Transfer Learning with Similar Domain Shift Classification of Oral Squamous Cell Carcinoma

Beatriz Matias Santana Maia
Department of Informatics
Federal University of Espírito Santo
Vitória, Brazil
beatriz.maia@edu.ufes.br

Vitor Bonella
Department of Informatics
Federal University of Espírito Santo
Vitória, Brazil
vitor.bonella@edu.ufes.br

Higor David Oliveira
Department of Informatics
Federal University of Espírito Santo
Vitória, Brazil
higor.d.oliveira@edu.ufes.br

***Abstract*—Oral cancer in 2019 accounted for more than 370,000 cases worldwide, with more than 170,000 cases resulting in death [27]. Among the different types of oral cancer, oral squamous cell carcinoma (OSCC) represents around 90% of malignant neoplasms that affect the oral cavity. Over the past few decades, we have seen how artificial intelligence (AI) can impact multiple areas across different types of tasks. With advances in deep learning (DL), the impact of informatics on medicine has also been growing, especially in aiding diagnosis. However, there are some limitations in CNN image classification. A high quality dataset in healthcare take years to be curated, depending on-demand patient with the diseases in questions, and, specially in the field of pathology, with the equipment necessary. In order to classify them or annotate information regarding these images also demands experts analysis. To face these adversities, several techniques have been developed. Considering the difficulties of curating a new, robust dataset, one such technique is transfer learning, which focuses on transferring knowledge between similar domains in scenarios where there is limited data or computing power. However, dealing with this introduces new challenges as well. The dataset being used for transferred learning will contain different features due to the fact the data was collected in a distinct setting, occurring a domain shift from the inference images. Using Densenet121, we analyze the impact of applying transfer learning, training on using P-NDB-UFES and Rahman et at. datasets to the problem of OSCC image classification. Using ImageNet transferred learning weights, Densenet121 was able to reach 89.39% of accuracy and 89.79% f1 score on Rahman et al. When using P-NDB-UFES, a similar domain, for transferred learning, the model reached 90.97% accuracy and 90.89% f1 score.**

***Index Terms*—Oral squamous cell carcinoma, convolution neural network, CNN, domain shift, transfer learning**

I. Introduction

Health is a fundamental human concern. Lethal diseases can have substantial impact to the individual, but also to those around them. Cancer is an example of this, affecting not only those directly impacted by it, but also family and friends [33]. Currently, it's one of leading cause of death in the world, being responsible for almost 10 millions deaths only in 2020 [32]. Alarming data also shows that cancer is starting earlier, with more cases before 50's increasing [30].

Oral cancer in 2019 accounted for more than 370,000 cases worldwide, with more than 170,000 cases resulting in death [27]. Among the different types of oral cancer, oral squamous cell carcinoma (OSCC) represents around 90% of malignant neoplasms that affect the oral cavity. Diagnosis of OSCC can be difficult as detection itself proves to be challenging when dysplasia and carcinoma can be difficult to recognize and differentiate on histopathological imaging [1, 6, 21]. It's predicted that by 2040, OSCC will increase by 40%, with a growth in mortality [26].

Over the past few decades, we have seen how artificial intelligence (AI) can impact multiple areas across different types of tasks. With advances in deep learning (DL), the impact of informatics on medicine has also been growing, especially in aiding diagnosis. As an example, Bejnordi et al. [3] applied different convolutional neural network models (CNN) to identify and classify metastases in histopathological images of breast cancer, obtaining results comparable to those of experienced pathologists. Companies that use AI have also been growing [2], with applications ranging from cancer detection to automated assistants, making doctors' daily lives easier.

However, there are some limitations in CNN image classification, specifically in the medical field. Some of the challenges faced when exploring digital pathology are: the scarcity of categorized images; unbalanced distribution of data sets; the lack of indication of the region of interest; the high dimension of histopathological images; the absence of clinical information such as family history and symptoms (metadata) in several datasets; and the need for faster hardware for training [8, 28]. A high quality dataset in healthcare take years to be curated, depending on-demand patient with the diseases in questions [4], and, specially in the field of pathology, with the equipment necessary. In order to classify them or annotate information regarding these images also demands experts analysis. Images such as whole slide images (WSIs) require large storage due to the size of pathology images. [8, 28].

To face these adversities, several techniques have been developed. Considering the difficulties of curating a new, robust dataset, one such technique is transfer learning, which focuses on transferring knowledge between similar domains in scenarios where there is limited data or computing power [17, 31]. However, dealing with this introduces new challenges as well. The dataset being used for transferred learning will contain different features due to the fact the data was collected in a distinct setting, occurring a domain shift from the

inference images [23].

Another technique is classification based on *patches*, which consists of using segmented parts of an image called *patches*, reducing the dimensionality of the images inserted into the [10] models without losing information with image compression, increasing the dataset size.

The main contribution of this paper is:

1) Using the best model and configuration in the paper in "Transformers, convolutional neural networks, and few-shot learning for classification of histopathological images of oral cancer" [15], we analyze the impact of applying transfer learning using P-NDB-UFES and Rahman et at. [19] datasets to the problem of OSCC image classification.

The remainder of this paper is organized in the following order: section 2 presents the datasets used in this research as well as an introduction to CNNs, Domain Shifts and Transfer Learning. In Section 3, we present the experimental setting and results obtained. Section discusses the results, analyzing the impact of transferred learning applied to OSCC, with concluding remarks in Section 5.

II. Material and methods

A. Oral cancer datasets

1) P-NDB-UFES: The NDB-UFES dataset [4] was obtained in cooperation with the Oral Pathological Anatomy Service of the Dentistry Course (also known as SAP Bucal) of the Federal University of Espírito Santo. Patients attending the extension project Oral Diagnosis Center. Slides were selected to capture histopathological images using a camera attached, with the 10x and 40x objectives, saved in the size of 2048 x 1536 pixels. Images from the NDB-UFES dataset were cropped into random patches and labeled by pathologists from SAP Bucal, resulting inthe P-NDB-UFES dataset, that contains only images categorized in one of the following 3 classes: OSCC, dysplasia and without dysplasia. There are in total 1126 images of OSCC (29.92%), 1930 images of dysplasia (51.29%), and 707 images without dysplasia (18.79%) in the dataset.

2) Rahman et al. Dataset: The histopathological image database from Rahman et al.[19] contains normal images of the oral cavity epithelium and images of oral squamous cell carcinoma (OSCC), divided into two sets: images magnified at 100x and images magnified at 400x. The database is highly unbalanced, with OSCC images comprising 73.31% of the data. In the first set, there are 89 images without OSCC and 439 images of OSCC. In the second set, there are 201 images without OSCC and 495 images of OSCC.

B. Convolutional Neural Network

Convolutional Neural Networks (CNNs), introduced by LeCun et al. [13], are designed to handle data with a grid-like structure, such as images, and have demonstrated significant success in computer vision applications. A typical CNN architecture consists of three primary types of layers: convolutional layers, pooling layers, and fully connected layers.

The convolutional layer is fundamental for feature extraction from input data. It consists of learnable filters, also known as kernels, which slide over the input data to produce feature maps, also referred to as activation maps. After convolution, an activation function is typically applied to introduce non-linearity, suppress negative values, and highlight important features. The Rectified Linear Unit (ReLU) is a widely used activation function in this context [7].

Pooling layers follow the convolutional layers to reduce the spatial dimensions of the feature maps while retaining their essential characteristics. This is achieved through a pooling operation, commonly downsampling. The two prevalent pooling methods are maximum pooling and average pooling. Maximum pooling selects the highest value from each region, whereas average pooling computes the average value [7].

Fully connected layers, also known as Multilayer Perceptrons (MLPs), are standard components in CNNs, typically serving as classifiers. They take the features extracted by previous layers as input. Since feature maps are tensors and fully connected layers require vector inputs, a flattening operation is performed to convert the feature maps into vectors. Alternatively, a CNN can output the flattened vector directly as an embedding, a technique often used in few-shot learning scenarios [7]. Figure 1 illustrates a CNN architecture comprising convolutional layers, pooling layers, and a multi-layer feedforward network.

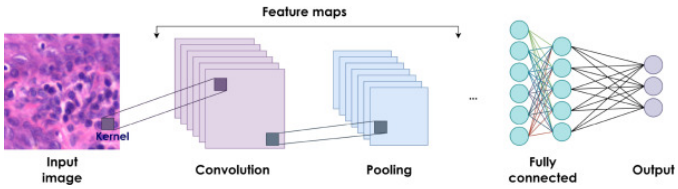

Fig. 1. A histopathology patch image being processed by a convolutional neural network (CNN), the following steps are delineated: Initially, the image undergoes transformation through convolutional layers, which utilize kernels to detect various features. This is followed by pooling layers, which perform dimensionality reduction while preserving the essential features extracted. As the process progresses, feature maps are generated, encapsulating the significant characteristics identified by the convolutional and pooling operations. Subsequently, the fully connected layers, structured as a multi-layer feedforward network, undertake the classification task. This culminates in the final output layer, which provides the classification results based on the processed features.

Over the years, various architectural elements such as residual blocks and dense blocks have been introduced to enhance layer communication. Consequently, several advanced architectures like ResNet [9], MobileNet [11], DenseNet [12], EfficientNet [25], and VGG [20] have been developed, achieving remarkable results in image classification.

Due to time and hardware constraints, DenseNet [12] was selected for this study, as it demonstrated superior performance in recent research [15] on the dataset under investigation.

C. Domain Shift

In their study, Torraba and Efros [29] highlighted the detrimental impact of domain shift on performance across various

public datasets sharing identical classes. This phenomenon isn't limited to these datasets alone but extends to histopathological image datasets as well, influenced by differences in lighting, contrast due staining methods, variations in technical equipment used for image capture, and other factors.[22] Strategies like image augmentation and normalization are commonly employed to mitigate these effects and align datasets. However, their effectiveness in improving model performance across all datasets remains uncertain.[22]

A detailed study on the model's sensitivity to learning and on the data correlation across different datasets, along with data collection to measure and understand domain shift between datasets, proves necessary for the image classification problem.[24]

Beyond dataset adjustments, recent approaches such as deep domain adaptation are being explored. These methods aim to fundamentally alter the learning process of convolutional neural network (CNN)-based models, fostering better transferable representations amidst datasets affected by domain shift.[16]

D. Transfer Learning

In the realm of convolutional neural networks (CNNs), transfer learning operates on the principle that a pre-trained network, originally designed for a specific task, can effectively be repurposed to address a related but different task [34]. This approach leverages the learned features from related domains, thereby reducing the requirement for extensive labeled data and computational resources [14]. The adaptation process typically involves fine-tuning the network on a new dataset or task by adjusting its weights and parameters.

One effective strategy in transfer learning is "pretrained feature extraction," where the pretrained convolutional network is employed without modifying its weights. Instead, it serves solely to extract features that are subsequently used by classifiers.

An alternative approach, termed "partial fine-tuning," involves freezing the convolutional layers of the pre-trained model while retraining only the fully connected layers. This strategy is advantageous when the pre-trained model has already learned generic, low-level features that are applicable to the new task. By maintaining the convolutional layers fixed, practitioners conserve significant computational resources and time, all while benefiting from the established knowledge embedded within the network. This method is particularly suitable when the lower-level features do not require substantial modification for the new task at hand.

"Full fine-tuning" represents a comprehensive strategy where both convolutional and fully connected layers of a pre-trained model are adapted to a new task using their existing weights. This method allows practitioners to customize the network architecture closely to the specific requirements of the new task, albeit at a higher computational cost compared to freezing layers. Fine-tuning is particularly beneficial when the lower-level features of the pre-trained model need adjustment or when the new task differs significantly from its original purpose.

It's important to recognize that fine-tuning requires a substantial amount of data to effectively adjust the model's parameters without overfitting. Therefore, the availability of ample and diverse datasets is crucial for successful fine-tuning endeavors. This ensures that the adapted model can generalize well to new inputs and perform optimally in practical applications.

In this study, full fine-tuning was exclusively employed due to the significant disparity between the previous dataset and the target task. The initial model was trained on a dataset vastly dissimilar to the current application domain, necessitating a thorough adaptation of both convolutional and fully connected layers.

III. Results

A. Experimental Setting

The experimental setup is divided into three parts. The first part involves training the DenseNet121 convolutional architecture [12] using pre-trained weights from ImageNet dataset [5], utilizing data from the P-NDB-UFES train dataset and the Rahman train dataset to generate two distinct models. In the second part, the test sets from each dataset are evaluated on both models to investigate potential domain shift between the datasets, as they belong to the same domain. In the third part, the best model, initially trained on the P-NDB-UFES data, is further trained with the Rahman train dataset. After this training, the new model is tested on the Rahman test set to verify if there was any transfer learning between the two datasets. Figure 2 shows this process.

The selection of the best model is performed using stratified 5-fold cross-validation for all tests. The model is subsequently trained using 4 folds and validated with the remaining 1 fold. The model with the best F1 macro score is chosen for the test.

During the experiments, it was observed that the model was having difficulties learning with P-NDB-UFES dataset, as shown in Figures 3 and 4. Upon further inspection, the model was unable to distinguish dysplasia images from the other classes, having specifically hard time with carcinoma images. This is to be expected due to the fact that dysplasia isn't the same as carcinoma, however it can be a transitional phase. Given the results shown in Figure 5, the problem was simplified for classification of images with carcinoma or no dysplasia.

The training process utilizes the Weighted Cross-Entropy loss function and the Stochastic Gradient Descent (SGD) optimizer to adjust the model's parameters. Additionally, a ReduceLROnPlateau scheduler [18] is employed, which decreases the initial learning rate of 10^{-3} by 10% every 5 epochs without improvement.

B. Experimental Results

Table I presents metrics regarding the training results for each experiment in terms of mean and standard deviation of each fold. Both model used ImageNet weights for transferred learning and reached accuracy above 85%. The models had a higher precision in comparison to recall, however the

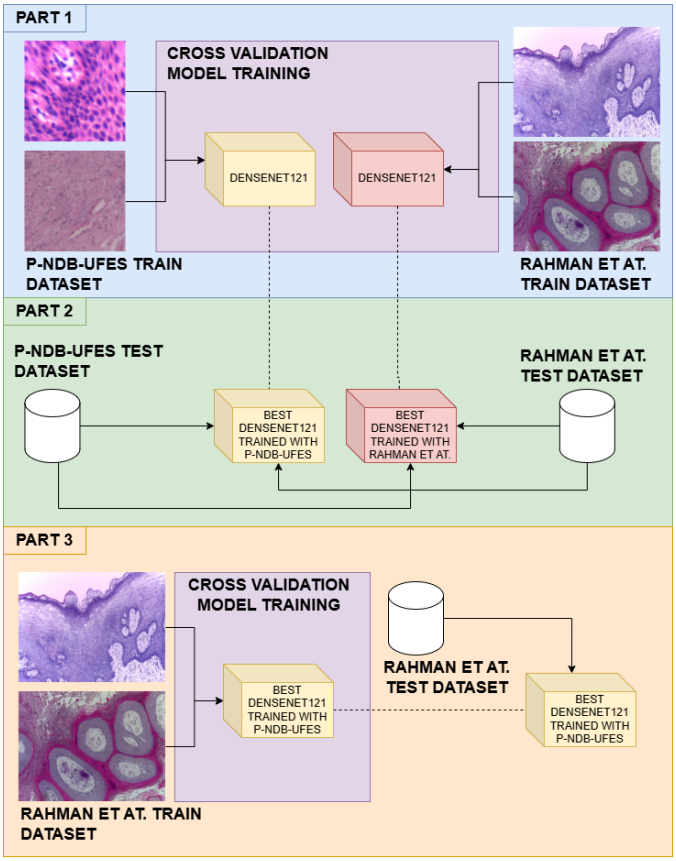

Fig. 2. Research and investigation process.

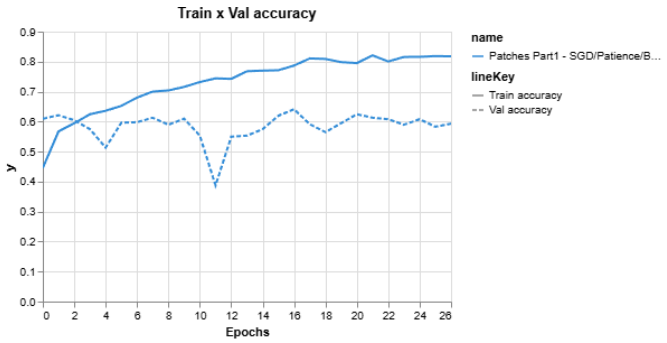

Fig. 3. Accuracy curves obtained of on the folds for train and validation set. The graphs show the model over-fitting to the training model, being unable to obtain better results in validation set.

difference was small. This could indicate that the models are predicting less cases as carcinoma. This is not ideal considering the scope of the problem and the lethal impact of classifying cancerous patients as non-cancerous.

Table II presents metrics regarding the various tests assembled. When using the same domain for training and testing, the results were similar to results obtained during Part 1. However, when using different datasets, applying a domain shift, the trained models had a significant impact in results. The model trained with Rahman et al. dataset and tested with the P-NDB-UFES dataset had a drop of 17% in accuracy compared to

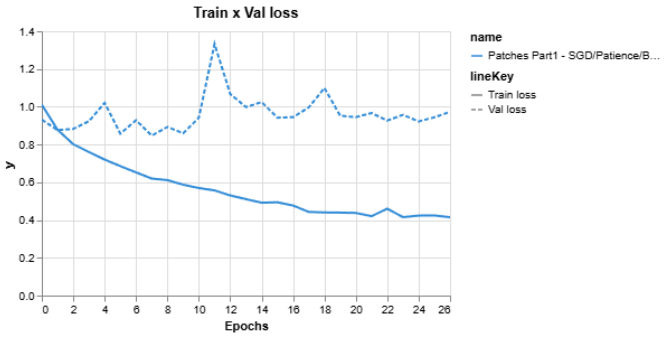

Fig. 4. Loss curves obtained of on the folds for train and validation set. The graphs show the model over-fitting to the training model, being unable to obtain better results in validation set.

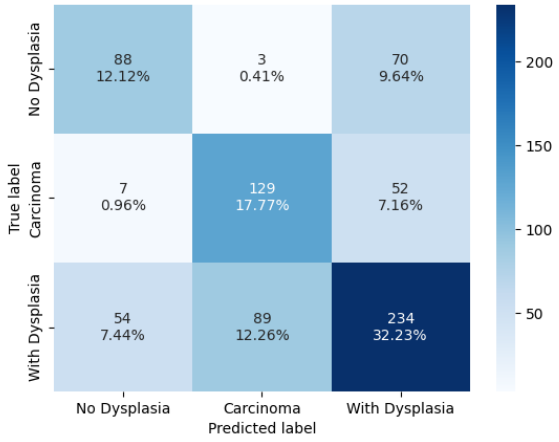

Fig. 5. Confusion matrix of P-NDB-UFES dataset with 3 classes: no dysplasia, carcinoma and with dysplasia. Model had difficulties distinguishing with dysplasia from other classes.

the model trained with P-NDB-UFES. The drop is even more significant when comparing the Rahman et al. test dataset, with a drop of 42%. A performance drop is expected given that the data is in a different domain, despite it being the same problem, a binary OSCC classification. The test with the different domain in this step of the experiment was done to measure how much of a drop would there be without any training beforehand.

Part 3 consisted of training and testing transferred learning in a similar domain. In Table III, the results obtained using Rahman et al. dataset with the P-NDB-UFES weights shown no improvement compared to results in Table I. The best model

Part 1 - Training Results (%)				
Dataset	**Accuracy**	**F1**	**Recall**	**Precision**
P-NDB-UFES	89.64 ± 1.83	89.59 ± 1.87	89.65 ± 1.83	90.21 ± 1.73
Rahman et al.[19]	87.35 ± 1.83	87.97 ± 1.63	87.35 ± 1.83	90.16 ± 0.79

TABLE I
AVERAGE RESULTS FOR MODELS TRAINED PER FOLD IN EXPERIMENTS OF PART 1.

Part 2 - Test Results (%)					
Model Weights	**Test set**	**Accuracy**	**F1**	**Recall**	**Precision**
P-NDB-UFES	P-NDB-UFES	90.97	90.89	90.97	91.07
Rahman et al.[19]	P-NDB-UFES	73.87	73.97	73.87	74.15
Rahman et al.[19]	Rahman et al.[19]	89.39	89.79	89.39	91.05
P-NDB-UFES	Rahman et al.[19]	47.35	51.26	47.35	61.31

TABLE II
TEST RESULTS WITH BEST MODEL WEIGHTS FROM EXPERIMENTS OF PART 1. WEIGHTS WERE LOADED AND TESTED ON SETS WITHOUT ANY ADDITIONAL TRAINING.

Part 3 - Training Results (%)				
Dataset	**Accuracy**	**F1**	**Recall**	**Precision**
Rahman et al.[19]	86.20 ± 2.01	86.96 ± 1.81	86.20 ± 2.01	89.25 ± 1.20

TABLE III
AVERAGE RESULTS FOR MODELS TRAINED PER FOLD IN EXPERIMENTS OF PART 3, USING TRANSFERRED LEARNING WITH P-NDB-UFES MODEL WEIGHTS.

results presented in Table IV were very similar to the results in Table II. This could indicate that the usage of the P-NDB-UFES weights didn't impact training any different than the usage of ImageNet weights.

IV. DISCUSSION

The results of this study highlight the effectiveness and limitations of using transfer learning for the classification of oral squamous cell carcinoma (OSCC) using histopathological images. The findings are divided into several critical observations.

The use of transfer learning, particularly with the DenseNet121 architecture, demonstrated considerable success. When leveraging ImageNet weights, the model achieved high accuracy and F1 scores with both the Rahman et al. and P-NDB-UFES datasets. This confirms the utility of pre-trained models in medical image analysis, where annotated datasets are often limited.

A significant drop in performance was observed when models trained on one dataset were tested on another, indicating a substantial domain shift. Specifically, the model trained on the P-NDB-UFES dataset and tested on the Rahman et al. dataset exhibited a sharp decline in accuracy and F1 scores. This drop underscores the challenge of domain shift in transfer learning applications, where variations in image acquisition techniques, staining methods, and patient demographics can impact model performance.

The experiment showed that the DenseNet121 model, when fine-tuned on a target dataset, could generalize well within the same domain. However, its generalization capacity across different domains without additional training was limited. This finding suggests that while transfer learning can jump-start the model training process, fine-tuning on the target domain is crucial for achieving optimal performance.

Part 3 - Test Results (%)				
Dataset	**Accuracy**	**F1**	**Recall**	**Precision**
Rahman et al.[19]	88.57	89.05	88.57	90.58

TABLE IV
TEST RESULTS WITH BEST MODEL WEIGHTS FROM EXPERIMENTS OF PART 3. WEIGHTS WERE LOADED AND TESTED ON RAHMAN TEST SET WITHOUT ANY ADDITIONAL TRAINING.

A full fine-tuning strategy was employed due to the significant difference between the pre-trained dataset (ImageNet) and the target datasets (P-NDB-UFES and Rahman et al.). The results indicate that full fine-tuning was necessary to adapt the model effectively to the specific characteristics of histopathological images of OSCC. This strategy allowed the model to leverage pre-learned low-level features while adapting high-level features to the target task.

Inherent challenges were highlighted in the field of digital pathology, including the high dimensionality of images, variability in staining techniques, and the need for expert annotations. These factors contribute to the complexity of developing robust and generalizable models for medical image analysis.

When using transfer learning of a similar domain, the model did not have a significant improvement compared to ImageNet weights. However, variables like image normalization based on ImageNet mean and standard deviant and possible data contamination in the Rahman et al. dataset introduced variance that is not possible to measure. Therefore, making it difficult to conclude whether the the usage of transfer learning of a similar domain has a positive impact in the results or whether it has the same effect as the usage of ImageNet weights, a domain of natural images vastly different from histopathology images.

Overall, the findings affirm the potential of transfer learning in histopathological image classification while also emphasizing the need for domain-specific fine-tuning. Addressing domain shift remains a critical challenge that requires further research and innovative approaches to improve model robustness across diverse datasets.

V. CONCLUSION

This study investigated the application of transfer learning for the classification of oral squamous cell carcinoma (OSCC) using histopathological images. By utilizing the DenseNet121 architecture and pre-trained weights from ImageNet, we demonstrated the efficacy of transfer learning in enhancing model performance within the domain of medical image analysis.

The results confirmed that transfer learning is a viable approach, achieving high classification accuracy (90.77% on P-NDP-UFES, 89.39% on Rahman et al.) and F1 scores (90.89% on P-NDP-UFES, 89.79% on Rahman et al.) for OSCC when pre-trained models are fine-tuned on domain-specific datasets. Despite these successes, the study highlighted a significant performance decline (Acc—F1: $47.35\%|51.26\%$ P-NDP-UFES weights on Rahman et al., $73.97\%|73.87\%$ Rahman et al. weights on P-NDP-UFES) when models trained on one dataset were tested on another, underscoring the challenge posed by domain shift. This indicates the necessity for domain-specific fine-tuning to achieve optimal model performance.

Furthermore, the full fine-tuning strategy employed was essential for adapting the pre-trained models to the distinct characteristics of histopathological images, enabling the effective transfer of learned features. This approach allowed for the leveraging of pre-trained low-level features while customizing high-level features to the target task.

To address the limitations posed by domain shift, future research should explore advanced techniques such as domain adaptation, data augmentation, and the development of more generalized feature extractors. Additionally, the creation and curation of larger, more diverse annotated datasets will be imperative for advancing digital pathology.

In summary, while transfer learning offers substantial benefits for medical image classification, addressing domain-specific challenges and ensuring robust model generalization across varied datasets remain critical areas for continued research and innovation.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT in order to improve language and readability. After using this tools, the authors reviewed and edited the content as needed and takes full responsibility for the content of the publication.

References

[1] Michael Awadalla, Matthew Idle, Ketan Patel, and Deepak Kademani. Management update of potentially premalignant oral epithelial lesions. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 125(6): 628–636, 2018.

[2] K Basu, R Sinha, A Ong, and T Basu. Artificial Intelligence: How is it changing medical sciences and its Future? *Indian Journal of Dermatology*, 65:365–370, Out 2020.

[3] Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes van Diest, Bram van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen A. W. M. van der Laak, and CAME-LYON16 Consortium. Assessment of deep learning algorithms for detection of lymph node metastases in women With breast cancer. *JAMA*, 318:2199–2210, 2017.

[4] Maria Clara Falcão Ribeiro de Assis, Leandro Muniz de Lima, Liliana Aparecida Pimenta de Barros, Tânia Regina Velloso, Renato Krohling, and Danielle Camisasca. Ndb-ufes: An oral cancer and leukoplakia dataset composed of histopathological images and patient data. *Mendeley Data*, 2023.

[5] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[6] A K El-Naggar, J K C Chan, J R Grandis, T Takata, and P J Slootweg. *WHO classification of head and neck tumours*, volume 9. IARC WHO Classification of Tumours Series, International Agency for Research on Cancer, 4 edition, 2017.

[7] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[8] Manu Goyal, Thomas Knackstedt, Shaofeng Yan, and Saeed Hassanpour. Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities. *Computers in Biology and Medicine*, 127:104065, 2020.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[10] Le Hou, Dimitris Samaras, Tahsin M. Kurc, Yi Gao, James E. Davis, and Joel H. Saltz. Patch-based convolutional neural network for whole slide tissue image classification. *arXiv e-prints, aXiv:1504.07947*, 2016.

[11] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. In *arXiv preprint arXiv:1704.04861*, 2017.

[12] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2018.

[13] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4): 541–551, 1989.

[14] Guoyu Lu, Yan Yan, Li Ren, Philip Saponaro, Nicu Sebe, and Chandra Kambhamettu. Where am I in the dark: Exploring active transfer learning on the use of indoor localization based on thermal imaging. *Neurocomputing*, 173:83–92, 2016. ISSN 0925-2312.

[15] Beatriz Matias Santana Maia, Maria Clara Falcão Ribeiro de Assis, Leandro Muniz de Lima, Matheus Becali Rocha, Humberto Giuri Calente, Maria Luiza Armini Correa, Danielle Resende Camisasca, and Renato Antonio Krohling. Transformers, convolutional neural networks, and few-shot learning for classification of histopathological images of oral cancer. *Expert Systems with Applications*, 241:122418, 2024. ISSN 0957-4174.

[16] Weihong Deng Mei Wang. Deep visual domain adaptation: A survey. *arXiv e-prints, aXiv:1802.03601*, pages 135–153, 2018.

[17] Behnam Neyshabur, Hanie Sedghi, and Chiyuan Zhang. What is being transferred in transfer learning? *arXiv e-prints, aXiv:2008.11687*, 2020.

[18] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing*

systems, 32, 2019.

[19] Tabassum Yesmin Rahman, Lipi B. Mahanta, Anup K. Das, and Jagannath D Sarma. Histopathological imaging database for oral cancer analysis. *Mendeley Data*, 2023.

[20] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2015.

[21] Paul M. Speight, Timothy J. Abram, Pierre N. Floriano, Robert James, Julie Vick, Martin H. Thornhill, Craig Murdoch, Christine Freeman, Anne M. Hegarty, Katy D'Apice, A. Ross Kerr, Joan Phelan, Patricia Corby, Ismael Khouly, Nadarajah Vigneswaran, Jerry Bouquot, Nagi M. Demian, Y. Etan Weinstock, Spencer W. Redding, Stephanie Rowan, Chih-Ko Yeh, H. Stan McGuff, Frank R. Miller, and John T. McDevitt. Interobserver agreement in dysplasia grading: toward an enhanced gold standard for clinical pathology trials. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 120(4): 474–482.e2, 2015.

[22] Karin Stacke, Gabriel Eilertsen, Jonas Unger, and Claes Lundström. A closer look at domain shift for deep learning in histopathology. *arXiv e-prints, aXiv:1909.11575*, 2019.

[23] Karin Stacke, Gabriel Eilertsen, Jonas Unger, and Claes Lundstrom. Measuring domain shift for deep learning in histopathology. *IEEE J. Biomed. Health Inform.*, 25(2): 325–336, February 2021.

[24] Karin et al. Stacke. Measuring domain shift for deep learning in histopathology. *IEEE journal of biomedical and health informatics*, pages 325–336, 2021. doi: 10.1109/JBHI.2020.3032060.

[25] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR, 2019.

[26] Yunhan Tan, Zhihan Wang, Mengtong Xu, Bowen Li, Zhao Huang, Siyuan Qin, Edouard C. Nice, Jing Tang, and Canhu Huang. Oral squamous cell carcinomas: state of the field and emerging directions. *International Journal of Oral Science*, 15, 2023.

[27] The Global Cancer Observatory. Lip, oral cavity - international agency for research on cancer, Dec 2020. Available at:¡https://gco.iarc.fr/today/data/factsheets/cancers/1-Lip-oral-cavity-fact-sheet.pdf¿. Accessed at: 14th of June, 2024.

[28] Hamid Reza Tizhoosh and Liron Pantanowitz. Artificial intelligence and digital pathology: challenges and opportunities. *Journal of Pathology Informatics*, 9:38, 2018.

[29] Antonio Torralba and Alexei A. Efros. Unbiased look at dataset bias. *IEEE*, pages 1521–1528, 2011. doi: 10.1109/CVPR.2011.5995347.

[30] Tomotaka Ugai, Naoko Sasamoto, Hwa-Young Lee, Mariko Ando, Mingyang Song, Rulla M. Tamimi, Ichiro Kawachi, Peter T. Campbell, Edward L. Giovannucci, Elisabete Weiderpass, Timothy R. Rebbeck, and Shuji Ogino. Is early-onset cancer an emerging global epidemic? Current evidence and future implications. *Nature Reviews Clinical Oncology*, 19(10):656–673, Out 2022.

[31] Karl Weiss, Taghi M. Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big Data*, 3, 2016.

[32] WHO. Cancer. Fev 2022. Available at: https://www.who.int/news-room/fact-sheets/detail/cancer#: :text=Cancer%20is%20a%20leading%20cause,and%20rectum%20and%20prostate%20cancers. Accessed on the: 14th of June.

[33] Katarzyna Woźniak and Dariusz Iżycki. Cancer: a family at risk. *Prz. Menopauzalny*, 13(4):253–261, September 2014.

[34] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? *Advances in neural information processing systems*, 27, 2014.