Computers in Biology and Medicine 172 (2024) 108267

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

An effective colorectal polyp classification for histopathological images based on supervised contrastive learning

Sena Busra Yengec-Tasdemir [a,*], Zafer Aydin [b,c], Ebru Akay [d], Serkan Dogan [e], Bulent Yilmaz [f,b]

[a] School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast, BT39DT, United Kingdom
[b] Department of Electrical and Computer Engineering, Abdullah Gul University, Kayseri, 38080, Turkey
[c] Department of Computer Engineering, Abdullah Gul University, Kayseri, 38080, Turkey
[d] Pathology Clinic, Kayseri City Hospital, Kayseri, 38080, Turkey
[e] Gastroenterology Clinic, Kayseri City Hospital, Kayseri, 38080, Turkey
[f] Department of Electrical Engineering, Gulf University for Science and Technology, Mishref, 40005, Kuwait

ARTICLE INFO

MSC:
41A05
41A10
65D05
65D17

Keywords:
Colonic polyp classification
Histopathology image classification
Computer-aided diagnosis
Big transfer
Supervised contrastive learning
Transfer learning

ABSTRACT

Early detection of colon adenomatous polyps is pivotal in reducing colon cancer risk. In this context, accurately distinguishing between adenomatous polyp subtypes, especially tubular and tubulovillous, from hyperplastic variants is crucial. This study introduces a cutting-edge computer-aided diagnosis system optimized for this task. Our system employs advanced Supervised Contrastive learning to ensure precise classification of colon histopathology images. Significantly, we have integrated the Big Transfer model, which has gained prominence for its exemplary adaptability to visual tasks in medical imaging. Our novel approach discerns between in-class and out-of-class images, thereby elevating its discriminatory power for polyp subtypes. We validated our system using two datasets: a specially curated one and the publicly accessible UniToPatho dataset. The results reveal that our model markedly surpasses traditional deep convolutional neural networks, registering classification accuracies of 87.1% and 70.3% for the custom and UniToPatho datasets, respectively. Such results emphasize the transformative potential of our model in polyp classification endeavors.

1. Introduction

Colorectal cancer is a leading cause of cancer-related deaths worldwide [1]. Most colon cancer cases originate from adenomatous polyps, making the early detection and removal of these pre-invasive polyps crucial in reducing mortality rates [2]. Conversely, hyperplastic polyps (HP) are generally considered non-malignant [3]. Therefore, accurately distinguishing adenomatous polyps from HP during the diagnostic process is of utmost importance.

Histopathological analysis and cancer screening require the expertise of pathologists, who face the challenge of differentiating polyp types accurately and efficiently [4]. The increasing demand for accurate polyp differentiation and the growing number of histopathological analysis requests have underscored the need for advanced decision-support systems. To carry out this process faster and more accurately, computer-aided diagnosis (CAD) can be employed. CAD provides valuable assistance in easing labor-intensive work and minimizing the errors associated with traditional approaches. By integrating CAD into the diagnostic workflow, pathologists can benefit from enhanced decision-making capabilities, leading to improved efficiency and effectiveness in the classification of polyps on histopathological images.

Recent advancements in machine learning and deep learning algorithms have sparked significant interest in CAD systems for medical image analysis. Several studies have been conducted on the individual diagnosis of colorectal cancer from histopathological images, including comprehensive studies on colon adenocarcinoma classification [5–17] and colon polyp classification [2,7,8,18–27].

Korbar et al. [19] proposed a framework for classifying colonic polyp types, including HP, sessile serrated polyp (SSP), serrated adenoma (SA), tubular adenoma (TBA), tubulovillous adenoma (TVA), and normal tissues. Their framework utilized the ResNet architecture and achieved a classification accuracy of 93%. Following their previous work, Korbar et al. [18] visualized attention maps on whole slide images (WSIs). Wei et al. [2] used the ResNet architecture to classify polyp types, including TBA, TVA, HP, and SSP, achieving an accuracy of 93.5%. Additionally, in [28] proposed a curriculum learning scheme for distinguishing HP polyps from SSA, achieving an Area Under the Curve (AUC) score of 88.2%. Nasir-Moin et al. [21] developed an AI-augmented diagnostic assistance tool using the ResNet-18 model for polyp type classification. Byeon et al. [24] utilized DenseNet-161

* Corresponding author.
E-mail addresses: sena.yengec@agu.edu.tr, s.yengectasdemir@qub.ac.uk (S.B. Yengec-Tasdemir).

https://doi.org/10.1016/j.compbiomed.2024.108267
Received 9 October 2023; Received in revised form 6 March 2024; Accepted 6 March 2024
Available online 8 March 2024

and EfficientNet-B7 to distinguish polyp types from cancerous and non-specific tissues. In a later work, Wei et al. [29] introduced a Confidence-Aware Label Smoothing scheme for classifying HP and SSP.

Contrastive learning (CLR) approaches have been widely used in histopathology image analysis, including classification [30–33], segmentation [34,35], and stain normalization [36]. These methods leverage contrastive loss to learn discriminative features and enhance the performance of histopathological image analysis tasks.

CLR methods have also been applied to other medical image classification tasks. Vu et al. [37] utilized self-supervised contrastive pre-training for classifying pleural effusion in chest X-ray images. Chen et al. [38] employed an encoder trained with CLR on public datasets to classify Covid-19 in chest X-rays. Zhang et al. [39] pre-trained medical image encoders using paired text data with contrastive loss. Tian et al. [40] proposed a Constrained Contrastive Distribution Learning methodology for detecting anomalies in medical images. Azizi et al. [41] used SimCLR with multiple instances of the same image and compared its performance on various medical image datasets with different baseline models, including Big Transfer. Stacke et al. [42] examined the performance of CLR approaches on histology images using in-domain pre-training and ImageNet pre-training.

Big Transfer (BiT), a transfer learning paradigm, leverages pre-trained models on large datasets to enhance performance in specific tasks [43]. This approach capitalizes on the transferability of learned features to improve classification performance. Mustafa et al. used BiT as a baseline alongside traditional ResNet architectures for medical image classification in mammography, chest X-rays, and dermatology images [44]. Galdran et al. [45] proposed a methodology to improve the classification performance on unbalanced medical image classification tasks by utilizing BiT. Lu et al. [46] combined BiT with SimSiam, demonstrating improved classification performance in skin cancer classification tasks. Shi et al. [47] improved the classification performance in whole-slide image classification of Eosinophilic esophagitis using BiT. Azizi et al. [41] presented a representation learning strategy for medical image classification by employing the weights of BiT as a backbone.

Alongside BiT, Supervised Contrastive Learning (Sup-Con) has emerged as a powerful learning strategy, especially in scenarios with limited labeled data, which is a common challenge in the medical image field. Sup-Con focuses on learning powerful representations by reducing the dependence on extensive labeled datasets and instead utilizing similarities between different but related samples [48]. Thus, facilitates faster learning and more accurate inference. Notably, this study is the first to apply the Sup-Con learning methodology with different state-of-the-art (SOTA) convolutional neural network (CNN) backbone architectures for colonic polyp classification on histopathology images.

This study expands on our previous research [4] by focusing on the multiclass classification of TBA and TVA, aiming to differentiate them from visually similar HP. The primary objective is to improve diagnostic accuracy and support better treatment decisions for these specific polyp types. To achieve this, we propose a CAD system that assists experts in classifying histopathology images, thereby enhancing the diagnostic process and aiding pathologists in their decision-making tasks. To achieve these objectives, we leverage advanced machine learning techniques, including Sup-Con and the BiT model, even in scenarios with limited labeled samples. These approaches significantly enhance the effectiveness of polyp classification on histopathological images. The main contributions of this study can be summarized as follows:

- We develop an improved Sup-Con model and apply it for polyp classification on colon histopathology images. Our method enhances visual task adaptation by utilizing a pre-trained BiT model as the encoder backbone, in contrast to classical Sup-Con models that typically use a conventional ResNet model.
- We comprehensively evaluate the performance of the Sup-Con method with various encoder structures for polyp classification problems on colon histopathology images, which, to the best of our knowledge, has not been extensively explored in previous studies.
- To support our experimental results, we curate a large custom dataset that is used in the experiments. We also provide baseline tests on the dataset using SOTA pre-trained Deep CNN algorithms to enable a fair comparison.
- Furthermore, we investigate the performance of in-domain pre-training for the classification of colonic polyps using a publicly available UniToPatho database.
- To comprehensively evaluate and assess the generalizability of the proposed model, we employ the publicly available UniToPatho database [49]. The proposed Sup-Con model achieves an accuracy of 87% and 70.12% on our custom dataset and UniToPatho, respectively, surpassing the performance of other SOTA methods.
- In the experiments, we compare the performance of the proposed method with traditional Sup-Con on both UniToPatho and our custom datasets.

The structure of the paper is as follows: The second section, *Materials and Methods*, includes explanations about the data collection process, and explains the proposed method and its components. The following section (Section 3) explains the experimental setup and the settings of the tests that are performed. Section 4 demonstrates the comprehensive test results. Finally, the paper is concluded with a discussion section (Section 5).

2. Material and methods

2.1. Data collection

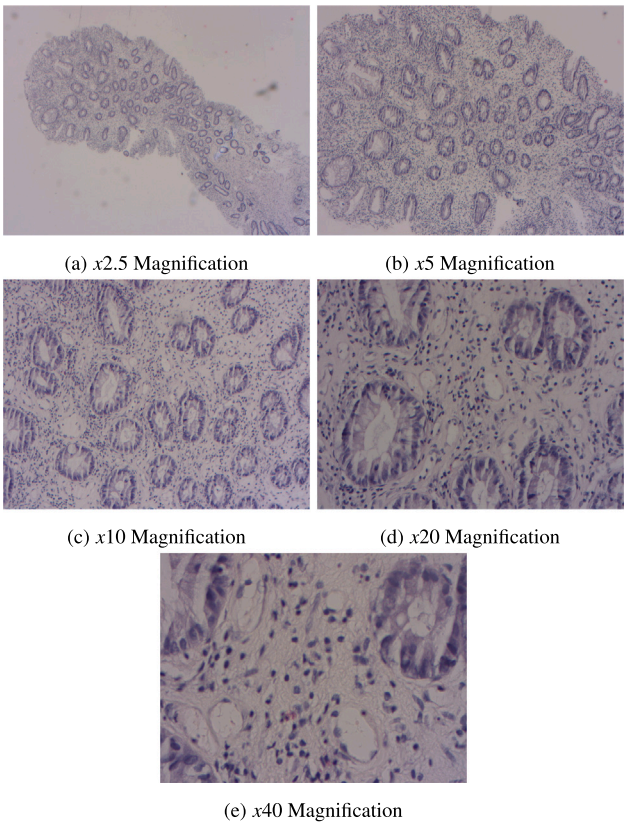
The histological slides used in this study were collected from 184 patients undergoing colorectal cancer screening at Kayseri City Hospital (Kayseri, Turkey) since May 2018.[1] Among the 184 patients, 82 were female and 102 were male. The age range was between 19 and 89, with an average age of 62.

The slides were obtained from patients who had undergone colonoscopy, with each slide representing a specific tissue sample. A total of 359 slides belonged to adenomatous polyps (including TBA and TVA), while 181 slides belonged to HP. The slides were examined under a microscope at five different magnification levels: x2.5, x5, x10, x20, and x40. Examples of tissue samples at each magnification level are shown in Fig. 1.

For each slide and across various magnifications, two expert pathologists performed labeling, and rectangular bounding boxes were manually annotated around the polyps, particularly at the higher magnifications of x2.5 and x5. This detailed annotation process was instrumental in identifying distinct areas of interest within each slide, which we refer to as ‘samples’ in our study. As a result, the total number of samples derived from these slides was significantly larger than the number of slides themselves. Specifically, our dataset comprised 346 samples of TVA, 340 of TBA, and 370 of HP, amounting to a total of 1056 samples. The distribution of these samples, reflecting the diversity and specificity of our study, is presented in Table 1.

To ensure a balanced representation of each class in the training, validation, and test sets, a class-wise split approach was employed. The dataset was initially divided according to class, after which an equal or proportional number of samples from each class were randomly selected for the subsets. Specifically, 718 samples were chosen for training, 157 for validation, and 181 for testing, with care taken to

[1] This study was approved by Kayseri City Hospital Ethics Committee and Erciyes University Clinical Research Ethics Committee.



(a) $x2.5$ Magnification (b) $x5$ Magnification

(c) $x10$ Magnification (d) $x20$ Magnification

(e) $x40$ Magnification

Fig. 1. Representative adenomatous tissue images at increasing magnifications, highlighting key morphological features for pathological classification.

Table 1
Number of samples for each of the classes and Training/Validation/Test sets.

	Hyperplastic	Tubular	Tubulovillous	Total
Training	254	222	242	718
Validation	56	51	50	157
Test	60	67	54	181
Total	370	340	346	1056

avoid any overlap between samples from the same patient. This method ensures that all classes are adequately represented across the subsets, thereby maintaining the diversity and integrity of the dataset.

In addition to this approach, 1657 slides were selectively drawn from the publicly available UniToPatho database, which includes 9536 hematoxylins and eosin-stained (H&E) patches from 292 whole slide images at 20x magnification [49]. This selection targeted slides pertinent to our study's focus – HP, TBA, and TVA – while excluding classes like normal and low-grade polyp classes not central to our research objectives. Moreover, the choice of this subset was also influenced by the practicality of computational resources, as handling the entire UniToPatho dataset would have significantly increased the computational load. Thus, a representative yet manageable subset of slides was chosen, balancing comprehensive data analysis with the available computational capacity. This approach ensured that efficient model training was conducted without compromising the relevance of the data.

2.2. Supervised contrastive learning

CLR techniques have demonstrated significant success in learning powerful representations for visual tasks by comparing positive and negative pairs of samples. However, the application of CLR in a supervised setting can further enhance the performance of inference models. This is achieved through Sup-Con, which bridges the gap between fully supervised learning (SL) and self-supervised learning (SSL) [48].

In Sup-Con, the primary objective is to bring closer the representations of positive samples (i.e., samples belonging to the same class). This goal is achieved by minimizing the contrastive loss function, which is carefully designed to consider both positive and negative pairs. The loss function encourages the model to maximize the distance between different classes in the normalized feature space. By doing so, the model not only learns to identify shared characteristics within the same class but also becomes adept at distinguishing between different classes. The contrastive loss ensures effective separation by calculating the similarity of a query sample with all other samples in the batch, emphasizing the need to differentiate between classes. The Sup-Con loss function (Eq. (1)) is given by:

$$\mathcal{L}(query) = \sum_{\substack{1 \le i \le M \\ y(query)=y(x_i)}} -\log \frac{\exp(query \cdot x_i/\tau)}{\sum_{j=1}^{M} \exp(query \cdot x_j/\tau)} \tag{1}$$

In this equation, the notation i specifically indexes samples that belong to the same class as the query within a batch of M total samples. The term $exp(query \cdot x_i/\tau)$ calculates the exponential similarity score between the query and a positive sample x_i, which is a sample from the same class as the query. This score quantifies the closeness between the query and samples from the same class. On the other hand, j iterates over all M samples in the batch, including both positive and negative samples (the latter being from different classes than the query). The term $exp(query \cdot x_j/\tau)$ computes the query's similarity with each sample in the batch. The function y() denotes the class label. The temperature parameter τ is crucial, as it modulates the Sup-Con's sensitivity, impacting the distinction between classes. By adjusting τ, the model's discriminative capacity is fine-tuned, enhancing its ability to distinguish between different classes. The equation's design, with the summation over i focusing on same-class samples and over j encompassing the entire batch, ensures a comprehensive learning of discriminative features [48,50].

Unlike traditional CLR, Sup-Con utilizes labeled data for positive and negative sample selection, addressing the issues associated with the random sampling of false negatives. The framework comprises an encoder network and a projector network. In the first stage, a batch of inputs undergoes data augmentation and is passed through the encoder network to generate embeddings. These embeddings are then fed into the projector network, which produces normalized outputs. The contrastive loss, defined in Eq. (1), is based on these normalized outputs, promoting similar embeddings for positive pairs and dissimilar ones for negative pairs. In the second stage, a linear classifier is trained on the frozen representations from the first stage, consisting of a fully-connected layer followed by a softmax layer with the target classes.

2.3. Big transfer

BiT is a transfer learning paradigm that incorporates architectural modifications to enhance the performance of both upstream and downstream tasks [43]. During pre-training, BiT utilizes upstream components such as scale, group normalization (GN), and weight standardization (WS). These components contribute to the learning of powerful representations in a supervised setting. Downstream components are then employed during fine-tuning for specific tasks.

BiT models, which are based on ResNet-v2 architectures, are designed in various sizes to cater to distinct computational and application requirements. Among these, the BiT-M model, representing the medium variant in the BiT family, strikes a balance between model complexity and computational efficiency. This medium variant is particularly relevant for scenarios that demand robust performance but are constrained by computational resources.

A key architectural modification in BiT is the replacement of batch normalization (BN) with GN. This decision is supported by the advantages GN offers over BN, particularly its robustness to variations in batch sizes. Batch Normalization, commonly used in neural networks, normalizes the input of a mini-batch by adjusting and scaling the activations. However, its effectiveness diminishes with smaller batch sizes, often encountered in resource-constrained environments. On the other hand, GN, which normalizes inputs across groups of channels within each layer, offers robustness to variations in batch sizes. This feature is particularly advantageous in scenarios where computational resources limit batch sizes [51].

In addition to group normalization, BiT models are further enhanced by integrating weight standardization (WS). WS optimizes the training process by standardizing the weights in the convolutional layers, a technique that involves normalizing these weights by subtracting their mean and dividing by the standard deviation. These architectural advancements enable BiT to demonstrate satisfactory performance in tasks that demand significant domain adaptation [43]. The adaptability and flexibility of BiT are also validated through its performance in the Visual Task Adaptation Benchmark (VTAB), which assesses the model across a wide range of diverse visual tasks. Furthermore, these improvements in domain adaptability are also crucial in more complex areas such as medical image classification, a topic analyzed in Section 1.

2.4. Proposed framework

Previous studies on polyp classification using histology images predominantly employ transfer learning methods. However, the domain mismatch between medical and natural images has been shown to hinder the effectiveness of transfer learning [52]. On the other hand, recent works have demonstrated that the BiT method improves the performance of transfer learning approaches in medical image classification tasks due to its remarkable domain adaptation capabilities [25,41,44–47]. Additionally, the loss function of Sup-Con accelerates the learning of hard negatives and hard positives, which is particularly beneficial for our histology database which contains challenging samples.

We have designed the proposed framework by combining the BiT and Sup-Con frameworks, as illustrated in Fig. 2. This figure depicts the comprehensive structure of our framework, including the two-stage Sup-Con approach and the integration of BiT as the encoder.

During the first stage, as depicted in Fig. 2, we train the encoder for 100 epochs using the Sup-Con loss with a batch size of 16. In this phase, we subject a batch of inputs to data augmentation before passing them through the encoder network to generate embeddings. These embeddings are then channeled into the projector network, resulting in normalized outputs. The contrastive loss (Eq. (1)), derived from these outputs, facilitates the comparison of a query with a set of samples from the same class [48,50]. Upon completion of the first stage, we progress to the second stage with the learned frozen representations. This stage encompasses a series of layers: a dropout layer, a fully connected layer equipped with L2 kernel regularizers, followed by another dropout layer, and a final softmax layer. This softmax layer is responsible for predicting the class of a given sample, such as “Hyperplastic Polyp”. To optimize this stage, the stochastic gradient descent (SGD) optimizer is employed, operating at a learning rate of 0.001 and a momentum of 0.9. We further refine the learning rate using the adaptive momentum optimization algorithm. The chosen loss function for this stage is categorical cross entropy, complemented by label smoothing set at a value of 0.1, to strike a balance in prediction confidence. The optimal settings described above were determined through extensive experiments and comparisons, which will be detailed in Section 3.

In contrast to the traditional Sup-Con model, our framework takes a different approach by integrating the BiT architecture as the encoder. This choice marks a significant departure from the traditional Sup-Con’s reliance on the simpler architecture of ResNet-50-V2 as the encoder. Our model capitalizes on BiT’s strengths, particularly in handling limited data scenarios, to enhance the learning process. This collective enhancements position our framework as a more tailored solution compared to the traditional Sup-Con model.

3. Experimental setup

To evaluate the performance of the proposed method, we compare it against commonly used ImageNet pre-trained Deep CNN models. We also employ these pre-trained models as encoders for the Sup-Con framework and compare the performance of SL and Sup-Con approaches. As explained in Section 2.4, the proposed Sup-Con framework is a modified version of the traditional Sup-Con. During the experiments, we also compare the performance of the proposed method with traditional Sup-Con. Hyperparameter optimization is conducted to explore the model’s performance under different settings, including learning rate (0.0001, 0.0005, 0.001, 0.005, 0.01), optimizer (ADAM, SGD, RMSprop, Adagrad), and temperature value (0.03, 0.05, 0.08, 0.1).

To assess the impact of in-domain pre-training, we altered the pre-training set and conducted ablation tests using standard performance metrics. These metrics include class-based F1-score, precision, recall, and overall accuracy, as defined in Eq. (2). In these equations, T, F, P, and N denote True, False, Positive, and Negative predictions, respectively. Specifically, FP_i represents the instances where the model incorrectly predicted class i, and FN_i denotes the cases where the model failed to identify class i, predicting other classes instead.

$$\begin{aligned} F_{1i} &: 2 \times \frac{Precision_i \times Recall_i}{Precision_i + Recall_i} \\ Precision_i &: \frac{TP_i}{TP_i + FP_i} \\ Recall_i &: \frac{TP_i}{TP_i + FN_i} \\ \text{Accuracy} &: \frac{\sum_c TP_c}{\text{Number of all Samples}} \end{aligned} \tag{2}$$

Here, F_{1i} is the F1-score for class i, calculated as the harmonic mean of precision and recall. Precision for class i ($Precision_i$) is the ratio of true positive predictions to the total predicted as class i, while recall ($Recall_i$) measures the ratio of true positive predictions to the actual number of samples in class i. The overall accuracy is determined by the proportion of true positive predictions across all classes (c) to the total number of samples.

Additionally, the weighted average for a metric M is computed as shown in Eq. (3). Here, M_i represents the metric value for class i, and C_i is the count of samples in class i. The weighted average of metric M is calculated using the formula:

$$\text{Weighted Average of metric } M : \frac{\sum_i M_i C_i}{\sum_i C_i} \tag{3}$$

Finally, to verify the generalizability of our proposed models, we conducted tests using the publicly available UniToPatho database. However, it is crucial to underline that this phase of testing does not represent a traditional external validation, as the models were re-estimated for compatibility with this specific dataset. While this approach provides valuable insights into the adaptability and performance of the models, it is more accurately described as an extended evaluation rather than a definitive external validation.

The models used for the ablation study are obtained from TensorFlow, including BiT-M (trained on ImageNet-21k), DenseNet-201 (trained on ImageNet), Inception-V3 (trained on ImageNet), ResnNet-50 (trained on ImageNet), ResNetV2-50 (trained on ImageNet), InceptionResNet-V2 (trained on ImageNet), and Xception (trained on ImageNet). All these models are pre-trained on natural images, and fine-tuning is applied to adapt them to the histological dataset. Fine-tuning is necessary because the pre-trained models are trained on a different domain, and transferring their knowledge to histology images requires fine-tuning on the specific dataset. The pre-trained models are used as encoders during the first stage of the Sup-Con framework, and they are trained using Supervised Contrastive loss.

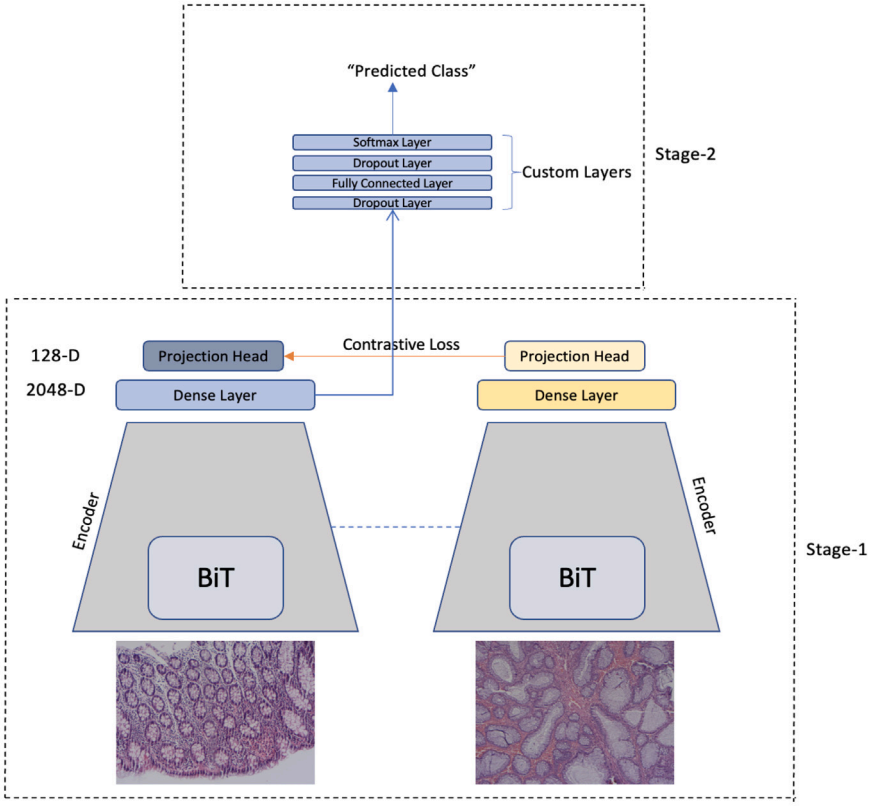


Fig. 2. Illustration of the proposed framework: This diagram outlines the architecture and workflow of the proposed supervised contrastive approach, which integrates the BiT encoder and demonstrates the two-stage training process for enhanced polyp classification.

All experiments are performed on the Google Colab platform, which provides 52 GB of RAM and NVIDIA Tesla K80, NVIDIA Tesla T4, and NVIDIA Tesla P100 GPU accelerators. The proposed experiments are implemented using Python v3.7.13 with the TensorFlow v2.8.0 framework.

The source code of the proposed algorithm and the experimental setup can be accessed online [53].

4. Results

Following the experimental setup, this section delves into the outcomes of our comprehensive testing and analysis. The efficacy of the proposed Sup-Con methodology, alongside various baseline models, is systematically evaluated in the polyp classification task. Our findings are encapsulated in Table 2 and visually summarized in Fig. 3, revealing the performance of these models in terms of Overall Accuracy, Precision, Recall, and F1 scores across different classes. Additionally, a weighted average of the performance metrics for each model is included in Table 2 to offer a holistic assessment.

In the analysis of Table 2, the BiT model emerges as the top performer in the supervised setting, followed by Inception-ResNet and DenseNet-201. Notably, the table also contrasts the performance of different model backbones, including the ResNet-50-v2 encoder as implemented in our proposed Sup-Con settings, alongside its traditional Sup-Con variant. A more detailed distinction between these two implementations is provided in Section 2.4.

Among the baseline models, the proposed method performs best when BiT-M is used as the encoder on our custom dataset. In this case, it achieves the highest accuracy, weighted average precision, weighted average recall, and weighted average F1-score, with values of 86.2%, 86.3%, 86.2%, and 86.1% respectively. Conversely, when ResNet-50 is used as the encoder, the performance of the model is relatively poor, with an accuracy, weighted average precision, weighted average recall, and weighted average F1-score of 75.7%, 75.7%, 75.6% and 75.6 respectively. The difference in performance between BiT-M and ResNet-50 can be attributed to the fact that BiT-M is an improved version of ResNet-50 specifically designed to ease domain adaptation. In addition to the proposed Sup-Con model, the performance of the Traditional Sup-Con setup, which uses the ResNet-50-v2 encoder, is also noteworthy. As seen in Table 2, the Traditional Sup-Con model shows a commendable improvement over the standard SL approach, achieving an overall accuracy of 75.2% compared to 70.2% in the supervised setting. This enhancement indicates the effectiveness of even the traditional Sup-Con approach in enhancing model performance, particularly in differentiating between the complex subtypes of polyps.

In addition to comparing the performance of different encoders, we also evaluated the impact of pre-training sets on the custom dataset. We employed three different settings: utilizing the publicly available UniToPatho dataset as the pre-training dataset of the encoders, fine-tuning ImageNet-trained models on our custom dataset, and fine-tuning ImageNet pre-trained models with the UniToPatho and custom database. It is crucial to highlight that the selection of models for comparison in Table 3 was meticulously based on their performance rankings as established in Table 2. Specifically, the top three performing models – DenseNet, Inception-ResNet, and BiT-M – were chosen from Table 2 for a focused analysis. The results, presented in Table 3, demonstrate varying impacts of the pre-training sets on these models. The ImageNet pre-trained DenseNet and Inception-ResNet-v2 models exhibit a 1.4% improvement in accuracy when fine-tuned on the UniToPatho database. In contrast, the BiT-M model shows a 3.7% decrease in performance under the same experimental conditions. This decrease in performance may be attributed to the fact that UniToPatho samples are cropped from only one magnification of a Whole Slide Image (WSI), while our custom dataset contains WSIs of four different magnification levels.

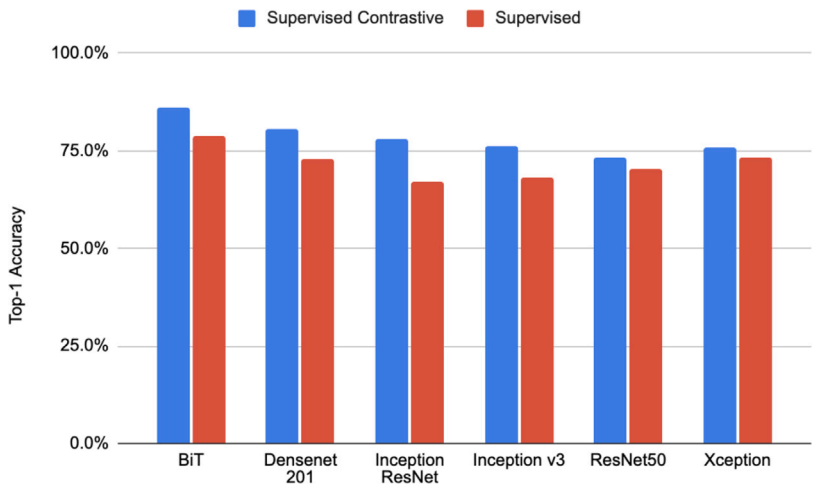


Fig. 3. Accuracy comparison of the supervised learning and supervised contrastive learning with state-of-the-art classifiers.

Table 2
Comparison of Supervised (SL) and Supervised Contrastive Learning (Sup-Con) performance measurements for various Deep CNN models. Acc., Pre., Rec., and F1 stand for Accuracy, Precision, Recall, and F1 Score, respectively. HP, TBA, and TVA represent Hyperplastic Polyps, Tubular Adenomas, and Tubulovillous Adenomas, respectively. The best results are highlighted in bold.

Model	SL				Sup-Con			
	Overall Acc. (%)	Pre. (%)	Rec. (%)	F1 (%)	Overall Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)
BiT-M	78.9	78.9	79.2	78.9	**86.2**	86.3	86.2	86.1
HP		78.2	80.4	79.3		83.6	93.3	88.2
TBA		81.0	75.2	78.0		87.9	86.6	87.2
TVA		77.0	82.7	79.8		87.5	77.7	82.4
DenseNet-201	73.0	73.0	73.0	73.0	**80.7**	80.2	81.6	80.4
HP		72.7	74.1	73.4		87.9	80.3	83.9
TBA		73.8	72.0	72.9		80.8	75.7	78.1
TVA		72.3	73.2	72.7		70.9	90.3	79.4
Inception ResNet	67.0	67.5	67.8	67.2	**77.9**	78.1	78.0	77.9
HP		58.3	70.8	64.0		74.8	81.6	78.1
TBA		74.0	60.6	66.7		80.8	74.3	77.4
TVA		69.6	73.3	71.4		78.5	78.5	78.5
Inception-v3	68.3	68.4	68.5	68.3	**76.2**	76.3	76.3	76.2
HP		66.0	70.7	68.3		73.8	79.8	76.7
TBA		65.1	69.7	67.3		78.9	72.6	75.6
TVA		75.0	64.7	69.5		76.0	76.9	76.4
ResNet-50	70.2	70.5	70.4	70.2	**75.7**	75.7	75.6	75.6
HP		66.0	75.3	70.4		76.1	83.6	79.7
TBA		71.4	71.4	71.4		77.4	72.7	75.0
TVA		74.4	63.7	68.6		73.1	70.4	71.7
ResNet-50-v2: Traditional Sup-Con	70.2	70.5	70.4	70.2	**75.2**	75.1	76.0	75.2
HP		66.0	75.3	70.4		73.8	77.5	75.6
TBA		71.4	71.4	71.4		81.7	69.1	74.9
TVA		74.4	63.7	68.6		68.4	83.1	75.0
Xception	73.5	73.6	73.4	73.4	**75.9**	75.9	75.8	75.8
HP		71.7	76.8	74.2		74.8	80.0	77.3
TBA		73.3	74.0	73.7		76.9	72.7	74.8
TVA		76.0	69.0	72.3		76.0	75.0	75.5

Furthermore, the unbalanced distribution of samples for certain polyp types in the UniToPatho dataset, such as TVA, TBA, and HP, may have a significant impact on the performance of BiT-M. On the other hand, the ImageNet pre-trained BiT-M achieves the highest accuracy of 86.2%.

An in-depth examination of the performance of various models across polyp subtypes, as detailed in Tables 2 and 3, provides valuable insights. We observed notable variations in model performance depending on the polyp subtype, which can largely be ascribed to the inherent architectural characteristics of each model and their interaction with the unique features of these subtypes. For example, BiT-M consistently exhibited superior performance across all subtypes, including the more complex TVA and TBA. Its higher accuracy and precision in handling these subtypes highlight its robustness.

In contrast, models like DenseNet-201 and Inception ResNet displayed varying efficacy levels across different polyp subtypes. DenseNet-201, for instance, was more effective in classifying HP subtype, potentially due to its architectural alignment with the specific patterns of HP polyps. On the other hand, Inception ResNet's performance fluctuated across subtypes, influenced by its feature extraction capabilities.

Moreover, the analysis of different pre-training sets, as shown in Table 3, revealed that the source of pre-training data significantly affects model performance for specific polyp subtypes. Models pre-trained on ImageNet and then fine-tuned on the UniToPatho database tended to excel in classifying HP, whereas those only pre-trained on ImageNet were slightly more inclined towards adenomatous polyps. This underscores the critical role of pre-training data in enhancing a model's adaptability and generalization to various subtypes. Thus, when developing models for polyp classification, it is essential to consider both

Table 3
Assessment of model performance pre-trained on different datasets with the adapted Sup-Con approach. Acc. stands for Accuracy, Pre. for Precision, Rec. for Recall, F1 for F1-Score, HP for Hyperplastic polyps, TBA for Tubular adenomas, and TVA for Tubulovillous adenomas. The best results are highlighted in bold.

Model	UniToPatho				ImageNet				ImageNet + UniToPatho			
	Overall Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)	Overall Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)	Overall Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)
DenseNet	63.1	62.5	64.4	62.8	80.7	80.2	81.6	80.4	82.1	82.1	82.7	82.3
HP		73.8	58.1	65.0		87.9	80.3	83.9		83.2	80.2	81.7
TBA		57.7	66.7	61.9		80.8	75.7	78.1		78.8	76.6	77.7
TVA		55.7	68.8	61.5		70.9	90.3	79.4		84.8	93.1	88.7
Inception-ResNet	72.4	72.0	73.0	72.0	77.6	77.8	78.1	77.8	80.7	80.3	81.1	80.5
HP		82.2	69.3	75.2		75.9	81.0	78.3		88.8	77.9	83.0
TBA		60.6	74.1	66.7		78.8	70.1	74.2		72.1	81.5	76.5
TVA		74.7	75.6	75.2		78.5	84.9	81.6		81.0	84.2	82.6
BiT-M	**79.6**	78.9	82.3	79.1	**86.2**	86.3	86.2	86.1	**82.4**	82.2	82.3	82.2
HP		88.7	74.0	80.7		83.6	93.3	88.2		86.0	86.0	86.0
TBA		84.6	78.6	81.5		87.9	86.6	87.2		81.7	78.7	80.2
TVA		60.8	96.0	74.4		87.5	77.8	82.4		78.5	82.7	80.5

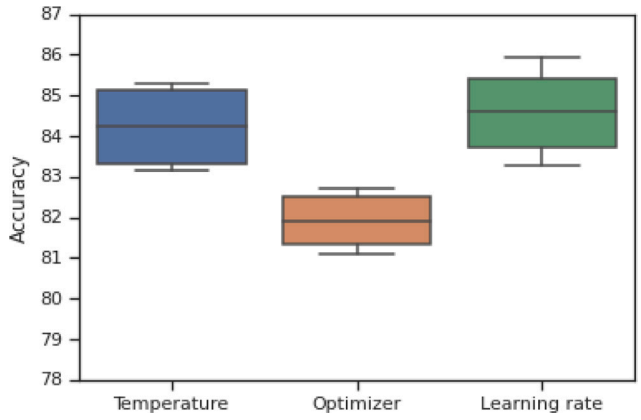


Fig. 4. Hyperparameter stability analysis of the proposed method.

the architecture and the pre-training data source, as these factors profoundly influence the model's effectiveness across different subtypes.

Additionally, we performed hyperparameter optimization on the proposed method. Various learning rates, optimization methods, and temperature values for the Sup-Con were tested. The values of each hyperparameter can be found in Section 3. Fig. 4 shows a standard box plot depicting the Top-1 accuracy changes for different hyperparameters: learning rate, optimization method, and temperature value. It can be observed that the proposed method exhibits low variance in accuracy across different hyperparameters. The best accuracy of **87.1**% for the proposed method was achieved with a learning rate of 0.0005, Adam optimizer, and a temperature value of 0.05.

Overall, the experimental results demonstrate that the proposed Sup-Con methodology, especially when combined with the BiT-M encoder, significantly improves the performance of polyp classification on colon histopathology images.

4.1. Generalization test

To evaluate the generalization ability of the model, we conducted experiments using both the UniToPatho dataset and the custom dataset. A comparison was made between the proposed Sup-Con methodology and traditional Sup-Con on these datasets. Separate fine-tuning was performed for each dataset, and the results are presented in Table 4. The receiver operating characteristic (ROC) curves of our proposed method on the custom database and the UniToPatho dataset can be found in Figs. 5 and 6, respectively. Additionally, the classification confusion matrices for each dataset are shown in Fig. 7.

Analyzing the results in Table 4, we observed that the proposed method achieved a 6.1% improvement in accuracy on the UniToPatho dataset compared to traditional SL approaches. Similarly, on the custom database, our method demonstrated a significant 11% improvement in accuracy. These results highlight the effectiveness of the Sup-Con methodology in enhancing the model's generalization ability across different datasets.

Furthermore, our proposed method surpasses the performance of other methods in the literature. Previous studies achieved accuracies of 64.3% [25] and 66.6% [26] on the UniToPatho database, whereas our proposed method achieved higher accuracies on both the UniToPatho dataset and the custom database.

The ROC curves in Figs. 5 and 6 depict the trade-off between true positive rate and false positive rate for our proposed method on the custom database and the UniToPatho dataset, respectively. The curves indicate the superior discriminative ability of our method, as it achieves higher true positive rates while maintaining lower false positive rates.

The classification confusion matrices in Fig. 7 provide additional insights into the performance of the proposed method. The matrices illustrate the distribution of predicted classes compared to the ground truth for each dataset. The higher values along the diagonal indicate accurate classification, while off-diagonal values represent misclassifications. Our proposed method demonstrates improved classification performance with fewer misclassifications.

These results demonstrate the strong generalization ability of the proposed Sup-Con methodology, as it consistently outperforms traditional SL approaches. Achieving higher accuracies on both the UniToPatho dataset and the custom database not only highlights its robust performance but also suggests that our method attains SOTA results in this domain.

5. Discussion

The relevance of CLR in histopathology image analysis has been underscored by numerous studies exploring tasks such as classification, segmentation, and stain normalization. However, the application of Sup-Con for colonic polyp classification on histology images remains a relatively unexplored territory, setting the stage for the contribution of our study. In comparison with existing methods for colonic polyp classification, which predominantly rely on conventional SL strategies employing Deep CNN models, our method offers significant improvements. Notable examples of these conventional methods include the work of Korbar et al. [19], Byeon et al. [24], and Iizuka et al. [9], who used ResNet-50 variants, DenseNet, and Inception-V3 respectively, and achieved considerable accuracy on their respective datasets. However, our approach has shown to deliver superior performance in distinguishing hard-positive and hard-negative samples.

Our experimental setup involved comprehensive comparisons with conventional models like ResNet50, EfficientNet, and Inception-v3. The performance gap observed between our custom dataset and the aforementioned studies could be attributed to variations in domain distribution. However, when incorporated into the Sup-Con approach,

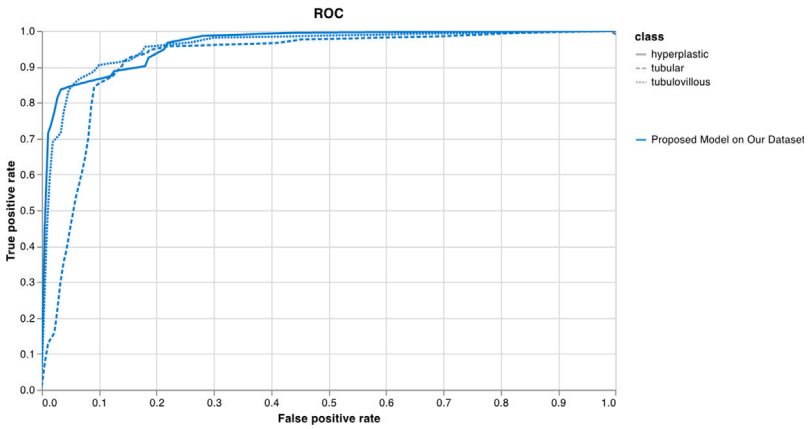


Fig. 5. ROC curve of the proposed method on custom dataset. It attains substantially high performance.

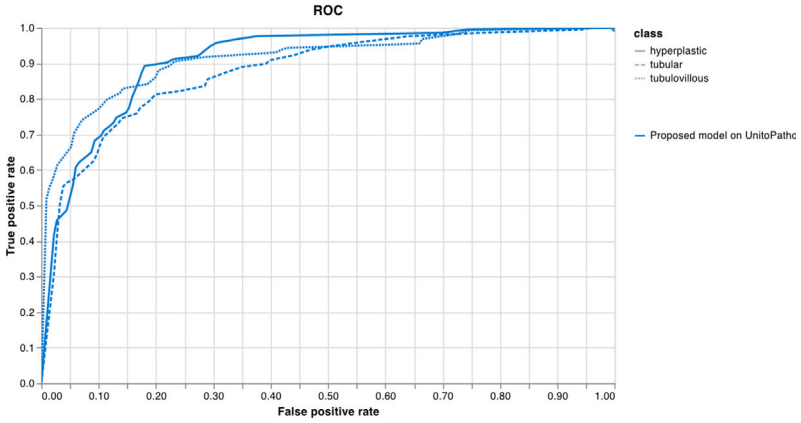
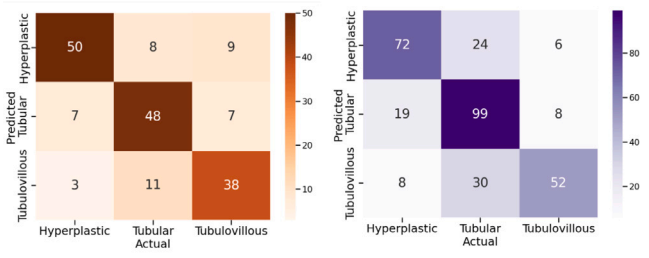


Fig. 6. ROC curve of the proposed method on public UnitoPatho dataset. This shows a high generalization ability of the model.

Table 4
Generalization performance of the proposed model and Traditional Sup-Con on the custom collected dataset and UniToPatho. Acc. stands for Accuracy, Pre. for Precision, Rec. for Recall, F1 for F1-Score, HP for Hyperplastic polyps, TBA for Tubular adenomas, and TVA for Tubulovillous adenomas. The best results are highlighted in bold.

Model/Per class	Custom collected dataset				UnitoPatho			
	Overall Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)	Overall Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)
Conventional Sup-Con	75.2	75.1	76.0	75.2	64.0	69.9	64.0	65.2
HP		73.8	77.5	75.6		88.2	90.9	89.6
TBA		81.7	69.1	74.9		74.0	50.3	59.9
TVA		68.4	83.1	75.0		32.7	55.2	41.1
Proposed method	**86.2**	86.3	86.2	86.1	**70.1**	71.8	70.1	70.3
HP		83.6	93.3	88.2		70.6	72.7	71.6
TBA		87.9	86.6	87.2		78.6	64.7	71.0
TVA		87.5	77.8	82.4		57.8	78.8	66.7



(a) *CustomDataset* (b) *UnitoPatho*

Fig. 7. Confusion matrix of the proposed method on our custom collected database and UnitoPatho. The vertical and horizontal axes are the predicted and the true labels, respectively.

these models significantly improved in performance, with our method utilizing BiT as a backbone achieving the highest accuracy.

It is important to note that many existing models are primarily evaluated on specific custom datasets, which may limit their generalization to diverse datasets encountered in real-world applications. Our model, however, has shown remarkable generalization capability across different datasets. Specifically, it has excelled not only on our custom-collected dataset but also on the publicly available UniToPatho dataset, achieving an impressive accuracy of 70.3%. When contrasted with other methods reported in the literature that garnered accuracies of 64.3% and 66.6% on the UniToPatho dataset [25,26], the generalization capability of our proposed model becomes even more apparent. These results underscore the potential of our model to address real-life challenges in colon adenomatous polyp classification, paving the way for its wider applicability in clinical settings. These results provide compelling evidence regarding the remarkable generalization capability of

our proposed model across different datasets, highlighting its potential applicability in real-life scenarios. This highlights the potential applicability of our method in real-world scenarios.

While our approach has demonstrated robust performance and generalizability, one potential drawback lies in its heavy computational requirement due to the use of the BiT model and Sup-Con framework. This could potentially limit its deployment in low-resource settings where computational power is a constraint. However, with ongoing advancements in model optimization techniques and hardware efficiency, this challenge could be mitigated in future implementations. As future work, the development of a comprehensive plan that includes incorporating other publicly available datasets for pre-training purposes, and conducting longitudinal follow-up studies of patients could enhance our method's robustness and applicability. These insights could significantly contribute to the development of more effective screening and diagnostic methods for colon adenomatous polyps and ultimately colorectal cancer.

6. Conclusion

This study introduces a novel CAD system that leverages enhanced Sup-Con for accurate polyp classification in colon histopathology images. The proposed method, incorporating the BiT model, effectively contrasts in-class from out-of-class images, significantly enhancing the model's capability to differentiate between polyp types. The evaluation of our model on both a custom-collected dataset and the UniToPatho dataset attests to its superior performance over conventional deep CNN models, thereby underscoring its potential for practical application. Furthermore, our analysis of different pre-training settings highlights the impressive visual task adaptation of the BiT model and indicates the potential benefits of in-domain pre-training. In conclusion, our research not only brings to light the potential of Sup-Con in polyp classification but also emphasizes the importance of domain-specific pre-training, thereby providing a roadmap for future advancements in the field.

CRediT authorship contribution statement

Sena Busra Yengec-Tasdemir: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Formal analysis, Data curation, Conceptualization. **Zafer Aydin:** Writing – review & editing. **Ebru Akay:** Data curation. **Serkan Dogan:** Data curation. **Bulent Yilmaz:** Writing – review & editing, Supervision, Resources, Project administration.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Sena Busra Yengec-Tasdemir reports financial support was provided by Scientific and Technological Research Council of Türkiye (TUBITAK)

Acknowledgments

This work was supported by the Scientific and Technological Research Council of Türkiye (TUBITAK) under Grant 120E204. The authors would like to thank Serdal Sadet Ozcan for her valuable contribution to detailed labeling of the dataset.

Ethics statement

This study was conducted ethically based on the standards of the Declaration of Helsinki, and it has been approved by Erciyes University, the Clinical Research Ethics Committee. We ensured that all research procedures adhered to the accepted ethical standards. The committee's approval on October 9, 2019, with the reference number 96 681 246, confirmed that our experimental protocols were thoroughly evaluated and endorsed by the appropriate institutional body.

Before participation, all subjects participating in this study were comprehensively informed about the research objectives, the procedures to be followed, potential risks, and the benefits that could arise from their participation. Following this briefing, informed voluntary consent was obtained from each participant. This meticulous process was designed to ensure that participants were fully aware of their rights and the nature of the research being conducted, thereby preserving their autonomy and well-being throughout the study.

In line with our commitment to ethical conduct, we have taken strict measures to ensure the rights to privacy of human subjects at all times. All the participants' personal data that were to be collected in the course of the study were treated with maximum confidentiality and used for this research alone. Identifiable information had been appropriately either anonymized or pseudonymized, and strict access was limited to the research team only, hence protecting all participants from the breach of either their privacy or integrity.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA: Cancer J. Clin. 71 (3) (2021) 209–249, http://dx.doi.org/10.3322/CAAC.21660, URL: https://pubmed.ncbi.nlm.nih.gov/33538338/.

[2] J.W. Wei, A.A. Suriawinata, L.J. Vaickus, B. Ren, X. Liu, M. Lisovsky, N. Tomita, B. Abdollahi, A.S. Kim, D.C. Snover, J.A. Baron, E.L. Barry, S. Hassanpour, Evaluation of a deep neural network for automated classification of colorectal polyps on histopathologic slides, JAMA Netw. Open 3 (4) (2020) e203398, http://dx.doi.org/10.1001/JAMANETWORKOPEN.2020.3398.

[3] E. Torlakovic, E. Skovlund, D.C. Snover, G. Torlakovic, J.M. Nesland, Morphologic reappraisal of serrated colorectal polyps, Am. J. Surg. Pathol. 27 (1) (2003) 65–81.

[4] S.B. Yengec-Tasdemir, Z. Aydin, E. Akay, S. Dogan, B. Yilmaz, Improved classification of colorectal polyps on histopathological images with ensemble learning and stain normalization, Comput. Methods Programs Biomed. 232 (2023) 107441, http://dx.doi.org/10.1016/j.cmpb.2023.107441, URL: https://www.sciencedirect.com/science/article/pii/S0169260723001074.

[5] D. Bychkov, N. Linder, R. Turkki, S. Nordling, P.E. Kovanen, C. Verrill, M. Walliander, M. Lundin, C. Haglund, J. Lundin, Deep learning based tissue analysis predicts outcome in colorectal cancer, Sci. Rep. 8 (1) (2018) 1–11, http://dx.doi.org/10.1038/s41598-018-21758-3, URL: https://www.nature.com/articles/s41598-018-21758-3.

[6] M. Bilal, Y.W. Tsang, M. Ali, S. Graham, E. Hero, N. Wahab, K. Dodd, H. Sahota, W. Lu, M. Jahanifar, A. Robinson, A. Azam, K. Benes, M. Nimir, A. Bhalerao, H. Eldaly, S.E.A. Raza, K. Gopalakrishnan, F. Minhas, D. Snead, N. Rajpoot, AI based pre-screening of large bowel cancer via weakly supervised learning of colorectal biopsy histology images, medRxiv (2022) http://dx.doi.org/10.1101/2022.02.28.22271565, 2022.02.28.22271565.

[7] P. Gupta, Y. Huang, P.K. Sahoo, J.F. You, S.F. Chiang, D.D. Onthoni, Y.J. Chern, K.Y. Chao, J.M. Chiang, C.Y. Yeh, W.S. Tsai, Colon tissues classification and localization in whole slide images using deep learning, Diagnostics (Basel Switz.) 11 (8) (2021) http://dx.doi.org/10.3390/DIAGNOSTICS11081398, URL: https://pubmed.ncbi.nlm.nih.gov/34441332/.

[8] C. Ho, Z. Zhao, X.F. Chen, J. Sauer, S.A. Saraf, R. Jialdasani, K. Taghipour, A. Sathe, L.Y. Khor, K.H. Lim, W.Q. Leow, A promising deep learning-assistive algorithm for histopathological screening of colorectal cancer, Sci. Rep. 12 (1) (2022) 1–9, http://dx.doi.org/10.1038/s41598-022-06264-x, URL: https://www.nature.com/articles/s41598-022-06264-x.

[9] O. Iizuka, F. Kanavati, K. Kato, M. Rambeau, K. Arihiro, M. Tsuneki, Deep learning models for histopathological classification of gastric and colonic epithelial tumours, Sci. Rep. 10 (1) (2020) 1–11.

[10] A. Kallipolitis, K. Revelos, I. Maglogiannis, Ensembling efficientnets for the classification and interpretation of histopathology images, Algorithms 14 (10) (2021) http://dx.doi.org/10.3390/a14100278.

[11] D. Sarwinda, R.H. Paradisa, A. Bustamam, P. Anggia, Deep learning in image classification using residual network (ResNet) variants for detection of colorectal cancer, in: Procedia Computer Science, Elsevier, 2021, pp. 423–431, http://dx.doi.org/10.1016/j.procs.2021.01.025.

[12] T.E. Tavolara, M.K.K. Niazi, V. Arole, W. Chen, W. Frankel, M.N. Gurcan, A modular cGAN classification framework: Application to colorectal tumor detection, Sci. Rep. 9 (1) (2019) 1–8, http://dx.doi.org/10.1038/s41598-019-55257-w.

[13] E. Terradillos, C.L. Saratxaga, S. Mattana, R. Cicchi, F.S. Pavone, N. Andraka, B.J. Glover, N. Arbide, J. Velasco, M.C. Etxezarraga, A. Picon, Analysis on the characterization of multiphoton microscopy images for malignant neoplastic colon lesion detection under deep learning methods, J. Pathol. Inform. 12 (1) (2021) 27, http://dx.doi.org/10.4103/jpi.jpi_113_20, URL:/pmc/articles/PMC8359734/ /pmc/articles/PMC8359734/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8359734/.

[14] M. Tsuneki, F. Kanavati, Deep learning models for poorly differentiated colorectal adenocarcinoma classification in whole slide images using transfer learning, Diagnostics 11 (11) (2021) 2074.

[15] M. Yildirim, A. Cinar, Classification with respect to colon adenocarcinoma and colon benign tissue of colon histopathological images with a new CNN MA_ColonNET, Int. J. Imaging Syst. Technol. 32 (1) (2022) 155–162, http://dx.doi.org/10.1002/ima.22623.

[16] C. Zhou, Y. Jin, Y. Chen, S. Huang, R. Huang, Y. Wang, Y. Zhao, Y. Chen, L. Guo, J. Liao, Histopathology classification and localization of colorectal cancer using global labels by weakly supervised deep learning, Comput. Med. Imaging Graph. 88 (2021) 101861, http://dx.doi.org/10.1016/J.COMPMEDIMAG.2021.101861.

[17] D. Albashish, Ensemble of adapted convolutional neural networks (CNN) methods for classifying colon histopathological images, PeerJ Comput. Sci. 8 (2022) e1031.

[18] B. Korbar, A.M. Olofson, A.P. Miraflor, C.M. Nicka, M.A. Suriawinata, L. Torresani, A.A. Suriawinata, S. Hassanpour, Looking under the hood: Deep neural network visualization to interpret whole-slide image analysis outcomes for colorectal polyps, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Vol. 2017-July, 2017, pp. 821–827, http://dx.doi.org/10.1109/CVPRW.2017.114.

[19] B. Korbar, A. Olofson, A. Miraflor, C. Nicka, M. Suriawinata, L. Torresani, A. Suriawinata, S. Hassanpour, Deep learning for classification of colorectal polyps on whole-slide images, J. Pathol. Inform. 8 (1) (2017) http://dx.doi.org/10.4103/JPI.JPI_34_17, URL: https://pubmed.ncbi.nlm.nih.gov/28828201/. arXiv:1703.01550.

[20] Z. Song, C. Yu, S. Zou, W. Wang, Y. Huang, X. Ding, J. Liu, L. Shao, J. Yuan, X. Gou, W. Jin, Z. Wang, X. Chen, H. Chen, C. Liu, G. Xu, Z. Sun, C. Ku, Y. Zhang, X. Dong, S. Wang, W. Xu, N. Lv, H. Shi, Automatic deep learning-based colorectal adenoma detection system and its similarities with pathologists, BMJ Open 10 (9) (2020) e036423, http://dx.doi.org/10.1136/BMJOPEN-2019-036423, URL: https://bmjopen.bmj.com/content/10/9/e036423 https://bmjopen.bmj.com/content/10/9/e036423.abstract.

[21] M. Nasir-Moin, A.A. Suriawinata, B. Ren, X. Liu, D.J. Robertson, S. Bagchi, N. Tomita, J.W. Wei, T.A. Mackenzie, J.R. Rees, S. Hassanpour, Evaluation of an artificial intelligence-augmented digital system for histologic classification of colorectal polyps, JAMA Netw. Open 4 (11) (2021) 1–12, http://dx.doi.org/10.1001/jamanetworkopen.2021.35271.

[22] D. Perlo, E. Tartaglione, L. Bertero, P. Cassoni, M. Grangetto, Dysplasia grading of colorectal polyps through CNN analysis of WSI, 2021, arXiv preprint arXiv:2102.05498.

[23] J. Wei, A. Suriawinata, B. Ren, X. Liu, M. Lisovsky, L. Vaickus, C. Brown, M. Baker, M. Nasir-Moin, N. Tomita, L. Torresani, J. Wei, S. Hassanpour, Learn like a pathologist: Curriculum learning by annotator agreement for histopathology image classification, 2021, CoRR.

[24] S.-j. Byeon, J. Park, Y.A. Cho, B.-J. Cho, Automated histological classification for digital pathology images of colonoscopy specimen via deep learning, Sci. Rep. 12 (1) (2022) 1–8.

[25] X. Wang, S. Yang, J. Zhang, M. Wang, J. Zhang, W. Yang, J. Huang, X. Han, Transformer-based unsupervised contrastive learning for histopathological image classification, Med. Image Anal. 81 (2022) 102559.

[26] X. Wang, Y. Du, S. Yang, J. Zhang, M. Wang, J. Zhang, W. Yang, J. Huang, X. Han, RetCCL: clustering-guided contrastive learning for whole-slide image retrieval, Med. Image Anal. 83 (2023) 102645.

[27] W.S. Liew, T.B. Tang, C.-H. Lin, C.-K. Lu, Automatic colonic polyp detection using integration of modified deep residual convolutional neural network and ensemble learning approaches, Comput. Methods Programs Biomed. 206 (2021) 106114, http://dx.doi.org/10.1016/j.cmpb.2021.106114, URL: https://www.sciencedirect.com/science/article/pii/S0169260721001899.

[28] J. Wei, A. Suriawinata, B. Ren, X. Liu, M. Lisovsky, L. Vaickus, C. Brown, M. Baker, M. Nasir-Moin, N. Tomita, L. Torresani, J. Wei, S. Hassanpour, Learn like a pathologist: Curriculum learning by annotator agreement for histopathology image classification, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, WACV, 2021.

[29] J. Wei, L. Torresani, J. Wei, S. Hassanpour, Calibrating histopathology image classifiers using label smoothing, 2022, arXiv preprint arXiv:2201.11866.

[30] K. Stacke, G. Eilertsen, J. Unger, C. Lundstrom, Measuring domain shift for deep learning in histopathology, IEEE J. Biomed. Health Inf. 25 (2) (2021) 325–336, http://dx.doi.org/10.1109/JBHI.2020.3032060.

[31] M.Y. Lu, R.J. Chen, J. Wang, D. Dillon, F. Mahmood, Semi-supervised histology classification using deep multiple instance learning and contrastive predictive coding, 2019, arXiv preprint arXiv:1910.10825.

[32] O. Ciga, T. Xu, A.L. Martel, Self supervised contrastive learning for digital histopathology, Mach. Learn. Appl. 7 (2022) 100198.

[33] P. Yang, Z. Hong, X. Yin, C. Zhu, R. Jiang, Self-supervised visual representation learning for histopathological images, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24, Springer, 2021, pp. 47–57.

[34] H. Wu, Z. Wang, Y. Song, L. Yang, J. Qin, Cross-patch dense contrastive learning for semi-supervised segmentation of cellular nuclei in histopathologic images, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 11666–11675.

[35] N. Boserup, R. Selvan, Efficient self-supervision using patch-based contrastive learning for histopathology image segmentation, 2022, arXiv preprint arXiv:2208.10779.

[36] J. Ke, Y. Shen, X. Liang, D. Shen, Contrastive learning based stain normalization across multiple tumor in histopathology, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24, Springer, 2021, pp. 571–580.

[37] Y.N.T. Vu, R. Wang, N. Balachandar, C. Liu, A.Y. Ng, P. Rajpurkar, Medaug: Contrastive learning leveraging patient metadata improves representations for chest x-ray interpretation, in: Machine Learning for Healthcare Conference, PMLR, 2021, pp. 755–769.

[38] X. Chen, L. Yao, T. Zhou, J. Dong, Y. Zhang, Momentum contrastive learning for few-shot COVID-19 diagnosis from chest CT images, Pattern Recognit. 113 (2021) 107826.

[39] Y. Zhang, H. Jiang, Y. Miura, C.D. Manning, C.P. Langlotz, Contrastive learning of medical visual representations from paired images and text, 2020, arXiv preprint arXiv:2010.00747.

[40] Y. Tian, G. Pang, F. Liu, Y. Chen, S.H. Shin, J.W. Verjans, R. Singh, G. Carneiro, Constrained contrastive distribution learning for unsupervised anomaly detection and localisation in medical images, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24, Springer, 2021, pp. 128–140.

[41] S. Azizi, B. Mustafa, F. Ryan, Z. Beaver, J. Freyberg, J. Deaton, A. Loh, A. Karthikesalingam, S. Kornblith, T. Chen, et al., Big self-supervised models advance medical image classification, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 3478–3488.

[42] K. Stacke, J. Unger, C. Lundström, G. Eilertsen, Learning representations with contrastive self-supervised learning for histopathology applications, 2021, arXiv preprint arXiv:2112.05760.

[43] A. Kolesnikov, L. Beyer, X. Zhai, J. Puigcerver, J. Yung, S. Gelly, N. Houlsby, Big transfer (bit): General visual representation learning, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16, Springer, 2020, pp. 491–507.

[44] B. Mustafa, A. Loh, J. Freyberg, P. MacWilliams, M. Wilson, S.M. McKinney, M. Sieniek, J. Winkens, Y. Liu, P. Bui, et al., Supervised transfer learning at scale for medical imaging, 2021, arXiv preprint arXiv:2101.05913.

[45] A. Galdran, G. Carneiro, M.A. González Ballester, Balanced-mixup for highly imbalanced medical image classification, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2021, pp. 323–333.

[46] Y. Lu, A. Jha, R. Deng, Y. Huo, Contrastive learning meets transfer learning: a case study in medical image analysis, in: Medical Imaging 2022: Computer-Aided Diagnosis, Vol. 12033, SPIE, 2022, pp. 715–722.

[47] Y. Shi, Q. Liu, J. Xu, Z. Asad, C. Cui, H. Correa, Y. Choksi, G. Hiremath, Y. Huo, Eosinophilic esophagitis multi-label feature recognition on whole slide imaging using transfer learning, in: Medical Imaging 2022: Digital and Computational Pathology, Vol. 12039, SPIE, 2022, pp. 277–284.

[48] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, D. Krishnan, Supervised contrastive learning, Adv. Neural Inf. Process. Syst. 33 (2020) 18661–18673.

[49] C.A. Barbano, D. Perlo, E. Tartaglione, A. Fiandrotti, L. Bertero, P. Cassoni, M. Grangetto, UniToPatho, a labeled histopathological dataset for colorectal polyps classification and adenoma dysplasia grading, 2021, http://dx.doi.org/10.48550/ARXIV.2101.09991, URL: https://arxiv.org/abs/2101.09991.

[50] A. Islam, C.-F.R. Chen, R. Panda, L. Karlinsky, R. Radke, R. Feris, A broad study on the transferability of visual representations with contrastive learning, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 8845–8855.

[51] Y. Wu, K. He, Group normalization, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 3–19.

[52] M. Raghu, C. Zhang, J. Kleinberg, S. Bengio, Transfusion: Understanding transfer learning for medical imaging, Adv. Neural Inf. Process. Syst. 32 (2019).

[53] S. Yengec-Tasdemir, Implementation of an effective colorectal polyp classification for histopathological images based on supervised contrastive learning, 2023, https://github.com/senabyengec/Supervised-Contrastive-for-Classification-of-Colorectal-Polyps-on-Histopathological-Images. [Online; accessed 23-May-2023].