

DeepCon: Unleashing the Power of Divide and Conquer Deep Learning for Colorectal Cancer Classification

Suhaib Chughtai¹, Zakaria Senousy¹, Ahmed Mahany², Nouh Sabri Elmitwally^{1,3},
Khalid N. Ismail^{1,4}, Mohamed Medhat Gaber¹, and Mohammed M. Abdelsamea^{1,5,*}

¹School of Computing and Digital Technology, Birmingham City University, Birmingham B4 7XG, UK

²Faculty of Computer and Information Sciences, Ain Shams University, Cairo 11566, Egypt

³Faculty of Computers and Artificial Intelligence, Cairo University, Giza 12613, Egypt

⁴Faculty of Computers and Information, Menoufia University, Shibin Al Kawm, Al Minufiyah, Egypt

⁵Department of Computer Science, Faculty of Computers and Information, Assiut University, Assiut 71515, Egypt.

*Corresponding author: Mohammed M. Abdelsamea (email: mohammed.abdelsamea@bcu.ac.uk).

ABSTRACT

Colorectal cancer (CRC) is the second leading cause of cancer-related mortality. Precise diagnosis of CRC plays a crucial role in increasing patient survival rates and formulating effective treatment strategies. Deep learning algorithms have demonstrated remarkable proficiency in the precise categorization of histopathology images. In this paper, we introduce a novel deep learning model, termed *DeepCon* which incorporates the divide-and-conquer principle into the classification task. *DeepCon* has been methodically conceived to scrutinize the influence of acquired composition on the learning process, with a specific application to the classification of histology images related to CRC. Our model harnesses pre-trained networks to extract features from both the source and target domains, employing a two-stage transfer learning approach encompassing multiple loss functions. Our transfer learning strategy exploits a learned composition of decomposed images to enhance the transferability of extracted features. The efficacy of the proposed model was assessed using a clinically valid dataset of 5000 CRC images. The experimental results reveal that *DeepCon* when coupled with the Xception network as the backbone model and subjected to extensive fine-tuning, achieved a remarkable accuracy rate of 98.4% and an F1 score of 98.4%.

INDEX TERMS Class Composition, Data Irregularity, Deep Learning, Colorectal Cancer, Image Classification

I. INTRODUCTION

COLORECTAL Cancer (CRC) is a prevalent form of cancer, ranking third in terms of global diagnosis rates. Unfortunately, it is also the second leading cause of cancer-related deaths, with a staggering 1.4 million new cases and 693,900 fatalities reported in 2020 alone [1]. However, there is hope for those affected by this disease. Early detection and precise diagnosis of CRC can significantly increase the chances of patient survival and pave the way for effective treatment plans.

Medical imaging is a crucial tool in the timely detection and diagnosis of colorectal cancer (CRC). Histopathological images are frequently employed in CRC diagnosis to identify abnormal cells and tissues in colon tissue samples. However, analyzing histopathology images is a complex task that demands a high level of expertise in pathology. To overcome

this challenge, deep learning algorithms have emerged as a promising solution, demonstrating remarkable accuracy in the classification of histology images. Among these algorithms, the Convolutional Neural Network (CNN) is one of the most widely used approaches, exhibiting exceptional performance in extracting salient features from images [2], [3].

One of the main challenges encountered in the medical field is the limited availability of samples. The collection and labeling of histology images can be challenging and time-consuming, and the size of available datasets in the field is often smaller than in other fields. As a result, transfer learning is a crucial technique that can be utilized to overcome the issue of limited availability of histology samples. Transfer learning is a machine learning technique where a pre-existing model that has been trained on a specific

task is used as a foundation for a new model to solve a different problem. It involves re-purposing pre-trained CNN models on larger datasets to improve the performance of models trained on smaller datasets [4].

In this paper, we present a novel deep learning approach, termed *DeepCon*, for the classification of CRC histopathology images. *DeepCon* is an advancement of the previously developed Decompose, Transfer, and Compose (DeTraC) model [5] that aims to investigate the impact of trained composition on the learning process. DeTraC relies on class decomposition by partitioning image classes into subsets based on specific guidelines and assigns new labels to these subsets. It then proceeds to composition, assigning each instance classified to a subclass to its original/parent class. In contrast, *DeepCon* retains class decomposition but takes a transformative step in the composition process. It introduces a learnable composition step that dynamically assembles subclass results using a pre-trained CNN, fine-tuning weights for precise data-driven composition.

The complex nature of histopathological images poses unique challenges, especially due to the intricate tissue structures. Accurately distinguishing between sub-classes within CRC is crucial for informing treatment decisions. *DeepCon* utilizes transfer learning with various modes by applying pre-trained models on ImageNet [6] to a dataset of 5000 colorectal cancer (CRC) images. Transfer learning from ImageNet-based models has been widely adopted and recognized as the efficient way of training deep learning models for classification tasks. A notable feature of *DeepCon* is that it incorporates learnable composition of the decomposed classes unlike the traditional composition presented in [5]. This is done by applying a pre-trained CNN to automatically learn and classify the input image into one of the original classes. The learned composition stage uses a pre-trained CNN which uses the CNN weights applied to the learning conducted for the subclass-level classification in addition to fine-tuning to cope with the original class-level classification. Our experiments show that the deep tuning mode of *DeepCon* outperforms the shallow tuning mode, achieving the highest accuracy among all pre-trained models employed. The use of learned composition in *DeepCon* is a significant advancement from the previous DeTraC model, which allows more efficient and effective learning. In a nutshell, during the decomposition phase of our divide-and-conquer approach in *DeepCon*, learning occurs in the divide stage, where classification is performed at the subclass level. Conversely, during the composition stage, learning takes place in the conquer stage, where it is done at the original class level.

DeepCon offers several advantages in this domain. Firstly, it achieves enhanced classification accuracy, enabling more precise diagnoses. Furthermore, its unique ability to perform subclass-level classification provides finer granularity, assisting clinicians in tailoring treatment plans. Additionally, *DeepCon* is characterized by its flexibility in transfer

learning, allowing it to adapt readily to various datasets and tasks. Lastly, it facilitates efficient learning, optimizing the utilization of available resources. However, it is important to acknowledge a limitation of our proposed solution. The inclusion of two transfer learning stages in *DeepCon*, while enhancing classification accuracy, may introduce computational overhead.

The contributions of the paper can be summarized as follows:

- Introduces *DeepCon*, a novel deep learning approach tailored for CRC histology image classification, addressing the critical need for accurate CRC diagnosis.
- Employs a unique divide-and-conquer methodology with a focus on learned composition, enhancing the transferability of features between domains and improving classification accuracy.
- Demonstrates the effectiveness of two-stage transfer learning with multiple loss functions, showcasing its potential for optimizing CRC histology image classification.

The paper is organized as follows. Section II provides an overview of the latest related work conducted for CRC histology image classification. In Section III, we describe the methodology of the proposed *DeepCon* model, including the different transfer learning modes utilized and the process of learned composition. Section IV presents the experimental study carried out on the CRC imaging dataset. Finally, in Section V, we draw conclusions from our findings.

II. Related Work

In this section, we review the latest work conducted for the classification of CRC histopathology images.

Several studies have been conducted in the field of CRC image analysis [7], [8] and more specifically in classification task using deep learning methods [9]–[12]. For example, Peng et al. [10] developed a multitask deep learning framework for simultaneous classification and retrieval of colorectal histopathological images, making use of the well-known concept of k-nearest neighbors to increase interpretation of the model. Their original framework can be built on top of any existing classification network (pre-trained models) by combining a triplet loss function with a novel triplet sampling strategy to compare distances between samples and adding a hashing loss function to accelerate searching for neighbors. Raczkowski et al. [13] introduced a Bayesian network called ARA-CNN, which is precise and reliable and employs an active approach to classify images of colorectal cancer histopathology. The network is developed based on residual network principles and utilizes variational dropout techniques in its design. Shaban et al. [9] presented a method for classifying colorectal images using a two-stacked CNN. Their approach involves incorporating a larger contextual view of the images through the use of a context-aware neural network. The model first converts the local representation of

a histopathology image into high-dimensional features and then combines these features while taking into account their spatial arrangement to make a final prediction.

Research conducted by [11] explored the application of the ResNet architecture in the detection of colorectal cancer through deep learning image classification. The study focused on training ResNet-18 and ResNet-50 models on colon glands images to differentiate between benign and malignant colorectal cancer. The work presented by [12] focused on using deep learning architectures to classify and identify colon cancer regions in sparsely annotated histopathological data. The study reviews and compares the latest CNNs and utilizes transfer learning techniques to overcome limited annotated data sets. The models were tested on the AiCOLO colon cancer and CRC-5000 datasets. The work presented by [14] aimed to automatically identify eight types of tissues in CRC histopathological evaluation using Transfer Learning from CNN architectures. CNN structures were modified to extract features from images, which were then fed into various machine learning methods, including naive Bayes, multilayer perceptron, k-nearest neighbors, random forest, and support vector machine (SVM). A total of 108 extractor-classifier combinations were evaluated, and the DenseNet169 with SVM achieved the best results.

Furthermore, Wang et al. [15] propose a transformer-based unsupervised contrastive learning strategy named semantically-relevant contrastive learning (SRCL), combined with a hybrid model CTransPath. This approach achieves exceptional results across diverse downstream tasks, highlighting its robustness and transferability. Kumar et al. [16] introduce CRCCN-Net, a lightweight convolutional neural network framework for automated colorectal tissue classification. The framework showcases impressive performance, positioning it as a potential diagnostic tool for clinicians. Additionally, Zhou et al. [17] developed the HCCANet method, a computer-aided diagnosis (CAD) system for grading colorectal cancer based on a CNN architecture and a novel attention mechanism named MCCBAM. The model's interpretability is improved through gradient-weighted class activation maps (Grad-CAM).

Moreover, Sabol et al. [18] contributed an explainable classifier to improve accountability in decision-making for colorectal cancer diagnosis from histopathological images. The model offers human-friendly explanations about the plausibility of decisions through a Cumulative Fuzzy Class Membership Criterion (CFCMC). The classifier is shown to be comparable to state-of-the-art neural networks in accuracy and is particularly suited for use by human experts in the medical domain.

Changjiang et al. [19] presented a framework that utilized features of varying magnifications of Whole Slide Images (WSI) to classify and localize colorectal cancer, relying solely on global labels. The work presented by Zhou et al. [20] developed a deep learning framework called the cell graph convolutional neural network (CGC-Net) to grade

colorectal cancer. Their method involves converting each large histopathological image into a graph representation, where the nuclei within the image are represented as nodes and the cellular associations are represented as edges based on the similarity between the nodes. The network utilizes the local features and spatial dependencies of the nodes to improve its accuracy. Haoyuan et al. [21] proposed the IL-MCAM framework for colorectal histopathology image classification, which involves two stages: automatic learning and interactive learning. The automatic learning stage uses three attention mechanism channels and CNNs to extract multiple channel features, while the interactive learning stage incorporates misclassified images into the training set to improve the model's classification performance.

Awan et al. [22] developed Best Alignment Metric (BAM), a gland-shape metric that correlates with the grade of colon cancer. Their model uses a Deep CNN to detect gland boundaries and an SVM classifier to determine cancer grade. Wang et al. [23] proposed a deep transferable semi-supervised domain adaptation model called HisNet-SSDA to classify histopathological WSIs with limited labeled data. The method uses a pre-trained network to extract features from both source and target domains, then matches the two domains via semi-supervised domain adaptation with multiple-weighted loss functions and a manifold regularization term. The final image-level classification is obtained by combining the estimated probabilities of the sampled patches.

Here, we present a deep learning approach, termed DeepCon, for the classification of histopathology images. DeepCon is a new divide and conquer deep learning technique that can address the challenge problem of data irregularity presented in the Histopathological Image because samples are characterized by high visual variability. DeepCon is working as a two-stage transfer learning approach to achieve coarse to fine transfer learning guided by a divide and conquer training approach to learn the composition of the decomposed images during the transfer learning.

III. Material and Methods

In this section, we explain in detail the different stages of *DeepCon*. First, we describe the dataset utilized in the study. Second, we explain the preprocessing stage performed to apply stain normalization for the dataset samples. Then, we describe the transfer learning modes used. We then explain the class decomposition and composition stages of *DeepCon*. Finally, we describe the training and hyperparameter settings utilized. Figure 1 describes the complete architecture of *DeepCon*, including all essential stages. *DeepCon* has three important stages. The first stage is to apply the decomposition to the samples of the normalized CRC dataset using the K-means clustering algorithm. The aim of applying decomposition is to simplify the complexity and data irregularity of dataset samples. This stage results in having 16 decomposed classes from the dataset. Then, as a second stage, we apply

initial transfer learning by utilizing pre-trained CNN trained on ImageNet. We applied fine-tuning to the last few layers of the network to cope with the specific problem we have (16-class classification task of the CRC dataset samples). Lastly, we introduce a novel automatic class composition strategy by applying second-round transfer learning. This is done by utilizing the pre-trained CNN on the decomposed CRC dataset in addition to fine-tuning to the last few layers to cope with the final classification of the original class labels (8-class classification task). See Figure 1.

A. Dataset

In this work, we used the dataset "CRC-VAL-HE-7K" from the Institute of Pathology (University Medical Center Mannheim, Heidelberg University, Mannheim, Germany) [24]. The dataset contains 5000 images, equally distributed in 8 classes of 625 images each; TUMOR, STROMA, COMPLEX, LYMPHO, DEBRIS, MUCOSA, ADIPOSE, and EMPTY/BACKGROUND. All images are 224×224 pixels at 0.5 microns per pixel with TIF format.

B. Pre-processing

Due to the high visual variability of the images, a stain normalization method [25] was used to combine the stain density maps with the stain color basis of a selected target image and consequently alter the color profile of the images in our dataset to preserve the structure described by the stain density maps; see Figure 2. Additionally, affine transformations were used for data augmentation to increase the size of the dataset. Moreover, random orthogonal rotations across were used in combination with horizontal and vertical flips. Although CNNs are not rotationally invariant, there is no correct orientation of histology images; therefore, the full 360-degree rotation was used. The reasoning for orthogonal rotations was to prevent any empty space because of rotation values other than multiples of 90 degrees. Although fill methods exist for such empty spaces, they could have a potentially negative effect and are also more computationally intensive than orthogonal rotation.

C. Class Decomposition

The research carried out by [5] explored a class decomposition approach in transfer learning known as DeTraC. The method involves clustering the classes within a dataset prior to training a network with transfer learning, with the aim of improving accuracy in situations where the data classes are unevenly distributed. After the training process, the labels are corrected back to their original superclass before evaluating performance metrics. The CRC dataset was used to evaluate the method and was tested in three different classes. The results demonstrated an increase in accuracy compared to other approaches. In this work, we developed a novel automated method for restoring the original classes from the decomposition stage with the aim of achieving higher accuracy.

The decomposition approach is implemented by following the steps outlined in the original paper [5]. For each of the 8 classes in the original dataset, a pre-trained Xception network was utilized to extract a 2048-dimensional feature vector for each of the 625 images in a particular class. We applied Principal Component Analysis (PCA) to project the high-dimensional feature space into a lower dimension, such that the highly correlated features were ignored. This step is crucial for the class decomposition to generate more homogeneous classes, minimize the requirement for memory, and increase the framework's effectiveness. In order to use PCA, the data was projected onto the top principal components after computing the eigenvectors of the feature covariance matrix. The feature vectors' covariance matrices were specifically constructed as follows after being centered:

$$\Sigma = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T, \quad (1)$$

where $n = 625$ is the number of images in the class, x_i is the feature vector of the i -th image, and μ is the mean feature vector of the class. T is the transpose applied to $(x_i - \mu)$. The eigenvectors of Σ were then computed, and the feature vectors were projected onto the upper d eigenvectors, where d was chosen so that the explained variance ratio was at least 0.95. This resulted in a reduced-dimension feature vector of approximately 150 dimensions per class. Let us assume that our feature space (PCA's output) is represented by a 2-D array, denoted as dataset A , and C is a set of class labels. We can express A and C as:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}, C = \{c_1, c_2, \dots, c_j\} \quad (2)$$

where j is the number of classes, m is the number of features, and n is the number of samples present in the dataset. Finally, to further categorize the images, the K-means clustering algorithm [26] with $k = 2$ was used to cluster each class into two subclasses.

Let $S = s_{11}, s_{12}, \dots, s_{1l}, s_{21}, s_{22}, \dots, s_{2l}, \dots, s_{jl}$ represent a set of subclasses/clusters, where $|S|$ denotes the number of elements in the set. This set is obtained through a homogeneous decomposition, where each class is divided into the same number of clusters/sub-classes, denoted by the variable l . Since we have j classes, the total number of elements in S is given by:

$$|S| = j \times l \quad (3)$$

Based on the above explanation of the new decomposed classes of the dataset, we can illustrate the feature space of both dataset A and B as follows:

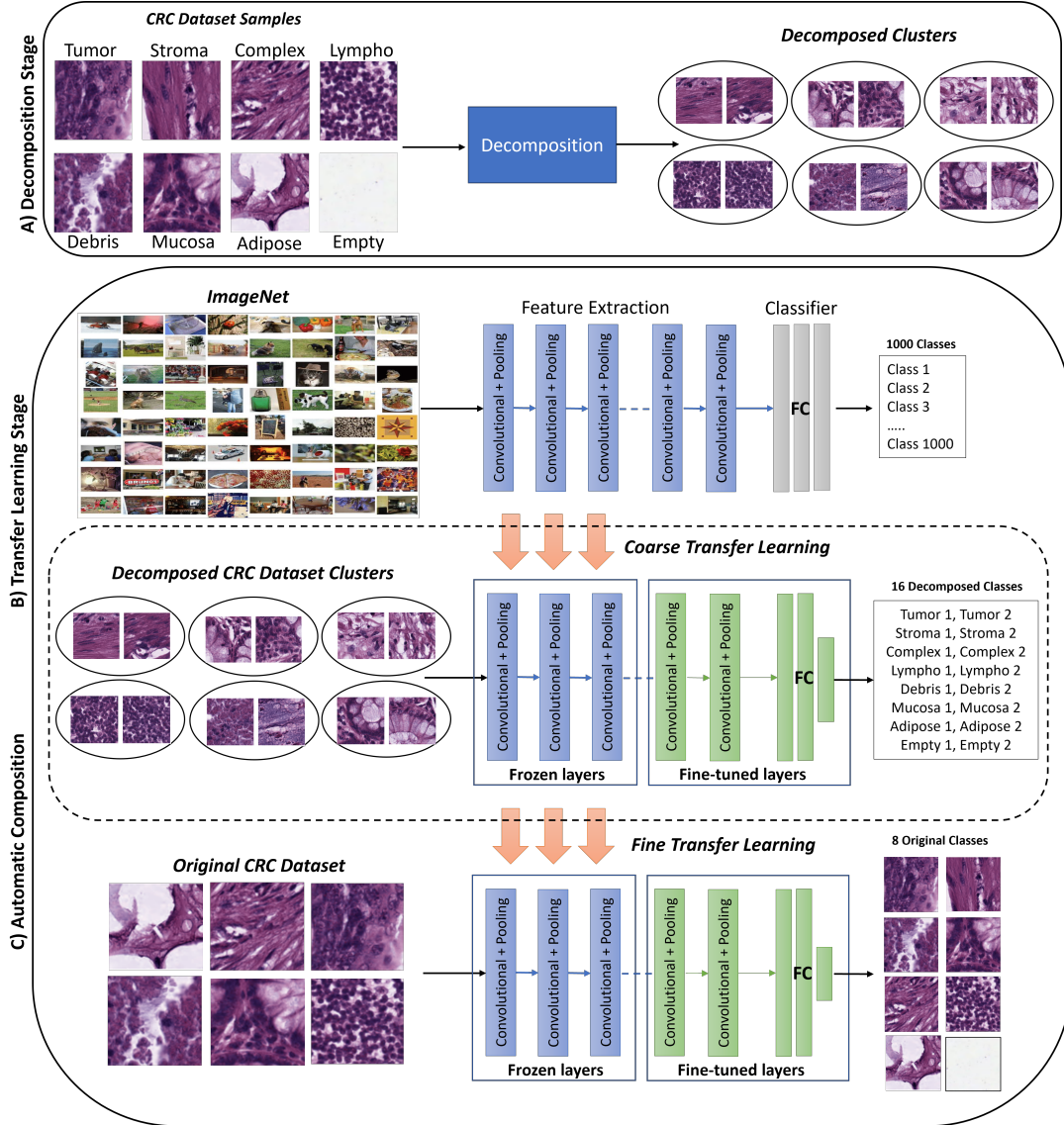


FIGURE 1. Full architecture of DeepCon model. A) Decomposition Stage: Applying clustering using KMeans on dataset samples resulted in 16 decomposed classes. B) Transfer Learning Stage: Transferring knowledge by using pre-trained CNN trained on ImageNet to the decomposed CRC dataset for classifying 16 decomposed classes. C) learned composition: Transferring knowledge by using pre-trained CNN trained on the decomposed CRC dataset to the original CRC dataset to classify 8 original classes.

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} & c_1 \\ a_{21} & a_{22} & \dots & a_{2m} & c_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & c_8 \\ a_{n1} & a_{n2} & \dots & a_{nm} & c_8 \end{bmatrix},$$

$$B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1m} & s_{11} \\ b_{21} & b_{22} & \dots & b_{2m} & s_{1j} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & s_{81} \\ b_{n1} & b_{n2} & \dots & b_{nm} & s_{8l} \end{bmatrix}.$$

(4)

Both matrices have n rows and $m + 1$ columns, where n is the number of instances or samples and m is the number of features. The last column of matrices A and B contain the class labels and sub-class labels of the corresponding instances, respectively.

Figure 3 shows a set of images that were randomly picked from the two clusters of the Stroma class. A clear visual difference can be observed between the images in each cluster, indicating that the clustering process was effective in separating the irregular images. Table 1 shows the distribution of dataset samples after the decomposition stage.

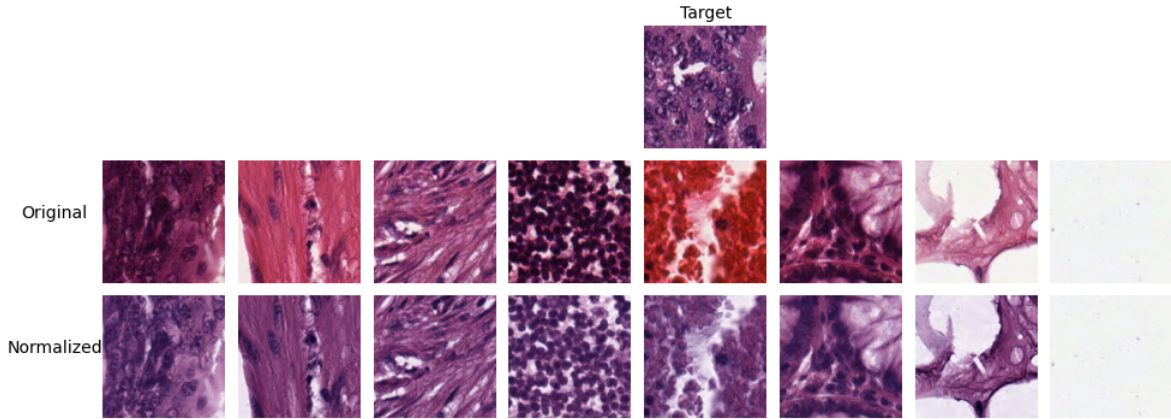


FIGURE 2. Stain normalization process. Transferring color from a target/reference image to source images.

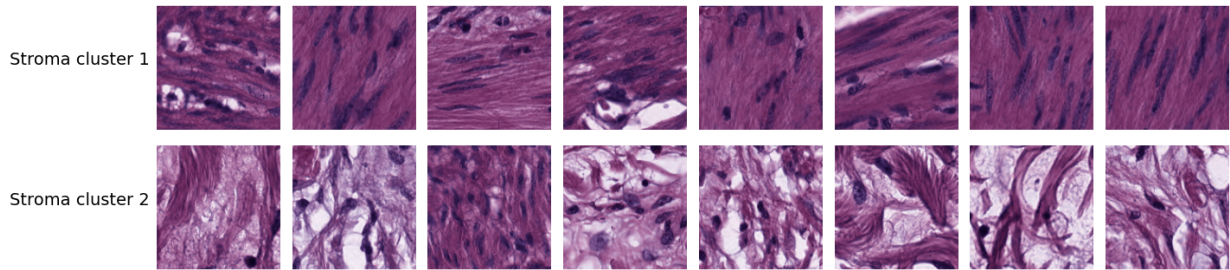


FIGURE 3. An example of images generated from the two clusters of the Stroma class.

TABLE 1. Samples Distribution in Datasets A and B

Dataset A	No. of Samples	Dataset B	No. of Samples
Tumor	625	Tumor 1	260
		Tumor 2	365
Stroma	625	Stroma 1	254
		Stroma 2	371
Complex	625	Complex 1	284
		Complex 2	341
Lympho	625	Lympho 1	226
		Lympho 2	399
Debris	625	Debris 1	390
		Debris 2	235
Mucosa	625	Mucosa 1	264
		Mucosa 2	361
Adipose	625	Adipose 1	378
		Adipose 2	247
Empty	625	Empty 1	307
		Empty 2	318

D. Coarse Transfer Learning

In this work, two models, namely Xception [27] and InceptionV3 [28], were chosen for transfer learning. The models

were initialized with pre-trained weights from ImageNet, but the top fully connected layers were not included. To reduce the number of parameters, the convolutional bases were passed through a GlobalAveragePooling2D layer instead of flattened. Then a new dense, fully connected layer with 1024 neurons and ReLU activation was added. The final output layer had eight neurons with SoftMax activation, corresponding to the number of classes. Consequently, to train the models, three transfer learning strategies were utilized: shallow, fine-tuning, and deep fine-tuning. In the shallow fine-tuning approach, only the fully connected layer is trained while the parameters of other layers in the convolutional network were frozen. The fine-tuning approach involves unfreezing the weights of layer 14 (25-35%) to adapt them to the dataset. Finally, all layers are trained in the deep fine-tuning approach.

As we are addressing a multi-class classification problem, the categorical cross-entropy loss function was used during the training process of the pre-trained CNN utilized for coarse transfer learning (Figure 1 B). In this stage, the decomposed classes generated from the decomposition stage were used. The cross-entropy loss is a measure of the difference between the predicted class probabilities and the true class probabilities, and it is defined as:

$$L_{Coarse}(y, \hat{y}) = - \sum_{i=1}^S y_i \log(\hat{y}_i), \quad (5)$$

where y is the true subclass probability distribution, \hat{y} is the predicted sub-class probability distribution, and S is the number of subclasses generated from the decomposition stage. The training process conducted for this stage is utilized in the next stage (learned composition) by transferring weights for decomposed classes classification in addition to fine-tuning to cope with the original classes classification.

E. Learnable Composition using fine transfer learning

One of the standout features of *DeepCon* is its ability to learn the composition to predict original classes. This is achieved by using one more stage of transfer learning in a pretrained CNN with the aim to learn some of the features generated for the decomposed classes and applying fine-tuning to fit to a classification task of the original classes. In other words, to achieve the original class-level classification, the learned composition stage of *DeepCon* uses a pretrained CNN that is fine-tuned to handle the original class-level classification using the CNN weights applied to the learning conducted for the subclass-level classification.

Similar to the previous pre-trained CNN utilized for the coarse transfer learning, categorical cross-entropy is used during the training process, but with the original class labels. The cross-entropy loss used in the learnable composition stage is defined as follows:

$$L_{fine}(y, \hat{y}) = - \sum_{i=1}^C y_i \log(p_i) \quad (6)$$

where L_{fine} is the categorical loss function applied during the fine transfer learning stage, C is the number of original classes, y_i is the ground truth label of the i th class, and p_i is the predicted probability of the i th class.

IV. Results

In this work, all experiments were performed using 10-fold cross-validation and the accuracy and F1-score were adopted to evaluate the performance of the models. The accuracy and F1 score are determined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$F1-score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (8)$$

where TP and TN represent the correct predictions by our model for the occurrence of a certain class or not, respectively, while FP and FN are the incorrect predictions of the model for all cases.

Each model was trained with the different transfer learning modes, where stratified cross-validation was used to ensure the same number of samples from each class was balanced in

TABLE 2. Hyperparameters setting for our experimental study.

Experiment Name	Epochs	LR Decay Rate (per 5 epochs)
Xception Shallow	50	0.9x
Xception Finetune	40	0.75x
Xception Deep	30	0.65x
InceptionV3 Shallow	50	0.9x
InceptionV3 Finetune	40	0.75x
InceptionV3 Deep	30	0.65x

each training and test fold. A batch size of 32 was selected, as this has been shown to increase generalizability and training stability with Stochastic Gradient Descent (SGD) in the case of transfer learning and especially for the fine-tuning mode. Moreover, all models were implemented using the TensorFlow 2.1 framework and trained and tested using a NVIDIA 2070 GPU with 8 GB of memory on CentOS 7.8.

Table 2 shows the configuration of the six different experiments for the selected models and the tuning strategies.

For the transfer learning stage of the *DeepCon* model, the Xception and InceptionV3 networks were tested with shallow, finetuning and deep transfer learning modes in the clustered data set with the same hyperparameters and decay rate. The number of epochs for shallow and finetuning modes was changed to 60 and 50, respectively. For evaluation, the labels for each cluster pair were corrected by collapsing them into one of the two pairs for both predictions and ground truth before the evaluation metrics were calculated. For each cross-validation iteration, the DeTraC model was trained on the decomposed dataset before replacing the output layer with 8 neurons. The model was then trained on the original dataset, making sure the same training and testing images were used across models for each iteration. For the second stage, the number of epochs was 7, 10, and 15 for the deep, finetuning, and shallow approaches. The learning rate was also reduced to 0.005, with no learning rate decay being used.

Table 3 compares the results obtained by the Xception and InceptionV3 networks on three different fine-tuning modes. The results of the initial transfer learning experiments show that deep transfer learning by using the pre-trained weights as weight initialization with the Xception model outperformed all the other models. However, despite a notable performance difference between the fine-tuning and shallow-tuning modes in the Xception model, there is no difference in the InceptionV3 model.

The DeTraC method degraded overall system performance across various fine-tuning modes and pre-trained models. On the other hand, our proposed *DeepCon* shows a significant improvement of 1.2% and 0.2%, respectively, for the Xception network with fine-tuning and deep-fine-tuning modes. Tables 4 and Table 5 illustrate the performance difference after applying the DeTraC and *DeepCon* models.

TABLE 3. Results of Transfer Learning Experiments on CNNs Pre-trained Models with several training modes.

Pretrained Model (training mode)	Accuracy (%)	F1-score (%)
Xception (Shallow)	92.8	92.8
Xception (Fine)	94.6	94.6
Xception (Deep)	98.2	98.2
InceptionV3 (Shallow)	91.4	91.5
InceptionV3 (Fine)	91.2	91.3
InceptionV3 (Deep)	97.4	97.4

Eventually, a paired student t test was used to determine whether the results obtained by DeTraC and *DeepCon* were statistically significantly different from the base experiments. The results showed that only the deep DeTraC model was significant with $p=0.027$. For *DeepCon*, the fine-tuning and deep models were significant, with both having $p=0.0002$.

V. CONCLUSION

Transfer learning based on state-of-the-art CNN image classification models has been widely researched in medical applications where data is often limited. This paper presents a novel deep learning model, called *DeepCon* to investigate the effect of divide-and-conquer applied to the original classes on the learning process. *DeepCon* introduces a two-stage transfer learning mechanism, in which knowledge was first transferred using decomposed subclasses of the original classes with coarse transfer learning and then an learned composition with fine transfer learning was employed to cope with the original classes. The proposed model has been validated on a clinically valid dataset of 5000 colorectal cancer (CRC) images. *DeepCon* based on deep-tuning mode offers better performance than the shallow-tuning mode with the highest accuracy achieved by all pre-trained models used in this work, confirming the impact of the composition learned at *DeepCon*.

REFERENCES

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 71, no. 3, pp. 209–249, 2021.
- [2] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, no. 1, pp. 221–248, 2017, pMID: 28301734.
- [3] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, C. I. Sánchez, Ardila, David et al., "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [5] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Detrac: Transfer learning of class decomposed medical images in convolutional neural networks," *IEEE Access*, vol. 8, pp. 74 901–74 913, 2020.
- [6] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 12 2015.
- [7] T. Zhou, Y. Zhou, K. He, C. Gong, J. Yang, H. Fu, and D. Shen, "Cross-level feature aggregation network for polyp segmentation," *Pattern Recognition*, vol. 140, p. 109555, 2023.
- [8] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "Pranet: Parallel reverse attention network for polyp segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, and L. Joskowicz, Eds. Cham: Springer International Publishing, 2020, pp. 263–273.
- [9] M. Shaban, R. Awan, M. M. Fraz, A. Azam, Y. W. Tsang, D. Snead, and N. M. Rajpoot, "Context-aware convolutional neural network for grading of colorectal cancer histology images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 7, pp. 2395–2405, 2020.
- [10] T. Peng, M. Boxberg, W. Weichert, N. Navab, and C. Marr, "Multi-task learning of a deep k-nearest neighbour network for histopathological image classification and retrieval," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, Eds. Cham: Springer International Publishing, 2019, pp. 676–684.
- [11] D. Sarwinda, R. H. Paradisa, A. Bustamam, and P. Anggia, "Deep learning in image classification using residual network (resnet) variants for detection of colorectal cancer," *Procedia Computer Science*, vol. 179, pp. 423–431, 2021, 5th International Conference on Computer Science and Computational Intelligence 2020.
- [12] A. Ben Hamida, M. Devanne, J. Weber, C. Truntzer, V. Derangère, F. Ghiringhelli, G. Forestier, and C. Wemmer, "Deep learning for colon cancer histopathological images analysis," *Computers in Biology and Medicine*, vol. 136, p. 104730, 2021.
- [13] L. Raczowski, M. Mozejko, J. Zambonelli, and E. Szczurek, "Ara: accurate, reliable and active histopathological image classification framework with bayesian deep learning," *Scientific Reports*, vol. 9, 10 2019.
- [14] E. F. Ohata, J. V. S. d. Chagas, G. M. Bezerra, M. M. Hassan, V. H. C. de Albuquerque, and P. P. R. Filho, "A novel transfer learning approach for the classification of histological images of colorectal cancer," *The Journal of Supercomputing*, vol. 77, no. 9, pp. 9494–9519, Sep 2021.
- [15] X. Wang, S. Yang, J. Zhang, M. Wang, J. Zhang, W. Yang, J. Huang, and X. Han, "Transformer-based unsupervised contrastive learning for histopathological image classification," *Medical Image Analysis*, vol. 81, p. 102559, 2022.
- [16] A. Kumar, A. Vishwakarma, and V. Bajaj, "Crcnn-net: Automated framework for classification of colorectal tissue using histopathological images," *Biomedical Signal Processing and Control*, vol. 79, p. 104172, 2023.
- [17] P. Zhou, Y. Cao, M. Li, Y. Ma, C. Chen, X. Gan, J. Wu, X. Lv, and C. Chen, "Hccanet: Histopathological image grading of colorectal cancer using cnn based on multichannel fusion attention mechanism," *Scientific Reports*, vol. 12, no. 1, p. 15103, 2022.
- [18] P. Sabol, P. Šinčák, P. Hartono, P. Kočan, Z. Benetinová, A. Blichárová, Ľudmila Verbóová, E. Štammová, A. Sabolová-Fabianová, and A. Jašková, "Explainable classifier for improving the accountability in decision-making for colorectal cancer diagnosis from histopathological images," *Journal of Biomedical Informatics*, vol. 109, p. 103523, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1532046420301519>
- [19] C. Zhou, Y. Jin, Y. Chen, S. Huang, R. Huang, Y. Wang, Y. Zhao, Y. Chen, L. Guo, and J. Liao, "Histopathology classification and localization of colorectal cancer using global labels by weakly supervised deep learning," *Computerized Medical Imaging and Graphics*, vol. 88, p. 101861, 2021.
- [20] Y. Zhou, S. Graham, N. Alemi Koohbanani, M. Shaban, P. Heng, and N. Rajpoot, "Cgc-net: Cell graph convolutional network for grading of colorectal cancer histology images," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 2019, pp. 388–398.
- [21] H. Chen, C. Li, X. Li, M. M. Rahaman, W. Hu, Y. Li, W. Liu, C. Sun, H. Sun, X. Huang, and M. Grzegorzczek, "Il-mcam: An interactive learning and multi-channel attention mechanism-based weakly supervised colorectal histopathology image classification approach," *Computers in Biology and Medicine*, vol. 143, p. 105265, 2022.
- [22] R. Awan, K. Sirinukunwattana, D. Epstein, S. Jefferyes, U. Qidwai, Z. Aftab, I. Mujeeb, D. Snead, and N. Rajpoot, "Glandular morphometrics for objective grading of colorectal adenocarcinoma histology images," *Scientific Reports*, vol. 7, 12 2017.

TABLE 4. Results of DeTrac and *DeepCon* using the Xception and InceptionV3 networks.

Backbone Network (Training mode)	DeTrac Acc (%)	DeTrac F1-score (%)	<i>DeepCon</i> Acc (%)	<i>DeepCon</i> F1-score (%)
Xception (Shallow)	90.2	90.2	91.8	91.8
Xception (Fine)	94.4	94.4	95.8	95.8
Xception (Deep)	97.6	97.6	98.4	98.4
InceptionV3 (Shallow)	89.2	88.9	91.2	91.2
InceptionV3 (Fine)	89.2	89	91.6	91.5
InceptionV3 (Deep)	94.8	94.8	96.8	96.8

TABLE 5. Performance difference between DeTrac and *DeepCon*.

Backbone Network (training mode)	DeTrac Acc	DeTrac F1-score	<i>DeepCon</i> Acc	<i>DeepCon</i> F1-score
Xception (Shallow)	-2.6	-2.6	-1	-1
Xception (Fine)	-0.2	-0.2	1.2	1.2
Xception (Deep)	-0.6	-0.6	0.2	0.2
InceptionV3 (Shallow)	-2.2	-2.6	-0.2	-0.3
InceptionV3 (Fine)	-2	-2.3	0.4	0.2
InceptionV3 (Deep)	-2.6	-2.6	-0.6	-0.6

- [23] P. Wang, P. Li, Y. Li, J. Xu, and M. Jiang, "Classification of histopathological whole slide images based on multiple weighted semi-supervised domain adaptation," *Biomedical Signal Processing and Control*, vol. 73, p. 103400, 2022.
- [24] J. N. Kather, C.-A. Weis, F. Bianconi, S. M. Melchers, L. R. Schad, T. Gaiser, A. Marx, and F. G. Zöllner, "Multi-class texture analysis in colorectal cancer histology," *Scientific reports*, vol. 6, no. 1, pp. 1–11, 2016.
- [25] A. Vahadane, T. Peng, A. Sethi, S. Albarqouni, L. Wang, M. Baust, K. Steiger, A. M. Schlitter, I. Esposito, and N. Navab, "Structure-preserving color normalization and sparse stain separation for histological images," *IEEE transactions on medical imaging*, vol. 35, no. 8, pp. 1962–1971, 2016.
- [26] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, no. 1, pp. 1–37, Jan 2008. [Online]. Available: <https://doi.org/10.1007/s10115-007-0114-2>
- [27] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.
- [28] X. Xia, C. Xu, and B. Nan, "Inception-v3 for flower classification," in *2017 2nd international conference on image, vision and computing (ICIVC)*. IEEE, 2017, pp. 783–787.