55:1–14. https://doi.org/10.1016/j.media.2019.03.014.

Raza SEA, Cheung L, Shaban M, Graham S, Epstein D, Pelengaris S, Khan M, Rajpoot NM. Micro-Net: a unified model for segmentation of various objects in microscopy images. Med Image Anal. 2019;52:160–73. https://doi.org/10.1016/j.media.2018.12.003.

Reis HC, Turk V. Transfer learning approach and nucleus segmentation with MedCLNet colon cancer database. J Digit Imaging. 2023;36(1):306–25. https://doi.org/10.1007/s10278-022-00701-z.

Ren G, Li B, Lam SK, Xiao H, Huang YH, Cheung ALY, Lu Y, Mao R, Ge H, Kong FM, et al. A transfer learning framework for deep learning-based CT-to-perfusion mapping on lung cancer patients. Front Oncol. 2022;12:1–11. https://doi.org/10.3389/fonc.2022.883516.

Rusak E, Schneider S, Pachitariu G, Gehler P, Bringmann O. ImageNet-D: a new challenging robustness dataset inspired by domain adaptation. In: International Conference on Machine Learning. 2022. pp. 1–5. https://openreview.net/pdf?id=LiC2vmzbpMO.

Rusia MK, Singh DK. A color-texture-based deep neural network technique to detect face spoofing attacks. Cybernet Inf Technol. 2022;22(3):127–45. https://doi.org/10.2478/cait-2022-0032.

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, et al. ImageNet large scale visual recognition challenge. Int J Comput vis. 2015;115(3):211–52. https://doi.org/10.1007/s11263-015-0816-y.

Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data. 2019;6:1. https://doi.org/10.1186/s40537-019-0197-0.

Sirinukunwattana K, Snead DRJ, Rajpoot NM. A stochastic polygons model for glandular structures in colon histology images. IEEE Trans Med Imaging. 2015;34(11):2366–78. https://doi.org/10.1109/TMI.2015.2433900.

Sirinukunwattana K, Raza SEA, Tsang YW, Snead DRJ, Cree IA, Rajpoot NM. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE Trans Med Imaging. 2016;35(5):1196–206. https://doi.org/10.1109/TMI.2016.2525803.

Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209–49. https://doi.org/10.3322/caac.21660.

Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. J Big Data. 2016;3(9):40. https://doi.org/10.1186/s40537-016-0043-6.

WHO, IARC. Estimated number of deaths in 2018, worldwide, all cancers, males, all ages. 2018. https://gco.iarc.fr/today/home.

Yao L, Li S, Tao Q, Mao Y, Dong J, Lu C, Han C, Qiu B, Huang Y, Huang X, et al. 2024. Deep learning for colorectal cancer detection in contrast-enhanced CT without bowel preparation: a retrospective, multicentre study. www.thelancet.com. Accessed 22 June 2024.

Zhou X, Lu Y, Wu Y, Yu Y, Liu Y, Chang W, Zhao Z, Chong W, Gao Z, Li Z, et al. Construction and validation of a deep learning prognostic model based on digital pathology images of stage III colorectal cancer. Eur J Surg Oncol. 2024;50:7. https://doi.org/10.1016/j.ejso.2024.108369.

Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, Xiong H, He Q. A comprehensive survey on transfer learning. Proc IEEE. 2021;109(1):43–76. https://doi.org/10.1109/JPROC.2020.3004555.

Zotin A, Hamad Y, Simonov K, Kurako M. Lung boundary detection for chest X-ray images classification based on GLCM and probabilistic neural networks. Procedia Comput Sci. 2019;159:1439–48. https://doi.org/10.1016/j.procs.2019.09.314.