Research Article

Raneem Y. Alsaedi*, and Hussam J. Alsharif

Classification of histopathological images for oral cancer in early stages using a deep learning approach

Abstract: Problem – Diagnosing early-stage oral squamous cell carcinoma (OSCC) is necessary for patient treatment and survival. The staining normalization and augmentation of histopathological images can play an important role in improving the accuracy of early oral cancer (OC) detection, especially with imbalanced and limited datasets, which is a difficult manual task because it is time-consuming, requires effort, and is subject to variation among different pathologists. Applying these methods can enhance the performance of deep learning models used for OSCC detection. **Aim –** This study addresses the aforementioned challenges by showing how the random stain normalization and augmentation (RandStainNA) technique can be applied to datasets as a preprocessing step along with transfer learning models to categorize histopathological images of OC into three classes. **Methods –** The performance of three models – ResNet-50, VGG-19, and an ensemble model – in image classification is compared. Using fivefold cross-validation for training, our framework compares two main tasks. The best results were achieved on the task in which the models were pretrained using patches, and the RandStainNA technique was applied to the images as a preprocessing step. We trained the models to extract features from the relevant images and then determine the final classifications. **Results –** Our proposed framework achieved the best results in the early diagnosis of OC using the NDB-UFES dataset and the ResNet-50 model, demonstrating 73.33% balanced accuracy, 74.68% precision, and 74.35% recall, as well as a 92.21% area under the curve. **Conclusion –** Our proposed framework was effective in improving the accuracy of early OC detection, especially when working with small and imbalanced datasets. In addition, this method may contribute to enhancing the generalizability of models and using them in diverse laboratories.

Keywords: oral cancer, cancer classification, oral leukoplakia, leukoplakia with dysplasia, deep learning, transfer learning, histopathological images

1 Introduction

Oral cancers (OCs) are common worldwide. Oral squamous cell carcinoma (OSCC) is a branch of head and neck squamous cell carcinoma [1]. In general, the prognosis and treatment of oral cavity cancer depend on the stage of the disease at diagnosis, and the survival period is about 5 years on average [2]. A late OC diagnosis can lead to a 30% decrease in the chance of survival, which increases the fatality rate. Therefore, early diagnosis is essential to optimize the chance of survival [3,4]. Currently, biopsy is the standard method for confirming the diagnosis of OC [5]. However, microscopy-based histopathological examination is often challenging, requires

* **Corresponding author: Raneem Y. Alsaedi,** Department of Computer Science, Jamoum University College, Umm Al-Qura University, Makkah 25375, Saudi Arabia, e-mail: s44280246@st.uqu.edu.sa
Hussam J. Alsharif: Department of Computer Science, Jamoum University College, Umm Al-Qura University, Makkah 25375, Saudi Arabia, e-mail: hjshareef@uqu.edu.sa

human expertise, and is time-consuming. The automated analysis of images of oral biopsies can help histopathologists optimize their performance and diagnose the images more accurately [6].

Oral leukoplakia is a disorder with a probability of becoming malignant. It appears in the form of a white patch in the mouth. The etiology of leukoplakia includes factors such as smoking cigarettes and consuming alcohol; irritation of the oral mucosa is the most prominent cause [7]. In review by Aguirre-Urizar et al. [8], 24 studies were analyzed to evaluate the risk factors for and estimate the likelihood of malignant transformation in oral leukoplakia. The study revealed that the percentage of malignant transformation ranged from 1.1 to 40.8%, and epithelial dysplasia was significantly associated with malignant transformations.

The complexity of histopathological data makes it challenging to classify oral lesions and OC. Oral leukoplasia also exhibits high intra-disease variability [9] within single classes, including heterogeneous histopathological patterns, such as differences in the thickness of the epithelium and keratinization; inter-class variability between dysplastic and non-dysplastic patients, and overlapping cytological features. This complexity can greatly hinder a model's ability to distinguish between classes. Furthermore, this issue contributes to the problem of class imbalances in datasets, as models tend to make predictions about majority classes. Unpredictable differences in staining protocols and imaging across laboratories introduce another obstacle to the generalizability of models. To address these challenges and optimize the accuracy of diagnoses, we must be able to extract high-resolution features from advanced computational models while simultaneously dealing with imbalanced datasets and variability in staining.

Transfer learning models have been used in several studies to address these limitations in the context of OC diagnosis [3,6,10]. These models offer the advantage of being pretrained on large datasets by learning various features. However, small dataset sizes, along with subtle variations in classes of histological images, limit the performance of such models. In most studies, the augmentation technique is used to increase the size of datasets – for instance, through horizontal and vertical flipping. While this can be effective in some datasets, it has proven inadequate elsewhere. For this reason, augmentation of histological images has been complemented by stain normalization. This offers a significant means of enhancing model training and can lead to crucial improvements, especially when used along with transfer learning models. Recently, the random stain normalization and augmentation (RandStainNA) [11] technique for classifying histopathological images was developed. RandStainNA combines stain normalization and stain augmentation with the objective of enhancing histopathological image analysis through the use of deep learning models, and it aims to optimize their generalizability. However, the present methods face many issues, including the following:

(1) Variation in histological features: a single sample may exhibit heterogeneity in the form of distinct features (such as non-dysplastic and dysplastic characteristics) – for instance, hyperplasia and keratinization for non-dysplastic features, or koilocytosis and intercellular bridges for dysplastic features – making it difficult for the model to learn the specific features to be classified.

(2) Subtle changes between categories: some dysplastic lesions exhibit mild or moderate changes yet remain similar to those that do not undergo dysplastic changes, leading to difficulty in determining the appropriate class.

(3) Ability to limited and multiclass datasets: many models have difficulty handling small datasets with multiple classes, limiting their applicability in certain research contexts.

(4) Efficiency issues due to imbalances between classes: most models have difficulty dealing with imbalances among classes, which leads to predictions of the bise majority class.

To address these challenges, this study develops a deep learning framework by subjecting models Visual Geometry Group (VGG)-19 and ResNet-50 to transfer learning in order to classify histopathological images of the oral cavity into three categories. Two categories represent the early stages of cancer, and the third category represents the malignant OSCC. This method is designed to ensure more accurate diagnoses of OC, which can lead to faster and more effective treatment plans. By helping to provide pathologists with efficient and accurate early diagnoses, this method could help reduce the number of potential deaths caused by OC. In this context, this study's objective is to facilitate the accurate diagnosis of OC in its early stages and enhance the generalizability of relevant models by adopting stain normalization and augmentation in tandem with the effective extraction of features by transfer learning models. More specifically, this research aims to classify

OSCC and oral lesions subject to potentially malignant transformations (i.e., oral leukoplakia with and without dysplasia) accurately and reliably by applying stain normalization and augmentation to histopathological images. The purpose of this work is to show how RandStainNA can be used to increase and feed training datasets in order to train transfer learning models in early-stage OSCC classification. Specifically, we present a framework that applies the RandStainNA technique as a preprocessing step to analyze the NDB-UFES dataset, while using the ResNet-50 and VGG-19 transfer learning models to improve the classification of OSCC in the early stages. The proposed framework's performance is evaluated in terms of balanced accuracy (BCC), precision, recall, and area under the curve (AUC). The primary findings indicate that the proposed framework improves classification performance, with the AUC increasing from 77 to 92% in the task with pretrained patches. These findings suggest that RandStainNA has the ability to improve the classification of early-stage OSCC, which could contribute to future cancer research. The following sections provide a literature review , a description of our methodology, and the results of the proposed framework. The study's main contributions are as follows:

- Improves the handling of dataset complexity: the subtle extraction of features from histopathological images requires an effective model. Histopathological images present several challenges in terms of intra-class variability and subtle differences between classes.
- Handling limited and imbalanced datasets: RandStainNA was applied to the NDB-UFES dataset to categorize images into three classes and investigate their effectiveness. Our study demonstrates that this application is effective because it improves the accuracy of classifications.
- Classification: the main objective was to accurately classify OSCC in the early stages using the proposed framework, which involves applying RandStainNA to the NDB-UFES dataset to facilitate feature extraction and classification using transfer learning models.
- The classification performance of our technique is compared to that of previous studies. On all performance metrics, our framework outperformed the other models.

This article is structured as follows: Section 2 presents the literature review, highlighting previous relevant works and their limitations. Section 3 describes the methodology, introducing our framework, the transfer learning models, the dataset, RandStainNA as a preprocessing approach, the data division and experiments, and the metrics we used for evaluation. Section 4 presents the results of the main tasks and discusses them, comparing the results with those of previous studies. Section 5 concludes the study and specifies potential future research directions.

2 Related works

This section presents a review of previous studies that have used artificial intelligence to classify cancer and OC. Given the rapid development in the field of cancer diagnosis and treatment, numerous studies have developed methods for using modern medical technologies to contribute to better treatment planning [12]. Artificial intelligence has exhibited promising capabilities for analyzing and processing medical data, such as omics and multi-omics, in order to accurately diagnose cancer [13]. In recent study, multi-omics has been used to improve cancer subclassification, with the best results achieved by applying feature selection along with machine learning algorithms, such as the K-means algorithm and support vector machine [14]. In the latter context, deep learning techniques have been used to analyze medical images to make early and accurate OC diagnoses. Most research has focused on the division of histopathological images into normal images and those indicative of OSCC [10,15–17]. Additionally, researchers have commonly employed transfer learning models, such as ResNet-50, VGG-16, and InceptionV3. On the other hand, some studies have used convolutional neural networks (CNNs) to identify high-risk epithelia dysplasia and normal epithelia [18] and to facilitate multi-class grading of dysplasia [19]. In addition, more recent research [3] has categorized histopathological images of oral leukoplakia into classes: leukoplakia without dysplasia (LW/oD), leukoplakia with dysplasia (LW/D), and the malignant OSCC. Some recent studies on these topics are listed in Table 1.

Table 1: Previous research on the classification of OC using deep learning

No.	Authors/Year	Dataset	Methodology	Results	Limitations
1	Camalan et al. [20] 2021	Photographs in two datasets (containing 30 and 43 images) were classified as either normal or indicative of oral dysplasia	Transfer learning (Inception- ResNet-V2) and decision- making automated heat maps were generated	Accuracy rates of 73.6 and 90.9% were achieved on the two datasets	• The sample size of the dataset was relatively small; the total number of patients was 54 in both datasets • The grade of dysplasia (mild, moderate, severe) could not be detected from the photographic images, and a histopathological examination was required to determine the grades.
2	Amin et al. [6] 2021	The RMDS dataset contains histopathological images of 290 normal and 934 OSCC samples	A concatenated model that employed a transfer learning technique by individually fine-tuning three pretrained deep learning models to detect OSCC, namely, ResNet50, VGG16, and InceptionV3, was used for feature extraction.	They achieved 96.66% accuracy	• The number of OSCC cases may not fully represent the real-life diversity of the condition and needs to be increased to generalize the model.
3	Deif et al. [10] 2022	The RMDS dataset contains histopathological images of 290 normal and 934 OSCC samples	Images were enhanced using the Reinhard method; a feature extraction deep learning model (InceptionV3) with binary particle swarm optimization was used to pick the best features, which were classified using the XGBoost model.	They achieved 96.3% accuracy	• The number of OSCC cases may not fully represent the true diversity of the condition and needs to be increased to generalize the model. • The specificity was 87.93%, and the true-negative value still needs improvement
4	Hag et al. [16] 2023	A public dataset in Kaggle that contains two classes of histopathological images, including 2,698 OSCC and 2,494 normal images	The researchers used two techniques for feature extraction, namely, ResNet50 and the Gabor filter, and then classified images using the CatBoost algorithm.	They achieved 94.92% accuracy	• Their model misclassified 14 cases of malignant OSCC tumors as normal.
5	de Lima et al. [3] 2023	The NDB-UFES dataset contains metadata on and histopathological images of OC, which are placed into three classes: leukoplakia without dysplasia, and leukoplakia with dysplasia, and OSCC	The CoaT, PiT, RegNetY, ResNetV2, and ViT transfer learning models were used for Task IV; the PiT model achieved the best results using only histopathological images	The balanced accuracy was 63.70% on Task IV, and it was 67.54% on Task IV with pretrained patches	• The results indicate that improvements in the accuracy and efficiency of OC diagnosis are still required.

Although they have made significant contributions to the classification of OC images using deep learning models, these studies have some limitations. Most studies have focused on binary classification, i.e., the simple determination of whether OC exists. However, the specific identification of early-stage OC is crucial to enable timely treatment and ensure the greatest chance of survival, especially with the deadly OSCC. In light of these gaps, the present study aims to facilitate the classification of OSCC and oral lesions subject to potentially malignant transformations – LW/D and LW/oD – as the early stages. Furthermore, recent studies have used histopathological images to classify OC images into three classes: LW/oD, LW/D, and OSCC. However, the results need further refinement so that accurate diagnoses can be ensured. Therefore, this study set out to fill these gaps by applying the RandStainNA technique to improve the performance of deep learning models and thus yield more accurate and reliable diagnoses.

3 Materials and methods

Figure 1 presents the general framework of our oral histopathological image classification system, which was used to improve the performance of deep learning models in classifying OC.
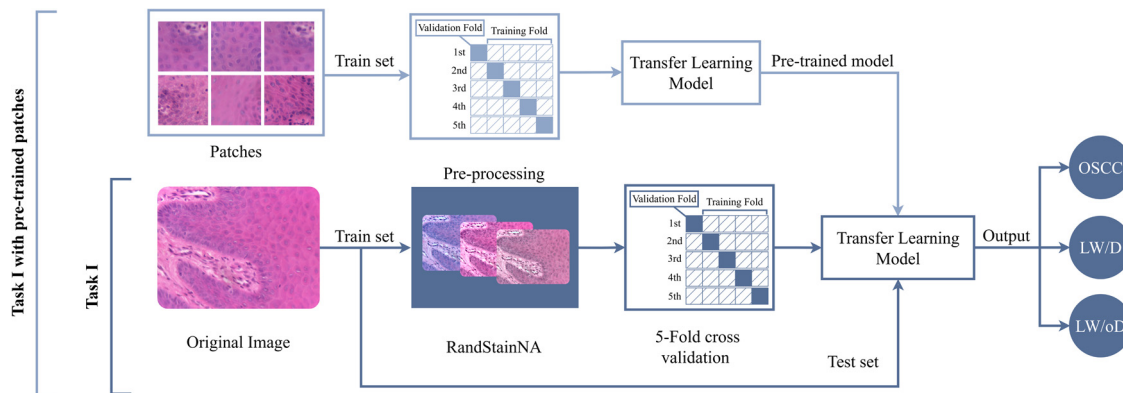

Figure 1: Our framework for the three-class categorization of histopathological images. The figure shows Task I and Task I with pretrained patches, which were used to distinguish three classes of images – OSCC: oral squamous cell carcinoma; LW/D: leukoplakia with dysplasia; LW/oD: leukoplakia without dysplasia. Source: Created by the authors.

3.1 Data description

The dataset we used was published in 2023 by the Oral Diagnosis Project of the Federal University of Espírito Santo (NDB-UFES) [21]. It is composed of high-resolution 2,048 × 1,536-pixel histopathological images of oral leukoplakia and OSCC from biopsies performed between 2010 and 2021 and stained with hematoxylin–eosin. A total of 237 images were obtained: 91 for OC, 89 for oral LW/D, and 57 for LW/oD. Figure 2 provides examples of
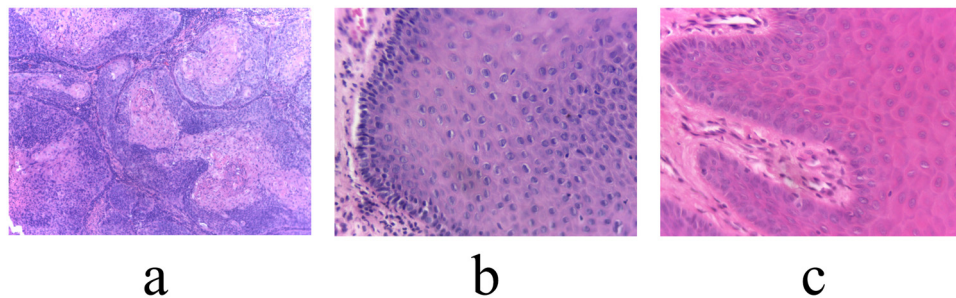

Figure 2: Examples of images from the three classes in the NDB-UFES dataset: (a) oral squamous cell carcinoma, (b) LW/D, and (c) LW/oD.

images from these three classes. In addition, the dataset includes patches of the images with 512 × 512 pixels. There were 3,763 patches in total, of which 1,126 were of OC, 1,930 were of LW/D, and 707 were of LW/oD.

3.2 Preprocessing

Model generalization is a challenge when analyzing histological images due to differences between digital scanners and between their responses, which can affect their performance. To address this problem, some research on stain normalization [22] has suggested transferring color characteristics from one image to another using statistical analysis. The technique suggested in the study by Shen et al. [11] unifies two methods of analyzing histological images: stain normalization, in which an image is used as a template to alleviate stain differences, and stain augmentation, which enriches stain styles by simulating variations of those stains. With the objective of decreasing generalization errors, this method, called RandStainNA, applies these techniques to the problem of changing stain styles in the workable range and with several color spaces selected randomly. The colors in image I_m Im are converted from RGB to LAB, and the pixel value can be written as $[l, a, b]$, where each channel in this image is represented by the average $G_m = [g_m^{(l)}, g_m^{(a)}, g_m^{(b)}]$ and the standard deviation $S_m = [s_m^{(l)}, s_m^{(a)}, s_m^{(b)}]$. The resulting virtual template associated with image I_m from Gaussian distribution F_G and F_S is represented by $G_t = [g_t^{(l)}, g_t^{(a)}, g_t^{(b)}]$ and $S_t = [s_t^{(l)}, s_t^{(a)}, s_t^{(b)}]$)]. The formula that represents the image-wise normalization by the random virtual template is as follows:

$$
\begin{cases}
l' = \dfrac{s_t^{(l)}}{s_m^{(l)}}(l - g_m^{(l)}) + g_t^{(l)}, \\[2mm]
a' = \dfrac{s_t^{(a)}}{s_m^{(a)}}(a - g_m^{(a)}) + g_t^{(a)}, \\[2mm]
b' = \dfrac{s_t^{(b)}}{s_m^{(b)}}(b - g_m^{(b)}) + g_t^{(b)}.
\end{cases}
\tag{1}
$$

The color space is then transferred again $[l', a', b']$ from the LAB to the RGB space. In this study, to increase the generalizability of the model and the number of training datasets, we used the RandStainNA technique with the color spaces in LAB. The number of datasets achieved after applying RandStainNA is shown in Table 2, and Figure 3 presents the steps of the RandStainNA technique when applied to the dataset.

Table 2: Dataset used in this study

Class	Leukoplakia without dysplasia	LW/D	Malignant OSCC	Total
Original images	57	89	91	237
Training set	48	74	76	198
Test set	9	15	15	39
RandStainNA	1,152	1,147	1,140	3,439

Note: This represents the number of images in the dataset after being split into training and test sets and after applying the RandStainNA technique to the training set.

3.3 Transfer learning

In most studies, transfer learning using CNNs has been applied to extract features from medical images. The researchers in the study by Li et al. [23] used ResNet-50 to extract features from breast cancer histopathological images, while the study by Liu et al. [24] used ResNet-50 to predict the primary tumor areas in spinal
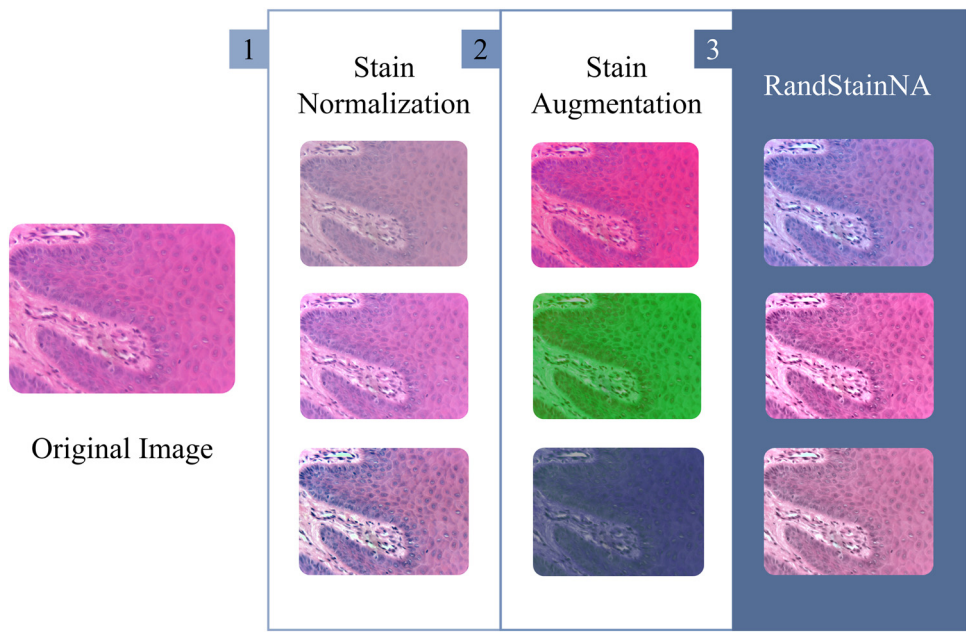
Figure 3: Steps involved in performing random stain normalization and augmentation on an image from the dataset used in this study. Source: Created by the authors.

metastases using magnetic resonance imaging. In this study, we used the pretrained VGG-19 and ResNet-50 models.

3.3.1 VGG-19

The VGG model is a CNN pre-trained model that was suggested in 2014 [25] by Simonyan et al. at Oxford University, UK, where the VGG was trained on ImageNet with 1.3 million images. The VGG-19 model contains 19 connected layers, which consist of fully and highly connected convolutional layers, and uses max pooling and the Softmax activation function. In this study, we used the same architecture with a simple modification in the classification block, which contains two linear layers and the ReLU activation function between them, followed by a dropout with a value of 0.5 (see Figure 4 for present the VGG-19 model architecture).
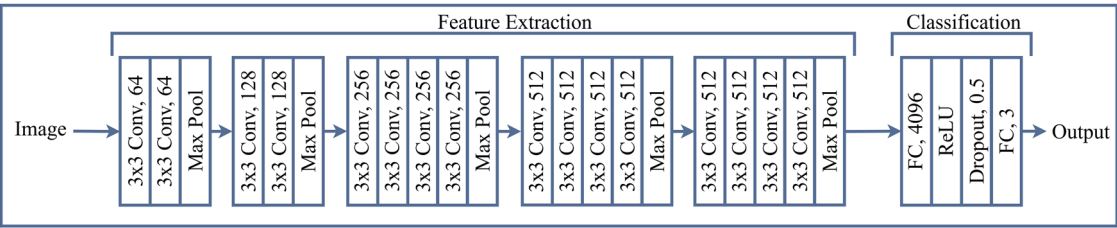

Figure 4: Graphic representation of the VGG-19 model architecture created by Simonyan et al. [25].

3.3.2 ResNet-50

ResNet-50 is a deep CNN [26] containing 50 deep layers. It is based on the ResNet architecture, which uses a residual block to add the output from one layer to the next, enabling skip connections and solving the vanishing gradient problem. The ResNet-50 model was trained on one million images from ImageNet. The

architecture consists of many CLs followed by BatchNorm and ReLU activation to extract features. The data are subjected to max pooling, which reduces the spatial dimensions. It also contains two types of blocks, i.e., identity and convolutional blocks, to process and transform these features. Both blocks pass on the input and are processed through the CL, which adds the input to the output. In the convolutional blocks, a 1 × 1 CL is added prior to a 3 × 3 CL to decrease the number of filters. In the last of the ResNet-50 networks, the final classification is performed by the fully connected layers, which use Softmax as an activation function to determine the probability of the final class. Figure 5 presents the ResNet-50 model architecture.
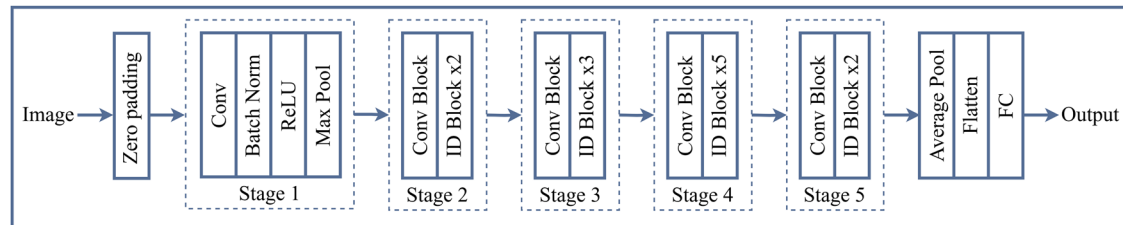

Figure 5: Graphic representation of the ResNet-50 model architecture created by He et al. [26].

3.3.3 Ensemble model

We also combined the pretrained VGG-19 and ResNet-50 models, which had been trained independently on patches of three classes from the dataset. We used fivefold cross-validation and stochastic gradient descent (SGD) optimization for each model, with 10 training epochs for VGG-19 and 15 for ResNet-50. We then saved the models to reuse as pretrained models. The resulting ensemble model used a bagging approach to train the two models together and then classify the images.

3.4 Data splitting and experiments

To evaluate the models, we divided the dataset into training and validation (5/6) and testing (1/6) portions and used standard forms of augmentation, including flipping, rotation, and range shifting. We experimented with various hyperparameters and settled on a 0.001 learning rate and fivefold cross-validation for all models. There were some differences in the settings for each model. With the VGG-19 model, we used an SGD optimizer with ten epochs. For the ResNet-50 model, we employed two SGD optimizers for the pretraining phase and Adam in the training on images using 15 epochs. For the ensemble model in the training phase, we used Adam optimization with five epochs. Table 3 displays the parameters of each model.

Table 3: Parameters of the transfer learning models

Parameter	VGG-19 model	ResNet-50 model	Ensemble model
Learning rate	0.001	0.001	0.001
Batch size	64	64	64
Epochs	10	15	5
Optimizer	SGD	SGD/Adam	Adam
Loss function	CrossEntropyLoss	CrossEntropyLoss	CrossEntropyLoss
Dropout rate	0.5	Not used	Not used
Number of layers	19	50	Combined (VGG-19 and ResNet-50 structure)

We performed the experiments for Task I using two methods to categorize images into three classes: LW/oD, LW/D, and OSCC. The steps for this task are explained in this section.

Task I: Preprocessing: we applied RandStainNA to the original images from all classes. Training: the transfer learning models were trained using the images after preprocessing, and fivefold cross-validation was applied during the training. Testing: the transfer learning models were tested using images unseen by the models, and the model's performance was evaluated using the metrics of BCC, precision, recall, and AUC.

Task I with pretrained patches: Pretraining: the transfer learning models were trained using the patches of training images from the NDB-UFES dataset. We applied fivefold cross-validation during the training and then saved the models. Training: the models we saved after pretraining using patches were used to train the model on images, and RandStainNA was applied. We again applied fivefold cross-validation during the training. Testing: the transfer learning models were tested using images unseen by the models, and the performance was again evaluated based on BCC, precision, recall, and AUC.

3.5 Evaluation metrics

To evaluate the models' performance on the training set, we used fivefold cross-validation to generate five trained models, and the average for the results of these five was computed after evaluation on the testing set. To assess the models' performance, we used the four metrics: BCC, precision, recall, and AUC. BCC is used to deal with imbalanced datasets. As shown in equation (2), it is defined as the mean of both the recall true-positive rate and the specificity true-negative rate, as shown in equations (4) and (5), respectively. Precision reflects the percentage of images diagnosed, as shown in equation (3). Recall indicates the proportion of images that the model correctly diagnosed. Finally, we used the macro average of the AUC values:

$$BCC = \frac{Recall + Specificity}{2}, \tag{2}$$

$$Precision = \frac{True\ Positive\ (TR)}{True\ Positive\ (TR) + False\ Positive\ (FP)}, \tag{3}$$

$$Recall = \frac{True\ Positive\ (TR)}{True\ Positive\ (TR) + False\ Negative\ (FN)}, \tag{4}$$

$$Specificity = \frac{True\ Negative\ (TN)}{True\ Negative\ (TN) + False\ Positive\ (FP)}. \tag{5}$$

4 Results and discussion

In this section, we present and discuss the results of Task I and Task I with pretrained patches.

In Task I, we investigated the effectiveness of using the RandStainNA technique on the NDB-UFES dataset. This effectively improved the model's classification performance based on our framework, as the results in Table 4 suggest. This led to improvement on all metrics, with an average improvement in BCC of 8.6%,

Table 4: Results for Task I: classifying images into three classes

Model	BCC	Precision	Recall	AUC
Original images				
VGG-19	0.6296	0.7041	0.6410	0.8416
ResNet-50	0.5259	0.5942	0.5897	0.7732
Ensemble	0.5111	0.5319	0.5897	0.7854
Images with RandStainNA technique				
VGG-19	0.6666	0.7271	0.7179	0.8919
ResNet-50	0.6370	0.6802	0.6666	0.8143
Ensemble	0.6296	0.6804	0.6410	0.7977

precision of 8.7%, recall of 7%, and AUC of 3.3%. The greatest improvement was observed in precision, followed by BCC, indicating that the models became more reliable in correctly determining positive predictions while effectively decreasing the false-positive rate. This indicates that the models became more effective in dealing with complex, imbalanced datasets. The best results were achieved on Task I using the VGG-19 model, where the BCC was 66.66%, the precision was 72.71%, the recall was 71.79%, and the AUC was 89.19%.

In Task I with pretrained patches, we aimed to improve the performance of the models by using patches of the images to pretrain the models; we also measured the effectiveness of this method along with that of our framework, including the RandStainNA technique. This methodology was highly effective with some models, such as ResNet-50, which achieved the best performance, followed by the ensemble model. As we can see in Table 5, the best results were achieved using the ResNet-50 model: the BCC was 73.33%, precision was 74.68%, recall was 74.35%, and AUC was 92.21%. The improvements in BCC and AUC indicate that the model's capacities to deal with imbalance in the dataset and ability to discriminate between classes were effectively improved, leading to more accurate diagnoses of images into the LW/oD class. In contrast, the VGG-19 model performed less well due to the structure of the model and its ability to effectively extract important features.

Table 5: Results for Task I with pretrained patches: classifying images into three classes

Model	BCC	Precision	Recall	AUC
VGG-19	0.6666	0.7021	0.6153	0.7640
ResNet-50	0.7333	0.7468	0.7435	0.9221
Ensemble	0.6740	0.7094	0.6923	0.8549

The AUC values on the tasks are compared in Figure 6. The best result achieved in Task I was for OSCC, followed by the LW/D class. In Task I with pretrained patches, the best result was for OSCC, followed by the LW/oD class. It seems that the OSCC class had the same value and the best results in both, and the pretrained patches improved the distinction between the LW/oD and LW/D classes, most significantly for the LW/oD class.
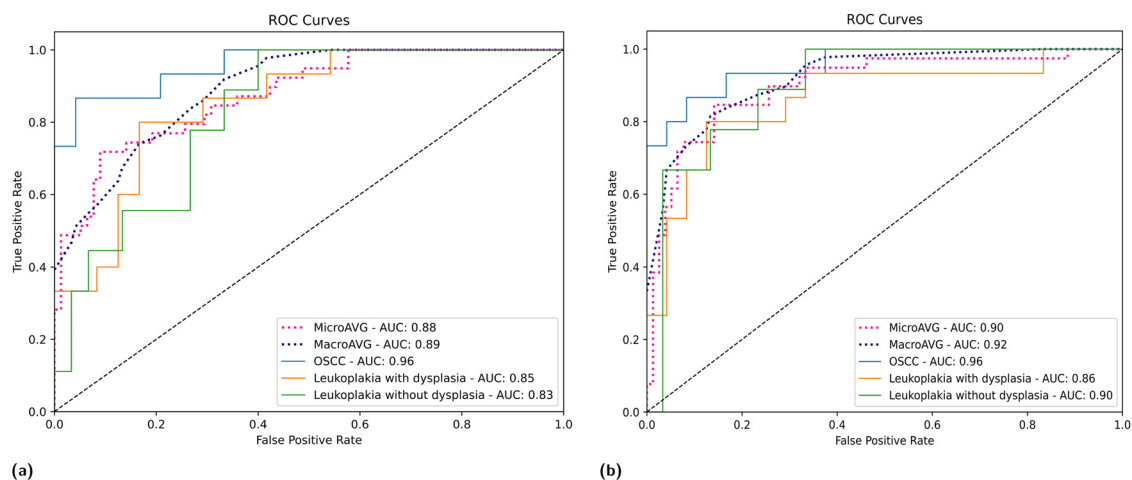

Figure 6: AUC results for both tasks: (a) Task I using the VGG-19 model and (b) Task I with pretrained patches using the ResNet-50 model. Source: Created by the authors.

4.1 Comparison with previous works

The performance of our framework can be compared with that of the methods used in the latest studies. The differences in performance between Task I and Task I with pretrained patches are shown in Table 6. Our

Table 6: Comparison of our results with state-of-the-art research

Model	BCC	Precision	Recall	AUC
Task I				
PiT [3]	0.6370	0.6760	0.6560	0.8406
VGG-19 (Ours)	0.6666	0.7271	0.7179	0.8919
Task I with pretrained patches				
PiT [3]	0.6745	0.6920	0.6760	0.8188
ResNet-50 (Ours)	0.7333	0.7468	0.7435	0.9221

framework achieved better results on Task I using both methods. To the best of our knowledge, only one study has used patches of images from the NDB-UFES dataset because it is new. Moreover, de Lima et al. [3] completed Task I using images and patches of images to classify the three classes. Their study achieved the best results using the PiT model for both procedures from Task I: the BCC was 63.70%, precision was 67.60%, recall was 65.60%, and AUC was 84.06% for Task I, while the BCC was 67.45%, precision was 69.20%, recall was 67.60%, and AUC was 81.88% for Task I with pretrained patches. The results using our framework demonstrated improved performance: the BCC was 66.66%, precision was 72.71%, recall was 71.79%, and AUC was 89.19% on Task I using the VGG-19 model, and the BCC was 73.33%, precision was 74.68%, recall was 74.35%, and AUC was 92.21% in Task I with pretrained patches using the ResNet-50 model.

4.2 Limitations

The proposed framework has proven effective in improving the results of classifications of OC in the early stages. However, our study has a few limitations that limit the learning ability of the models, including imbalanced and limited datasets and limited resources available to train the models. Regarding the issue of imbalanced and limited datasets, we used NDB-UFES, which contains histopathological images categorizable into three classes, namely, OSCC, LW/D, and LW/oD, of which LW/oD is a minority class by a difference of 32 samples. Additionally, the complexity of histopathology increases the difficulty of the learning process, particularly with respect to intra-disease heterogeneities in oral leukoplakia and the subtle differences between no dysplasia (LW/oD) and mild-to-moderate dysplasia (LW/D). This may have led to uncomprehensive data for each class in training and classification, which may have affected the accuracy of the outcomes. In addition, the process of training the models requires high-performance resources, particularly when dealing with the high resolution of histopathological images, which, in the NDB-UFES dataset, had a size of 2,048 × 1,536 pixels. This problem emerged when we applied the RandStainNA technique, wherein the number of images was increased by stain augmentation. This study used Colab Pro resources and an A100 processor with 83.5 gigabytes of random access memory. Therefore, limited resources were available to train the models.

5 Conclusions

This study demonstrates the efficacy of using stain normalization and augmentation to improve the performance of transfer learning models in classifying histopathological images, comparing these methods with existing research. This approach facilitated more accurate diagnosis of OC, which could improve the survival rate and treatment of the disease. Applying stain normalization and augmentation provides a promising method for processing limited and complex datasets, which makes it valuable in the histopathological field. These enhancements help improve the performance of models used to classify OC in its early stages, which further contributes to diagnosis and treatment. The theoretical outcome of this study was to prove the efficiency of the RandStainNA technique in handling limited datasets with subtle differences between classes;

the practical outcome is to highlight the improvement in the accuracy of diagnoses of OC in the early stages. The main contribution of this study was to propose a framework based on the use of the RandStainNA technique, along with transfer learning models, to enhance the accuracy of OC diagnosis in the early stages, as we have demonstrated how such classifications can be effectively improved. This framework is appropriate for dealing with the challenges of histopathological images, such as subtle differences between classes and heterogeneous intra-class features. This proposed method allows for the efficient preprocessing of this type of data. In this respect, we found that the use of RandStainNA is justified by certain practical advantages. Specifically, the accuracy of OC classification models is enhanced by the possibility of preprocessing complex and limited histopathological images. This improvement in accuracy can lead to more accurate diagnoses and treatment of OC in the early stages. However, this approach has limitations, such as application to an imbalanced and limited dataset, and the intra-class heterogeneity of the histopathological images, wherein there are subtle differences between categories, especially in terms of the differences between mild dysplasia and no dysplasia. In addition, limited resources were available to train the models, as the high resolution of histopathological images requires high-performance resources, particularly when augmenting and thus increasing the dataset. For future research, we recommend exploring the fusion of new transfer learning models, such as CaiT, InceptionNeXt, and BeiT, with morphological texture-extracting methods to extract features. Combined with the use of algorithms designed to select important features, this may help address the complexity of histopathological images by teaching models to understand the patterns in data and select important features, which can further improve classifications. Integrating clinical and demographic data, such as those provided in the NDB-UFES dataset of histopathological images, and using multimodal fusion, such as late plus intermediate fusion, may provide a better understanding of these models, so that they can efficiently handle the limitations of dataset sizes.

Author contributions: Raneem Y. Alsaedi: conceptualization, formal analysis, investigation, methodology, software, writing, and editing. Hussam J. Alsharif: review and supervision.

Conflict of interest: The authors state no conflict of interest.

Ethical approval: No ethical approval was required for this study, as it used the publicly available NDB-UFES dataset, which was ethically approved by the original authors.

Data availability statement: The dataset NDB-UFES is publicly available at http://doi.org/10.17632/bbmmm4wgr8.

References

[1] He S, Chakraborty R, Ranganathan S. Proliferation and apoptosis pathways and factors in oral squamous cell carcinoma. Int J Mol Sci. 2022;23(3):1562. doi: 10.3390/ijms23031562.

[2] Ray CS. IARC handbooks on cancer prevention, vol. 19, Oral cancer prevention. Mumbai: Medknow; 2024. doi: 10.4103/IJC.IJC_234_24.

[3] de Lima LM, de Assis MCFR, Soares JP, Gräo-Velloso TR, de Barros LAP, Camisasca DR. Importance of complementary data to histopathological image analysis of oral leukoplakia and carcinoma using deep neural networks. Intel Med. 2023;3(4):258–66. doi: 10.1016/j.imed.2023.01.004.

[4] Myriam H, Abdelhamid AA, El-Kenawy ESM, Ibrahim A, Eid MM, Jamjoom MM, et al. Advanced meta-heuristic algorithm based on Particle Swarm and Al-biruni Earth Radius optimization methods for oral cancer detection. IEEE Access. 2023;11:23681–700. doi: 10.1109/ACCESS.2023.3253430.

[5] Lingen MW, Abt E, Agrawal N, Chaturvedi AK, Cohen E, D'Souza G, et al. Evidence-based clinical practice guideline for the evaluation of potentially malignant disorders in the oral cavity: a report of the American Dental Association. J Am Dental Assoc. 2017;148(10):712–27. doi: 10.1016/j.adaj.2017.07.032.

[6]   Amin I, Zamir H, Khan FF. Histopathological image analysis for oral squamous cell carcinoma classification using concatenated deep learning models. medRxiv. 2021. doi: 10.1101/2021.05.06.21256741.

[7]   Jurczyszyn K, Gedrange T, Kozakiewicz M. Theoretical background to automated diagnosing of oral leukoplakia: a preliminary report. J Healthc Eng. 2020; 2020:8831161. doi: 10.1155/2020/8831161.

[8]   Aguirre-Urizar JM, Lafuente-Ibáñez de Mendoza I, Warnakulasuriya S. Malignant transformation of oral leukoplakia: Systematic review and meta-analysis of the last 5 years. Oral Diseases. 2021;27(8):1881–95. doi: 10.1111/odi.13810.

[9]   Cirillo N. Precursor lesions, overdiagnosis, and oral cancer: A critical review. Cancers. 2024;16(8):1550. doi: 10.3390/cancers16081550.

[10]   Deif MA, Attar H, Amer A, Elhaty IA, Khosravi MR, Solyman AA, et al. Diagnosis of oral squamous cell carcinoma using deep neural networks and binary Particle Swarm optimization on histopathological images: an AIoMT approach. Comput Intell Neurosci. 2022;2022:6364102. doi: 10.1155/2022/6364102.

[11]   Shen Y, Luo Y, Shen D, Ke J. Randstainna: Learning stain-agnostic features from histology slides by bridging stain augmentation and normalization. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham, Switzerland: Springer; 2022. p. 212–21. doi: 10.1007/978-3-031-16434-7_21.

[12]   Benameur N, Awadi R, Bouabidi A, Mohammed MA, Rehman MU, Ounalli L. Numerical study of two microwave antennas dedicated to superficial cancer hyperthermia. Proc Comput Sci. 2024;239:470–82. doi: 10.1016/j.procs.2024.06.195.

[13]   Ali AM, Mohammed MA. A comprehensive review of artificial intelligence approaches in omics data processing: evaluating progress and challenges. Int J Math Stat Comput Sci. 2024;2:114–67. doi: 10.59543/ijmscs.v2i.8703.

[14]   Mohammed MA, Ali AM. Enhanced cancer subclassification using multi-omics clustering and quantum cat swarm optimization. Iraq J Comput Sci Math. 2024;5(3):37. doi: 10.52866/ijcsm.2024.05.03.035.

[15]   Ahmed IA, Senan EM, Shatnawi HSA. Analysis of histopathological images for early diagnosis of oral squamous cell carcinoma by hybrid systems based on CNN fusion features. Int J Intel Syst. 2023;2023:2662719. doi: 10.1155/2023/2662719.

[16]   Haq IU, Ahmad M, Assam M, Ghadi YY, Algarni A. Unveiling the future of oral squamous cell carcinoma diagnosis: an innovative hybrid AI approach for accurate histopathological image analysis. IEEE Access. 2023;11:11281–90. doi: 10.1109/ACCESS.2023.3326152.

[17]   Albalawi E, Thakur A, Ramakrishna MT, Bhatia Khan S, SankaraNarayanan S, Almarri B, et al. Oral squamous cell carcinoma detection using EfficientNet on histopathological images. Frontiers in Medicine. 2024;10:1349336. doi: 10.3389/fmed.2023.1349336.

[18]   Nandini C, Basha S, Agarawal A, Neelampari RP, Miyapuram KP, Nileshwariba RJ. Deep learning approach to detect high-risk oral epithelial dysplasia: A step towards computer-assisted dysplasia grading. Adv Human Biol. 2023;13(1):57–60. doi: 10.4103/aihb.aihb_30_22.

[19]   Peng J, Xu Z, Dan H, Li J, Wang J, Luo X, et al. Oral epithelial dysplasia detection and grading in oral leukoplakia using deep learning. BMC Oral Health. 2024;24(1):434. doi: 10.1186/s12903-024-04191-z.

[20]   Camalan S, Mahmood H, Binol H, Araujo ALD, Santos-Silva AR, Vargas PA, et al. Convolutional neural network-based clinical predictors of oral dysplasia: Class activation map analysis of deep learning results. Cancers. 2021;13(6):1291. doi: 10.3390/cancers13061291.

[21]   Ribeiro-de Assis MCF, Soares JP, de Lima LM, de Barros LAP, Gräo-Velloso TR, Krohling RA, et al. NDB-UFES: An oral cancer and leukoplakia dataset composed of histopathological images and patient data. Data Brief. 2023;48:109128. doi: 10.1016/j.dib.2023.109128.

[22]   Reinhard E, Adhikhmin M, Gooch B, Shirley P. Color transfer between images. IEEE Comput Graph Appl. 2001;21(5):34–41. doi: 10.1109/38.946629.

[23]   Li Y, Wu J, Wu Q. Classification of breast cancer histological images using multi-size and discriminative patches based on deep learning. IEEE Access. 2019;7:21400–8. doi: 10.1109/ACCESS.2019.2898044.

[24]   Liu K, Qin S, Ning J, Xin P, Wang Q, Chen Y, et al. Prediction of primary tumor sites in spinal metastases using a ResNet-50 convolutional neural network based on MRI. Cancers. 2023;15(11):2974. doi: 10.3390/cancers15112974.

[25]   Simonyan K. Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learn Represent ICLR 2015-Conf track proc. Published online. 2015. p. 1. doi: 10.48550/arXiv.1409.1556.

[26]   He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016. doi: 10.1109/CVPR.2016.90.