

# An Enhanced Histopathology Analysis: An AI- Based System for Multiclass Grading of Oral Squamous Cell Carcinoma and Segmenting of Epithelial and Stromal Tissue

---

Musulín, Jelena; Štífanić, Daniel; Zulijani, Ana; Čabov, Tomislav;  
Dekanić, Andrea; Car, Zlatan

Source / Izvornik: **Cancers**, 2021, 13, 1 - 21

Journal article, Published version

Rad u časopisu, Objavljena verzija rada (izdavačev PDF)

<https://doi.org/10.3390/cancers13081784>

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:271:015974>

Rights / Prava: [Attribution 4.0 International](#)/[Imenovanje 4.0 međunarodna](#)

Download date / Datum preuzimanja: **2025-04-01**







Repository / Repozitorij:

[Repository of the University of Rijeka, Faculty of  
Dental Medicine](#)



## Article

# An Enhanced Histopathology Analysis: An AI-Based System for Multiclass Grading of Oral Squamous Cell Carcinoma and Segmenting of Epithelial and Stromal Tissue

Jelena Musulin <sup>1</sup>, Daniel Štifanić <sup>1,\*</sup>, Ana Zulijani <sup>2</sup>, Tomislav Čabov <sup>3,\*</sup>, Andrea Dekanić <sup>4,5</sup> and Zlatan Car <sup>1</sup>

- <sup>1</sup> Faculty of Engineering, University of Rijeka, Vukovarska 58, 51000 Rijeka, Croatia; jmusulin@riteh.hr (J.M.); car@riteh.hr (Z.C.)
- <sup>2</sup> Department of Oral Surgery, Clinical Hospital Center Rijeka, Krešimirova Ul. 40, 51000 Rijeka, Croatia; ana.zulijani@sz.uniri.hr
- <sup>3</sup> Faculty of Dental Medicine, University of Rijeka, Krešimirova Ul. 40, 51000 Rijeka, Croatia
- <sup>4</sup> Department of Pathology and Cytology, Clinical Hospital Center Rijeka, Krešimirova Ul. 42, 51000 Rijeka, Croatia; andrea.dekanic@medri.uniri.hr
- <sup>5</sup> Faculty of Medicine, University of Rijeka, Ul. Braće Branchetta 20/1, 51000 Rijeka, Croatia
- \* Correspondence: dstifanic@riteh.hr (D.Š.); tomislav.cabov@fdmri.uniri.hr (T.Č.)

**Simple Summary:** An established dataset of histopathology images obtained by biopsy and reviewed by two pathologists is used to create a two-stage oral squamous cell carcinoma diagnostic AI-based system. In the first stage, automated multiclass grading of OSCC is performed to improve the objectivity and reproducibility of histopathological examination. Furthermore, in the second stage, semantic segmentation of OSCC on epithelial and stromal tissue is performed in order to assist the clinician in discovering new informative features. Proposed AI-system based on deep convolutional neural networks and preprocessing methods achieved satisfactory results in terms of multiclass grading and segmenting. This research is the first step in analysing the tumor microenvironment, i.e., tumor-stroma ratio and segmentation of the microenvironment cells.

**Abstract:** Oral squamous cell carcinoma is most frequent histological neoplasm of head and neck cancers, and although it is localized in a region that is accessible to see and can be detected very early, this usually does not occur. The standard procedure for the diagnosis of oral cancer is based on histopathological examination, however, the main problem in this kind of procedure is tumor heterogeneity where a subjective component of the examination could directly impact patient-specific treatment intervention. For this reason, artificial intelligence (AI) algorithms are widely used as computational aid in the diagnosis for classification and segmentation of tumors, in order to reduce inter- and intra-observer variability. In this research, a two-stage AI-based system for automatic multiclass grading (the first stage) and segmentation of the epithelial and stromal tissue (the second stage) from oral histopathological images is proposed in order to assist the clinician in oral squamous cell carcinoma diagnosis. The integration of Xception and SWT resulted in the highest classification value of 0.963 ( $\sigma = 0.042$ ) AUCmacro and 0.966 ( $\sigma = 0.027$ ) AUCmicro while using DeepLabv3+ along with Xception\_65 as backbone and data preprocessing, semantic segmentation prediction resulted in 0.878 ( $\sigma = 0.027$ ) mIOU and 0.955 ( $\sigma = 0.014$ ) F1 score. Obtained results reveal that the proposed AI-based system has great potential in the diagnosis of OSCC.

**Keywords:** AI-based system; data preprocessing; histopathological images; oral squamous cell carcinoma



**Citation:** Musulin, J.; Štifanić, D.; Zulijani, A.; Čabov, T.; Dekanić, A.; Car, Z. An Enhanced Histopathology Analysis: An AI-Based System for Multiclass Grading of Oral Squamous Cell Carcinoma and Segmenting of Epithelial and Stromal Tissue. *Cancers* **2021**, *13*, 1784. <https://doi.org/10.3390/cancers13081784>

Academic Editor: Max Witjes

Received: 13 March 2021

Accepted: 7 April 2021

Published: 8 April 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Cancer is a major public health problem and the second leading cause of death in the developed world. Oral cancer (OC; Supplementary Materials Table S1 lists descriptions of the abbreviations and acronyms used) is among the ten most common cancers in Europe

and the United States, where more than 90% are squamous cell carcinomas [1,2]. According to the GLOBOCAN database, there were an estimated 377,713 new cases diagnosed, and 177,757 deaths were recorded in 2020 [3]. Despite diagnostic and therapeutic development in OC patients', mortality and morbidity rates remain high with no advancement in the last 50 years, largely due to late-stage diagnosis when tumor metastasis has occurred [4]. Often, oral squamous cell carcinoma (OSCC) arises from pre-existing lesions of oral mucosa with increased risk for malignant transformation in cancer. Diagnosis and management at the "precancerous" stage and early detection of OSCC improves survival rates and morbidity accompanying the treatment of OSCC [5]. Typically, OSCC is treated primarily by surgical resection with or without adjuvant radiation, which has a major impact on patient quality of life [6]. Until now great strides have been made in understanding the complex process in carcinogenesis, but no reliable tool for prognostic prediction has been found. The tumor-node-metastasis (TNM) staging is widely used in the prognosis, treatment plan, and prediction outcomes of oral cancer in patients with OSCC. However, the inability to incorporate clinical characteristics and individual characteristics of the patient such as lifestyle behaviours reflects on the limitation of the TNM staging in prognostic prediction [7]. Currently, standard methods for detection of oral cancer and the gold standard are clinical examination, conventional oral examination (COE), and histopathological evaluation following biopsy, which can detect cancer in the stage of established lesions with significant malignant changes [8]. However, the main problem in using histopathological examination for tumor differentiation, and like prognostic factor, is the subjective component of the examination, respectively inter- and intra-observer variability [9]. Improving objectivity and reproducibility, respectively reducing inter- and intra-observer variability using Artificial Intelligence (AI) algorithms could directly impact patient-specific treatment intervention by identifying patients' outcomes. Furthermore, it could assist the pathologist in terms of reducing the load of manual inspections as well as making fast decisions with higher precision.

Recently, many AI algorithms have been proven successful in medical image analysis [10–13] as well as other various fields of science and technology [14–16]. Medical image analysis is one of the areas where histological tissue patterns are used with computer-aided image analysis to facilitate detection and classification of disease [17]. Numerous studies describe the successful implementation of AI algorithms for the detection, staging, and prognosis of all types of cancer. Based on histopathological images Sharma and Mehra proposed a model for automatic multiclass classification of breast cancer using a deep learning approach. Among the different combinations of classifiers VGG16, a convolutional neural network (CNN) architecture which consists of 16 layers that have weights, in a combination with support-vector machine (SVM) classifier resulted in the highest micro- and macro-average (0.93) [18]. Wu et al. used a deep CNN framework in order to predict the risk of lung cancer. CNNs are considered as the most popular deep learning approach in terms of analysing visual imagery. In their study, authors presented that features extracted from histopathological images can be used for the prediction of lung cancer recurrence [19]. By analysing histopathological images Tabibu et al. presented how deep learning algorithms can be used for pan-renal cell carcinoma classification as well as prediction of survival outcome. After breaking the multiclass classification task into multiple binary tasks, deep CNN in a combination with directed acyclic graph (DAG) and SVM proved to be the best performing method and resulted in 0.93 slide-wise AUC [20].

In this research, a two-stage AI-based system for automated multiclass classification of OSCC into three classes, in the first stage, and semantic segmentation of the epithelial and stromal tissue in the second stage is proposed.

The main contributions in this research are as follows:

- The first stage of an AI-based system for multiclass grading of OSCC which can potentially improve objectivity and reproducibility of histopathological examination, as well as reduce the time necessary for pathological inspections.
- The second stage of an AI-based system for segmentation of tumor on epithelial and stromal regions which can assist the clinician in discovering new informative features.

It has great potential in the quantification of qualitative clinic-pathological features in order to predict tumor invasion and metastasis.

- A new preprocessing methodology based on the stationary wavelet transform (SWT) is proposed to enhance high-frequency components in the case of multiclass classification and to extract low-level features in the case of semantic segmentation. This approach allows more effective predictions and improves the robustness of the entire AI-based system.

#### *Related Work*

Many studies have used various models, techniques, and methodologies in order to develop AI-solutions for classification and segmentation of oral cancer.

Ariji et al. demonstrated the use of artificial intelligence for the diagnosis of lymph node metastasis in patients with OSCC. The dataset consisted of 441 computed tomography images in total. The performance of the proposed system resulted in 75.4% sensitivity, 81% specificity, and 78.2% accuracy. The obtained results were similar to those found by the radiologists which prove that the proposed AI system can be valuable for diagnostic support [21]. Halicek et al. presented deep learning methods for detection of OSCC. Their dataset was collected from Emory University Hospital Midtown and consisted of 293 tissue specimens. The tissue specimens were imaged with reflectance-based HSI and autofluorescence imaging. With HSI-based methods, OSCC detection may be obtained in less than 2 min with AUC upwards of 0.80 to 0.90 [22]. Horie et al. used CNN to diagnose outcomes of esophageal cancer. The dataset was collected at the Cancer Institute Hospital (Toyko, Japan) and consisted of 8428 training images and 1118 test images. The proposed method could detect oral esophageal cancer with a sensitivity of 98% and distinguish esophageal cancer from advanced cancer with an accuracy of 98% [23]. Tamasiro et al. showed the diagnostic ability of an AI-based system for the detection of oral pharyngeal cancer. The dataset consists of 5400 training images and 1912 validation images, obtained from the Cancer Institute Hospital. CNNs detected pharyngeal cancer with high sensitivity (85.6%) [24]. Jeyaraj et al. used a partitioned deep CNN to detect oral cancer in the early stages. The performance of this partitioned deep CNN resulted in 94% sensitivity, 91% specificity, and 91.4% accuracy. After comparison with other medical image classification algorithms, from obtained results, it can be concluded that the quality of diagnosis was increased [25]. Bhandari et al. proposed a system that consists of a CNN with a modified loss function to minimize the error in predicting and classifying oral cancer. Dataset contains magnetic resonance imaging (MRI) scans which were used for training and testing the proposed system. Based on the results, the proposed system achieved an accuracy of 96.5% [26]. In order to detect oral cancer in the early stages, Xu et al. established a three-dimensional CNN algorithm. Their results show that 3DCNNs outperform 2DCNNs in identifying benign and malignant lesions [27]. Welikala et al. presented the automated detection and classification of oral cancer in the early stages. Images were gathered from clinical experts as a part of the MeMoSA project. Image classification algorithm based on deep CNN resulted in 87.07% F1 score, and object detection with Faster R-CNN resulted in 41.18% F1 score [28].

The literature reveals that most of the researchers have applied CNN in the retrospective study to detect and classify oral cancer. It can be observed that most of the classification tasks were binary classification, where the features from the image like colour, shape, and texture are used. Even though, CNN is a powerful visual model for image recognition tasks, some researchers used deep CNNs which outperform conventional CNN in classification tasks.

Chan et al. proposed an innovative deep CNN combined with a texture map for oral cancer detection. The proposed model consists of two collaborative branches, one to perform oral cancer detection, and a second one to perform region of interest (ROI) marking and semantic segmentation. The experimental results of detection, based on a wavelet transform, are up to 96.87% for sensitivity and 71.29% for specificity [29]. Fraz et al. demon-

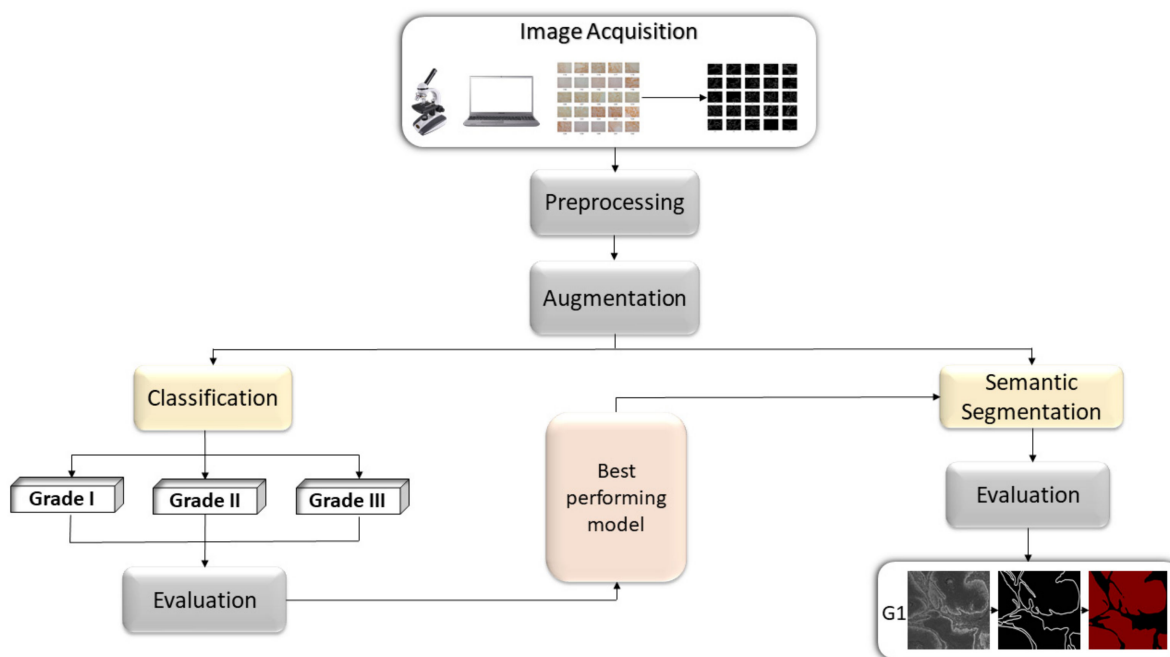
strated a network for simultaneous segmentation of microvessels and nerves of oral cancer. The dataset used consisted of 7780 H&E-stained histology images, size  $512 \times 512$  pixels, extracted from 20 WSIs of oral carcinoma tissue. The proposed block-based pyramid pooling deep neural network outperforms other deep CNN for semantic segmentation [30].

Most of the segmentation tasks were performed using deep CNN architectures on histopathological images. A shortcoming of the prementioned studies is that they were trained to determine a binary state of carcinoma or non-carcinoma. Based on a literature study only Das et al. proposed a deep learning model for the classification of cells into multiple classes in epithelial tissue of OSCC. The dataset consisted of image patches derived from whole slide biopsy images. Proposed CNN model resulted in 97.5% accuracy [31].

According to the presented studies different AI approaches have proven successful for clinical analysis [21–29]. However, in the terms of histological analysis of OC, fewer studies have been conducted since histopathological examination is highly invasive [30,31]. Exhaustive literature study reveals that, at the time when this research was performed, no work has been done on multiclass grading along with segmenting of OSCC using whole-slide histopathology images obtained by biopsy and stained with marker protein.

## 2. Materials and Methods

The overall workflow of the research can be described as follow; first, the original data will be augmented and used as an input variable in order to perform multiclass classification by utilizing AI-based models. After evaluating the performance of the models, the obtained results will be compared, and the best performing model will be selected. Second, the input data will be decomposed using stationary wavelet transform in order to obtain wavelet coefficients. The obtained coefficients will be preprocessed and the data will be reconstructed. Afterwards, the data preprocessed with SWT will be used to train the selected model, and the obtained results will be compared. Third, the selected model will be used as Deeplabv3+ backbone, by which, semantic segmentation into two classes will be performed. Fourth, the impact of data preprocessing on the model performance and robustness will be examined. Lastly, the best performing configuration in terms of semantic segmentation will be used to show predictions of the epithelial and stromal tissue on three samples from new, unseen data. The overview of the proposed framework is given in Figure 1.



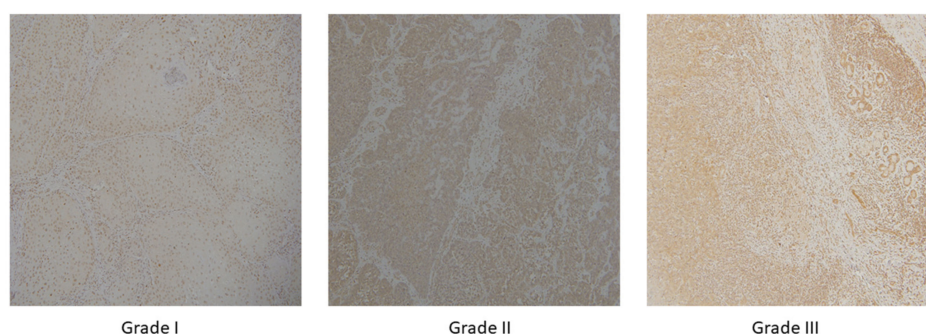
**Figure 1.** Block diagram representation of the proposed methodology.

### 2.1. Dataset Description

For this research, 322 histology images with  $768 \times 768$ -pixel size have been used to create a dataset. The formalin-fixed, paraffin-embedded oral mucosa tissue blocks of histopathologically reported cases of OSCC were retrieved from the archives of the Clinical Department of Pathology and Cytology, Clinical Hospital Center in Rijeka. Sample slides were reviewed by two unbiased pathologists and classified following the 4th edition of the World Health Organization (WHO) classification of Head and Neck tumors [32] and 8th edition of the AJCC Cancer Staging Manual [33].

The paraffin blocks tissues were selected from the patients with primary invasive squamous cell carcinomas, where the invasion of the epithelium of the oral cavity through the basement membrane was confirmed with certainty. All patients who received preoperative chemo- or radiation-therapy, as well as patients with in situ carcinoma and relapsed or second primary OSCC were excluded from the research.

Kappa coefficient was used to determine the degree of agreement between the pathologists. The value of Kappa coefficient was found to be 0.94. In accordance with the aforementioned classification, images have been divided into three classes: well-differentiated (grade I), moderately differentiated (grade II), and poorly differentiated (grade III) OSCC, as shown in Figure 2. Additionally, the segmentation masks were prepared by a health professional and validated by another independent pathologist.



**Figure 2.** OSCC group of well-differentiated OSCC (**grade I**), moderately differentiated OSCC (**grade II**) and poorly differentiated OSCC (**grade III**) with magnification  $\times 10$ .

Briefly,  $4 \mu\text{m}$  sized paraffin-embedded tissue sections were deparaffinized in tissue clear solution, rehydrated in a series of different concentrations of alcohol and stained using monoclonal mouse anti-MT I + II antibody ((clone E); DAKO, Santa Clara, CA, USA; diluted 1:100 in PBS with 1% BSA) and polyclonal rabbit anti-megalin (Santa Cruz Biotechnology, Dallas, TX, USA; diluted 1:100 in PBS with 1% BSA) using a standard protocol [34]. Immunoreaction was visualized by adding peroxidase substrate solution containing diaminobenzidine (DAB). Slides were afterwards stained with hematoxylin (Sigma-Aldrich, Munich, Germany), dehydrated and mounted with Entellan (Sigma-Aldrich).

Images were captured using the light microscope (Olympus BX51, Olympus, Tokyo, Japan) equipped with a digital camera (DP50, Olympus) and transmitted to a computer by CellF software (Olympus). Furthermore, images were captured at  $10 \times$  objective lenses.

Corresponding clinic-pathological reports of the patients were collected and used for pTMN classification [33]. Patient demographic information included age at the time of diagnosis, sex, smoking. Patients were adults, where the median age was 64, and 69% of them were smokers. As seen in Table 1, more patients were male (65%) while 35% were female. The least number of patients were diagnosed with grade III (17%) whereas 50% were diagnosed with grade I. In more patients (54%) presence of metastases in the lymph nodes were excluded.

Deep CNNs are heavily reliant on a large number of samples in order to achieve satisfactory performance and avoid overfitting. Since domains, such as medical image analysis, do not have access to a large number of samples very often, it is necessary

to perform augmentation techniques, by which, the size and quality of the data can be significantly increased [35]. Due to the aforementioned neural network demand and limited availability of the data, augmentation techniques are performed to artificially increase the number of samples.

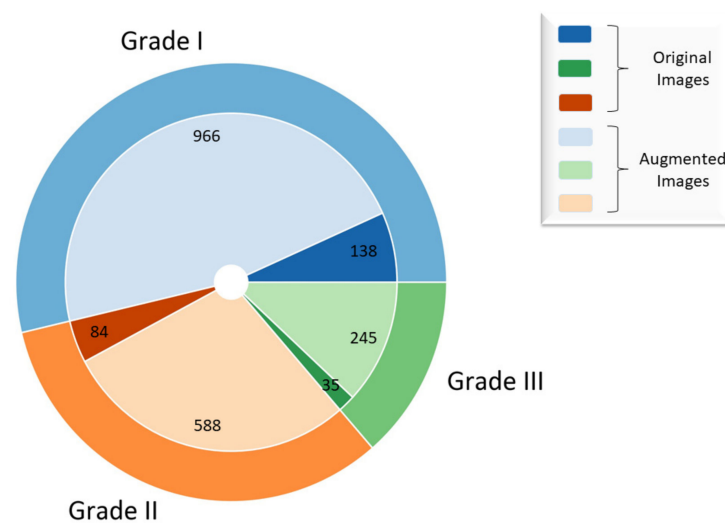
**Table 1.** Characteristic of the patients include sex, age, smoking habits, presence of metastases in the lymph nodes, and histological grade of carcinoma.

Characteristic of the Patients		%
Sex	F	35
	M	65
Age	To 49	6
	50–59	13
	60–69	58
	+70	23
Smoking	Y	69
	N	31
Lymph Node Metastases	Y	46
	N	54
Histological Grade (G)	I	50
	II	33
	III	17

Geometrical transformations used for the augmentation procedure are: 90 degrees anticlockwise rotation, 180 degrees anticlockwise rotation, 270 degrees anticlockwise rotation, horizontal flip, horizontal flip combined with 90 degrees anticlockwise rotation, vertical flip, and vertical flip combined with 90 degrees anticlockwise rotation. Since newly generated data are variations of the original data, the augmentation procedure is utilized only for the creation of the training samples, thereby, testing samples are not augmented.

Due to the high-imbalance of OSCC classes, stratified 5-fold cross-validation is used to estimate the performance of AI-based models. This way, the representation of each class across each test fold is approximately the same [36]. Accordingly, within each fold, the first part (approximately 80% of each class) is augmented and used for model training, while the second part (approximately 20% of each class) is used to evaluate the performance of the trained models.

By utilizing the aforementioned transformations, a new training set with additional 1799 images has been created, which gives a total of 2056 images as shown in Figure 3.



**Figure 3.** Representation of original and augmented dataset.

## 2.2. Preprocessing Method Based on Stationary Wavelet Transform and Mapping Function

The wavelet transform (WT) is a powerful technique commonly used in data preprocessing [37]. Applying WT, data can be represented at different scales and frequencies. Furthermore, it is a useful tool for describing the image in multiple resolutions. Wavelet transform of signal  $x(t)$  can be defined as [38]:

$$X(\tau, a) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t) \psi^* \left( \frac{t - \tau}{a} \right) dt, \quad (1)$$

where  $a$  is the dilation,  $\psi$  is analysing wavelet, and  $\tau$  is translation parameter. By performing WT, low-frequency components (i.e., approximation coefficients) and high-frequency components (i.e., detail coefficients) can be obtained [39]. The discrete wavelet transform (DWT) of signal  $x[m]$  can be calculated as follows [38]:

$$X[k, l] = 2^{-\frac{k}{l}} \sum_{m=-\infty}^{\infty} x[m] \psi \left[ 2^{-k} m - l \right]. \quad (2)$$

When dealing with images, DWT is applied in each dimension separately; therefore, the image is divided into four subbands LL, LH, HL, and HH where LL represents approximation coefficients while LH, HL, and HH represent detail coefficients of an image [40]. DWT reduces the computation time and can be implemented easily, but it also suffers in terms of shift-invariance and decimation. In order to overcome aforementioned drawbacks, the research utilises stationary wavelet transform (SWT) by which histopathology images can be decomposed.

The advantages of SWT are as follows [41]:

- no decimation step—provides redundant information,
- better time-frequency localization, and
- translation-invariance.

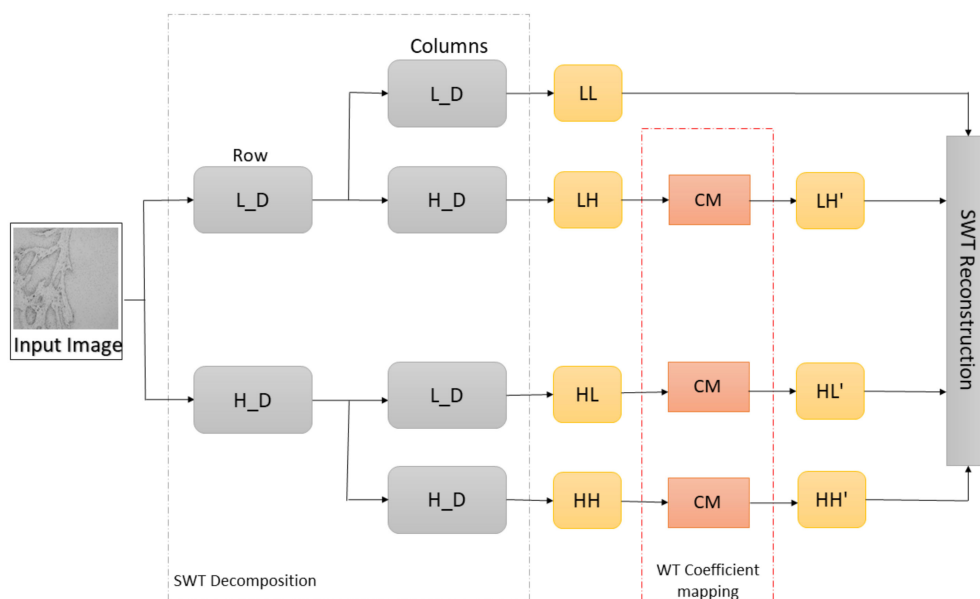
After the SWT decomposition process, obtained coefficients are weighted using a mapping function. This way important features of an image can be further enhanced. The following considerations are taken into account when determining the mapping function; First-coefficient mapping is performed only on detail coefficients. Second—details with high, as well as details with low coefficient values preserve valuable information, thereby, they are heavily weighted. Wavelet coefficient mapping function can be mathematically defined as follows:

$$y_{i,j} = aw_{i,j}^3 + bw_{i,j}^2 + cw_{i,j} + d, \quad (3)$$

where  $a$ ,  $b$ ,  $c$ , and  $d$  represent constants,  $w_{i,j}$  is an input coefficient, while  $y_{i,j}$  is a coefficient after mapping. After the coefficient mapping process, the SWT reconstruction is performed with approximation and weighted SWT coefficients in order to obtain an enhanced image. The process of SWT decomposition-coefficient mapping-SWT reconstruction is shown in Figure 4.

The quality of weighted coefficients directly depends on mapping function constants and wavelet function; therefore, a careful selection of these parameters is necessary. Additionally, the selection of the parameters also depends on the type of input data and expected output. More traditional approaches for parameter determination such as random search or grid search can sometimes be infeasible since evaluating each point in large search-spaces is extremely costly (computationally) [42]. These methods do not consider evaluated performances of past iterations when selecting the next configuration of parameters, often resulting in spending time on evaluating function with a poor selection of parameters. In contrast, the Bayesian approach for determining parameters of mapping function selects the next configuration of parameters based on results obtained in past iterations [43]. This way, convergence to the best solution can be achieved in fewer iterations, thus outperforming more traditional methods.





**Figure 4.** Representation of SWT decomposition, wavelet coefficient mapping, and SWT reconstruction (L\_D—low pass filter, H\_D—high pass filter, LL—approximation coefficients, LH—horizontal coefficients, HL—vertical coefficients, HH—diagonal coefficients, and CM—coefficient mapping function).

In order to determine optimal values of wavelet coefficient mapping function constants (a, b, c, d) and wavelet function, Bayesian optimization has been used. The domain of mapping function constants over which to search is defined and shown in Table 2.

**Table 2.** List of hyperparameters used in the Bayesian optimization process.

Hyperparameter	Possible Parameters
a	0–0.1
b	0–0.1
c	0–0.1
d	0.001–1
Wavelet function	Haar, sym2, db2, bior1.3

### 2.3. AI-Based Models

In this subsection, a brief overview of deep CNN architectures suitable for OSCC classification problem is given.

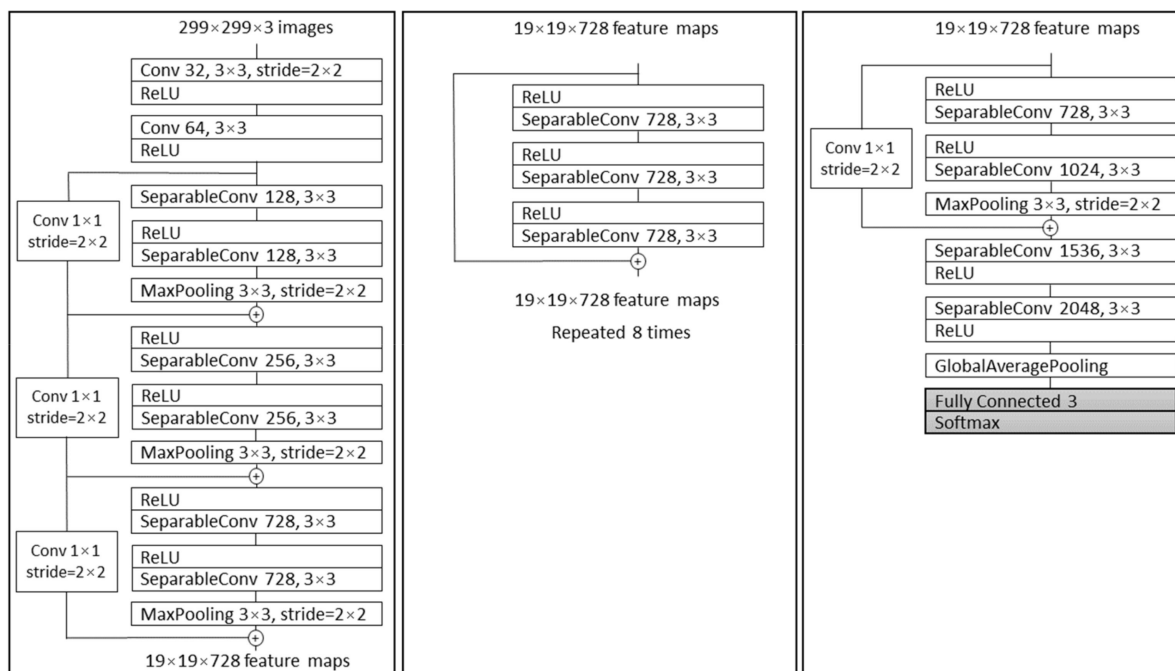
#### 2.3.1. Xception

Chollet demonstrated a novel architecture named Xception, based on Inception V3. Compared to Inception V3 on ImageNet, Xception resulted in larger performance improvement [44]. In a conventional CNN, convolutional layers search through depth and space for correlation. Xception takes few steps further with mapping the spatial correlations for each output channel separately and performing  $1 \times 1$  depth-wise convolution to capture cross-channel correlation. Xception architecture consists of 36 convolutional layers which are structured in 14 modules [44]. All the modules have linear residual connections around them, not including the first and last module, as shown in Figure 5.

#### 2.3.2. ResNet50 and –101

As neural networks become deeper, they become difficult to train due to the notorious vanishing gradient problem. In order to ease the training of deep neural networks, He et al. propose a residual network (ResNets) [45]. They refined the residual block along with the pre-activation variant of the residual block where through the shortcut connections vanishing gradients can flow unimpededly to any other previous layer. In ResNet50 architecture, every 2-layer block is replaced in the 34-layer network with a 3-layer bottleneck block, which results

in 50 layers, while ResNet101 architecture is constructed using more 3-layer blocks as shown in Table 3. The number of parameters for ResNet50 and ResNet101 totals 23,888,771 and 42,959,235, respectively. On the ImageNet dataset, He et al. proved that ResNets with achieved 3.57% error outperforms other architectures on the ILSVRC classification task [45].



**Figure 5.** The Xception architecture; first, the data propagate through entry flow (first box), then through middle flow (second box) and repeats eight times. In the end, data propagate through the third box which represents exit flow [44].

**Table 3.** ResNet50 and ResNet101 architecture representation.

Layer	Output	Layers	ResNet50	ResNet101
			Number of Repeating Layers	
Conv1	112 × 112	7 × 7, 64, stride 2	×1	×1
		3 × 3 max pool, stride 2	×1	×1
Conv2_x	56 × 56	1 × 1, 64	×3	×3
		3 × 3, 64		
		1 × 1, 256		
Conv3_x	28 × 28	1 × 1, 128	×4	×4
		3 × 3, 128		
		1 × 1, 512		
Conv4_x	14 × 14	1 × 1, 256	×6	×23
		3 × 3, 256		
		1 × 1, 1024		
Conv5_x	7 × 7	1 × 1, 512	×3	×3
		3 × 3, 512		
		1 × 1, 2048		
	1 × 1	Flatten 3-d Fully Connected Softmax	×1	×1

### 2.3.3. MobileNetv2

Sandler et al. introduced in their article [46] the mobile architecture called MobileNetv2 which enhances the performance of mobile models. It builds on MobileNetv1’s ideas by utilizing depthwise separable convolution as efficient building blocks. Compared to conventional residual models that use an extended input representation, MobileNetv2 as input to the residual block use thin bottleneck layers. The MobileNetV2 architecture includes the initial fully convolution layer with 32 filters and 19 residual bottleneck layers. The detailed architecture structure is shown in Table 4.

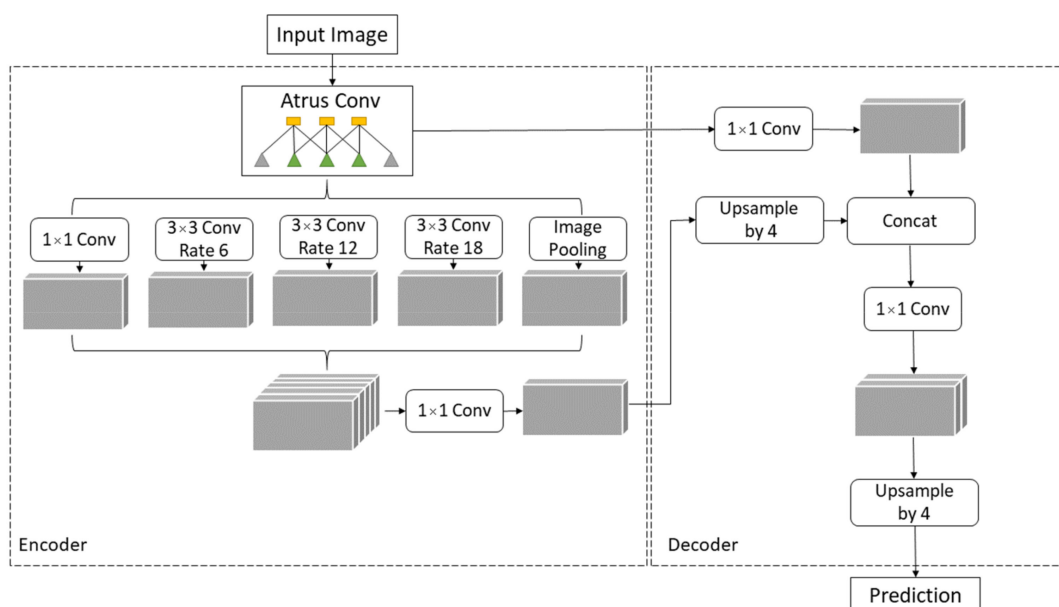
**Table 4.** MobileNetv2 architecture; each row represents a sequence of at least 1 identical layer, repeated n times. The number c of output channels is the same for each layer in the same sequence. The first layer of each sequence consists of a stride s while all the rest use stride 1. The expansion factor t is used for the input size.

Input	Operator	Expansion Factor (t)	Number of Output Channels (c)	Repeating Number (n)	Stride (s)
224 × 224 × 3	conv2d	-	32	1	2
112 × 112 × 32	bottleneck	1	16	1	1
112 × 112 × 16	bottleneck	6	24	2	2
56 × 56 × 24	bottleneck	6	32	3	2
28 × 28 × 32	bottleneck	6	64	4	2
14 × 14 × 64	bottleneck	6	96	3	1
14 × 14 × 96	bottleneck	6	160	3	2
7 × 7 × 160	bottleneck	6	320	1	1
7 × 7 × 320	conv2d 1 × 1	-	1280	1	1
7 × 7 × 1280	avgpool 7 × 7	-	-	1	-
1 × 1 × 1280	fully connected (Softmax)	-	3	-	-

MobileNetv2 for classification task was compared with MobileNetv1, NASNet-A, and ShuffleNet on the ImageNet dataset. The results show that presented architecture with specific design variations improves the state-of-the-art for a wide range of performance points [45].

#### 2.4. DeepLabv3+

The process of assigning a semantic label to each part of the image is called semantic segmentation. Chen et al. proposed the DeepLab system as state-of-the-art at PASCAL VOC 2012 semantic segmentation task and based on the experimental results they achieved 79.7% mIOU on the test set [47]. Later, Chen et al. presented the DeepLabv3 system that improves their previous versions of DeepLab. Without DenseCRF postprocessing, the presented system attains a performance of 85.7% mIOU on the PASCAL VOC 2012 dataset [48]. DeepLabv3+ is an extended version of DeepLabv3 that includes a decoder module to refine the result of segmentation. Due to Chen et al.’s paper, DeepLabv3+ without any postprocessing achieves a performance of 89% mIOU on the PASCAL VOC 2012 dataset [49]. DeepLab and DeepLabv3 use Atrous Spatial Pyramid Pooling (ASPP) to encode multi-scale contextual information while DeepLabv3+ combines encoder-decoder pathway and ASPP, as shown in Figure 6, intending to achieve more precise delineation of object boundaries [50].



**Figure 6.** DeepLabv3+ Encoder-Decoder architecture.

The architectures described in Section 2.3 (Xception, ResNet50, ResNet101, MobileNetv3) can be used as DeepLabv3+ backbones.

### 2.5. Evaluation Criteria

In order to evaluate and analyse the obtained results, it is necessary to describe evaluation metrics. Statistical measures such as micro- and macro- area under the curve (AUC) are adopted to evaluate the classification performance of models. In the second stage, semantic segmentation performances are evaluated using mean intersection-over-union (mIOU), Dice coefficient (F1), accuracy (ACC), precision, sensitivity, and specificity.

The AUC is an evaluation metric for calculating the performance of the binary classifier. To extend AUC to multiclass classification, it is necessary to conceptualize the problem as a binary classification problem, by using One vs. All technique, which means that one class is classified against all other classes. Micro averaging of true positive rate (TPR) calculates the number of correct classifications for each class and uses it as the numerator, while for the denominator it uses the total number of samples. Furthermore, fallout or false positive rate (FPR) calculates the ratio of incorrect classifications for each class and the total number of samples [51]. The mathematical representation of micro averaging is defined as follows

$$\text{TPR}_{\text{micro}} = \frac{\sum_{i=1}^k \text{TP}_i}{\sum_{i=1}^k \text{TP}_i + \sum_{i=1}^k \text{FN}_i} \quad (4)$$

and:

$$\text{FPR}_{\text{micro}} = \frac{\sum_{i=1}^k \text{FP}_i}{\sum_{i=1}^k \text{FP}_i + \sum_{i=1}^k \text{TN}_i} \quad (5)$$

by which,  $\text{AUC}_{\text{micro}}$  can be calculated. TP represents true positives, i.e., cases where the predicted and actual values are positive. TN represents true negatives, cases where the actual and predicted values are negative. False negatives (FN) capture cases when the prediction is negative and the actual value is positive. Furthermore, FP represents false positives, where the prediction is positive, and the actual value is negative [52].

Macro averaging for  $k$  classes compute the metrics individually for each class and averages results together.  $\text{AUC}_{\text{macro}}$  is based on the calculation of  $\text{TPR}_{\text{macro}}$  as well as  $\text{FPR}_{\text{macro}}$  and can be calculated as follows

$$\text{TPR}_{\text{macro}} = \frac{\sum_{i=1}^k \text{TPR}_i}{k} \quad (6)$$

and:

$$\text{FPR}_{\text{macro}} = \frac{\sum_{i=1}^k \text{FPR}_i}{k} \quad (7)$$

Higher values of  $\text{AUC}_{\text{macro}}$  and  $\text{AUC}_{\text{micro}}$  measure will result in better classification performance of the model.

One of the most used metrics for semantic segmentation is IOU, also known as the Jaccard Index. If the previously introduced terms TP, TN, FP, and FN are used, the Jaccard index can be defined as [53]:

$$\text{IOU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (8)$$

Since this research deals with multiple classes (more than one), mean IOU (mIOU) is used as a metric. It can be calculated as a ratio between a total number of IOU-s for each semantic class and total number of classes. Dice coefficient is positively correlated with the IOU. It is an overall measure of a model's accuracy and can be defined as [54]:

$$\text{F1} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \quad (9)$$

F1 score ranges in 0–1 where the model with low false positives and low false negatives performs well. Accuracy measure points out what percentage of the pixels in the image are assigned to the correct class, and can be calculated as follows [55]:

$$ACC = \frac{TP + TN}{TN + TP + FN + FP} \tag{10}$$

while precision shows the percentage of the results which are relevant and can be expressed as [54]:

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

Mathematically, sensitivity and specificity can be calculated as the following [56]

$$Sensitivity = \frac{TP}{TP + FN} \tag{12}$$

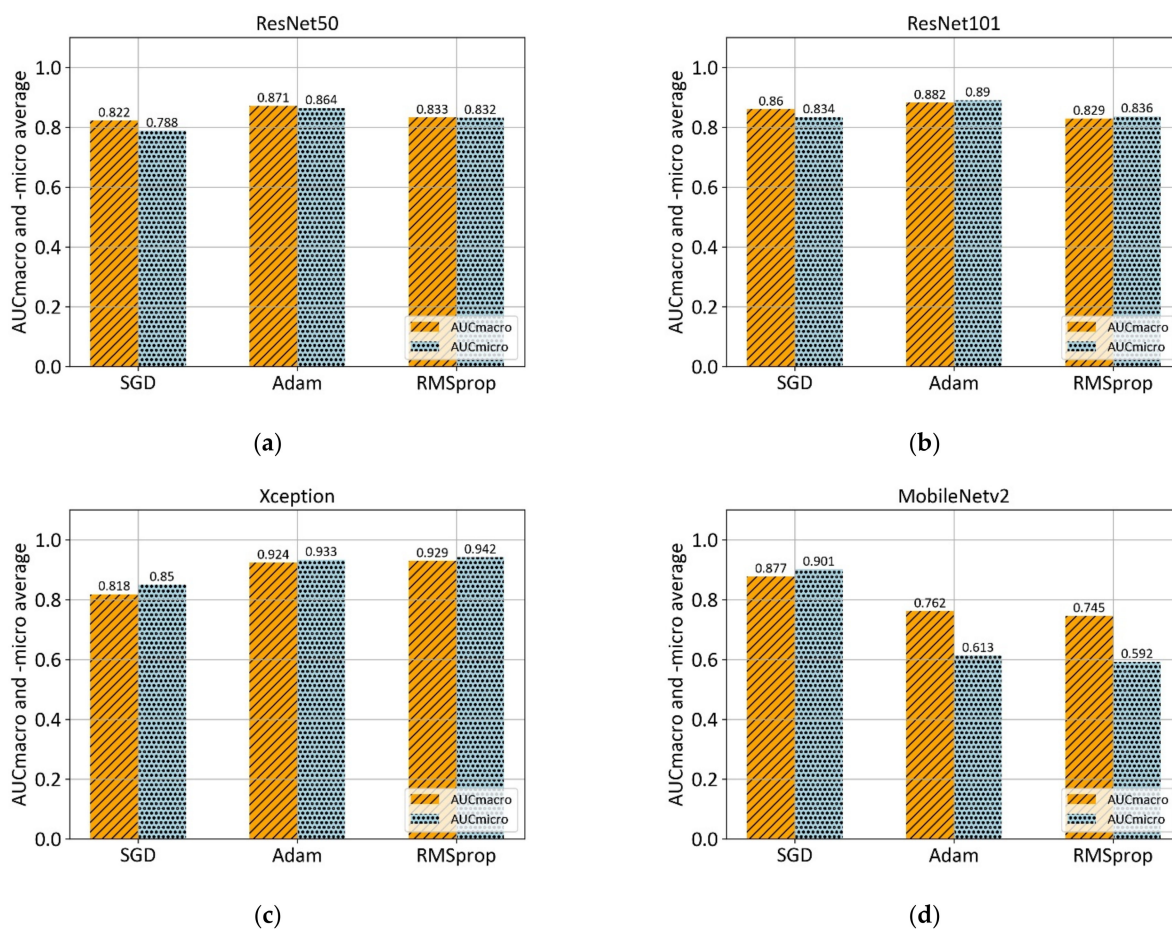
and:

$$Specificity = \frac{TN}{TN + FP} \tag{13}$$

Higher values of performance measures defined by Equations (8)–(13) mean better segmentation performance of the model.

### 3. Results

This section demonstrates the experimental results obtained at each step of the proposed methodology. The first experimental results are achieved with Xception, ResNet50, ResNet101, and MobileNetv2 architectures which are pre-trained on ImageNet. Each model architecture is trained with three optimizers: stochastic gradient descent (SGD), Adam, and RMSprop, as shown in Figure 7.



**Figure 7.** Comparison of mean AUC<sub>macro</sub> and -<sub>micro</sub> values of three different optimizers (SGD, ADAM, and RMSprop) on pre-trained models: (a) ResNet50; (b) ResNet101; (c) Xception; and (d) MobileNetv2.

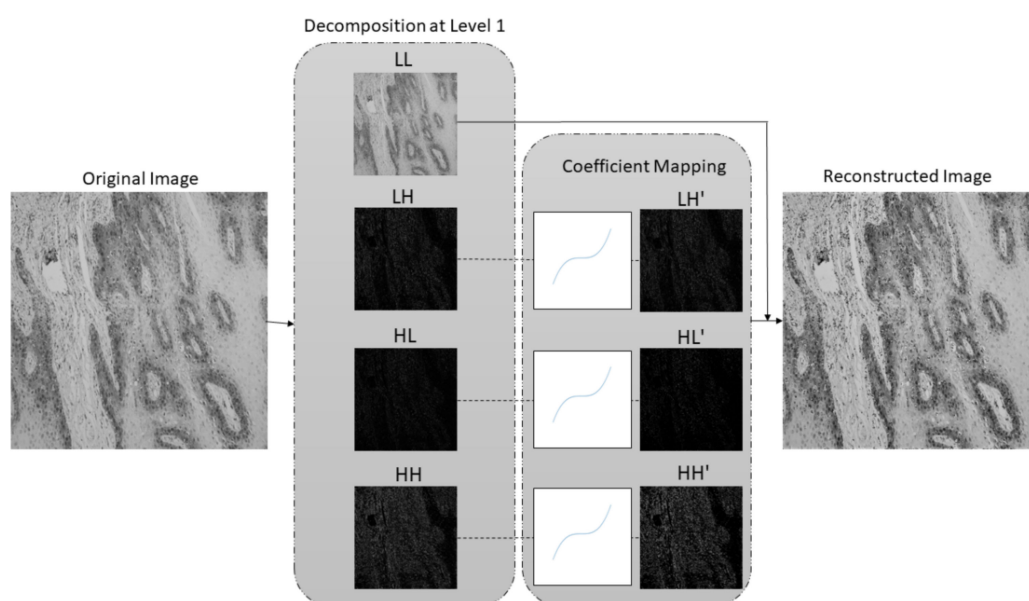
According to the 5-fold cross-validation results, the Adam optimizer achieves the highest  $AUC_{macro}$  and  $AUC_{micro}$  values in the case of ResNet50, ResNet101, and MobileNetv2 architectures. However, RMSprop optimizer in a combination with Xception architecture achieves the overall highest values of  $AUC_{macro}$ , and  $AUC_{micro}$ . Summarized mean values of performance measure along with corresponding standard deviation for each model architecture is shown in Table 5.

**Table 5.** Performance of different algorithms using  $AUC_{macro}$  and  $AUC_{micro}$  as evaluation metrics along with standard deviation ( $\sigma$ ).

Algorithm	$AUC_{macro} \pm \sigma$	$AUC_{micro} \pm \sigma$
ResNet50	$0.871 \pm 0.105$	$0.864 \pm 0.090$
ResNet101	$0.882 \pm 0.125$	$0.890 \pm 0.112$
Xception	$0.929 \pm 0.087$	$0.942 \pm 0.074$
MobileNetv2	$0.877 \pm 0.062$	$0.900 \pm 0.049$

The best results obtained in the first step were achieved when two additional layers were added at the end of the base Xception architecture. After the separable convolutional layer, and ReLU activation function, the global average pooling layer followed by the output layer with three neurons and Softmax activation function were added. This way, a classification of OSCC into three classes was enabled. The training process was divided into two stages; the first stage where only added layers were trainable while the others were frozen, and the second stage, where all layers were trainable except the last two added layers. In the first stage, the training process was performed with a learning rate of 0.001 and learning rate decay of  $1 \times 10^{-6}$ . The training process in the second stage was performed with a learning rate of 0.0001 and the same learning rate decay of  $1 \times 10^{-6}$ .

The second step of the proposed methodology is data preprocessing using Stationary Wavelet Transform, where original histopathology images were decomposed at level 1 using Haar, sym2, db2, and bior1.3 wavelet functions. After the decomposition process, high-frequency wavelet coefficients LH, HL, HH are weighted using mapping function defined by Equation (3), which resulted in new, modified LH', HL', HH' subbands. Modified subbands along with unmodified LL subband were used for SWT reconstruction in order to obtain input image for AI model, as shown in Figure 8.



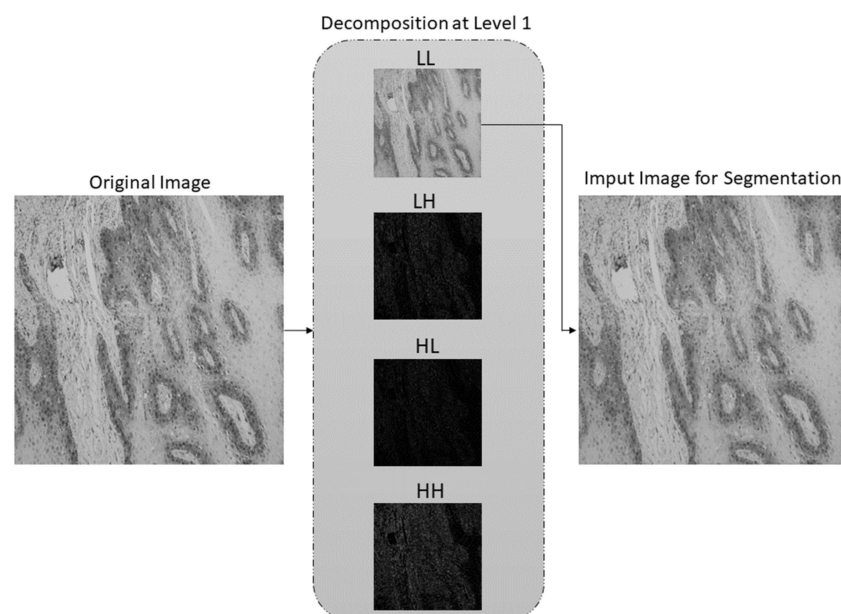
**Figure 8.** SWT decomposition at level 1 using Haar wavelet along with coefficient mapping, and SWT reconstruction.

By utilizing Bayesian optimization, the goal was to find optimal values of wavelet mapping function constants, by which, the maximal value of performance measure is achieved. In the case of this research,  $AUC_{\text{micro}}$  performance measure was monitored during the process of optimizing. Each Bayesian iteration involved data preprocessing with a defined set of mapping function constants, model training process, and performance evaluation. After 25 steps of random exploration and 20 steps of Bayesian optimization, the best performing constant configuration was obtained as shown in Table 6.

**Table 6.** Constants of coefficient mapping function obtained using Bayesian optimization along with corresponding 5-fold cross-validation performance.

Parameters				Xception + SWT		
a	b	c	d	Wavelet	$AUC_{\text{macro}} \pm \sigma$	$AUC_{\text{micro}} \pm \sigma$
0.0084	0.0713	0.0599	0.0566	sym2	$0.956 \pm 0.054$	$0.964 \pm 0.040$
0.0091	0.0301	0.0086	0.3444	db2	$0.963 \pm 0.042$	$0.966 \pm 0.027$
0.0063	0.0021	0.0771	0.3007	db2	$0.947 \pm 0.092$	$0.954 \pm 0.069$
0.0081	0.0933	0.0469	0.2520	haar	$0.952 \pm 0.056$	$0.958 \pm 0.050$
0.0053	0.0575	0.0649	0.1694	bior1.3	$0.962 \pm 0.050$	$0.965 \pm 0.046$

After multiclass grading of OSCC from histopathological images, the third step is semantic segmentation of the epithelial and stromal tissue. In contrast to the aforementioned preprocessing approach for multiclass classification, the data preprocessing in the case of semantic segmentation utilizes only a low-frequency subband. Therefore, after one-level decomposition utilizing the SWT with Haar, sym2, db2, and bior1.3 wavelet functions, only low-frequency coefficients with applied Greys colormap were used as input for the AI model. SWT decomposition of an image at level one using Haar wavelet is shown in Figure 9.



**Figure 9.** SWT decomposition at level 1 using Haar wavelet function. LL subband is used as an input image for semantic segmentation.

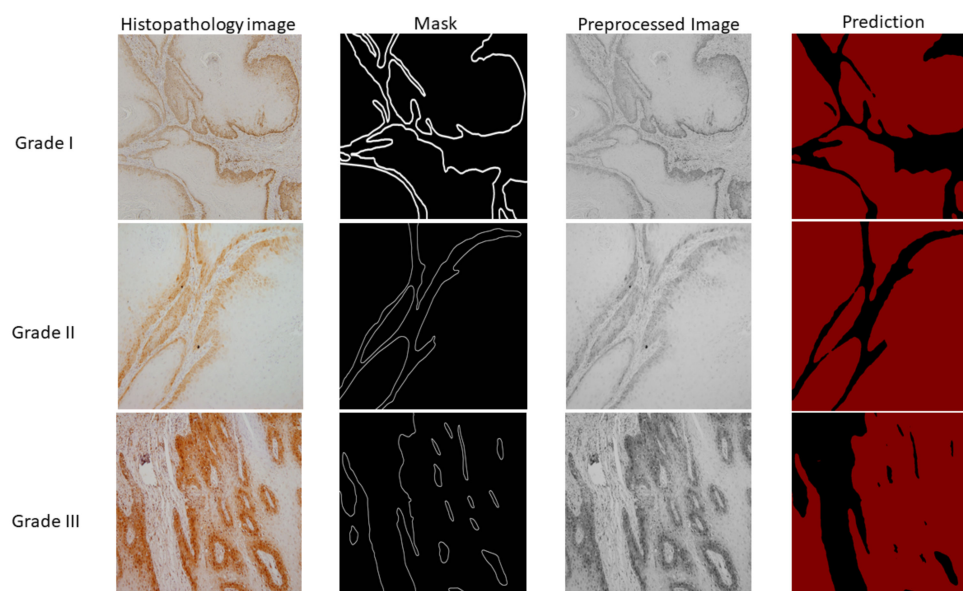
According to the results presented in Table 5, Xception\_65 was used as DeepLabv3+ backbone in order to perform semantic segmentation. The model was pre-trained on the Cityscapes dataset before the training on the oral carcinoma dataset was performed. As model input, original images and SWT approximations were used along with corresponding ground truth masks. In the training process, ASPP in a configuration with atrous rates

of 12, 24, 36 was used while the output stride was set to 8. Additionally, decoder output stride was set to 4. Performances of the models achieved on 5-fold cross-validation are shown in Table 7.

**Table 7.** Performance of DeepLabv3+ with Xception\_65 as backbone trained with data preprocessed with different wavelet functions.

		mIOU $\pm \sigma$	F1 $\pm \sigma$	Accuracy $\pm \sigma$	Precision $\pm \sigma$	Sensitivity $\pm \sigma$	Specificity $\pm \sigma$
DeepLabv3+ & Xception_65	Original	0.864 $\pm$ 0.020	0.933 $\pm$ 0.058	0.934 $\pm$ 0.012	0.933 $\pm$ 0.019	0.967 $\pm$ 0.013	0.873 $\pm$ 0.017
	sym2	0.874 $\pm$ 0.037	0.953 $\pm$ 0.016	0.939 $\pm$ 0.019	0.950 $\pm$ 0.025	0.956 $\pm$ 0.012	0.908 $\pm$ 0.040
	db2	0.876 $\pm$ 0.032	0.953 $\pm$ 0.016	0.940 $\pm$ 0.017	0.952 $\pm$ 0.019	0.955 $\pm$ 0.014	0.911 $\pm$ 0.031
	Haar	0.879 $\pm$ 0.027	0.955 $\pm$ 0.014	0.941 $\pm$ 0.015	0.951 $\pm$ 0.018	0.958 $\pm$ 0.016	0.910 $\pm$ 0.026
	bior1.3	0.874 $\pm$ 0.030	0.953 $\pm$ 0.015	0.939 $\pm$ 0.016	0.948 $\pm$ 0.020	0.958 $\pm$ 0.021	0.904 $\pm$ 0.027

By using data preprocessed with SWT at level one using a Haar wavelet function and a fully-trained DeepLabv3+ model, the predictions were made for the three cases of new and unseen data, as visually represented in Figure 10. Each sample from visual representation corresponds to a specific grade of OSCC.



**Figure 10.** Visual representation of histopathology images, ground truth masks, preprocessed images, and semantic segmentation results. The first column represents samples of OSCC obtained by the clinician while the second column is corresponding ground truth mask. The third column represents samples after preprocessing which are afterwards used as input variables for semantic segmentation. Finally, the last column shows the prediction for three cases (Grade I, II, and III) where the black colour represents stromal tissue and the red colour represents epithelial tissue.

AI experiments were performed using Python on a GPU-based High Performance Computing (HPC) server. The server consists of two Intel Xeon Gold CPUs (24 C/48 T, at 2.4 GHz), 768 GB of ECC DDR4 RAM, and five Nvidia Quadro RTX 6000 GPUs, with 24 GB of RAM, 4608 CUDA and 576 Tensor cores.

#### 4. Discussion

The diagnosis of OSCC is based on a histopathological assessment of altered oral mucosa which despite high subjectivity, still is the most reliable method of diagnosing oral carcinoma. The pathologist evaluates and grades the tumor depending on resemblance to the normal oral squamous epithelium as well, moderately or poorly differentiated [57]. Histopathological characteristics of well differentiated carcinoma are tumor islands with a central substantial amount of keratinization, keratin pearls. Moderately and poorly differentiated squamous cell carcinoma show more prominent nuclear and cellular pleo-



morphism, and prominent mitotic figures that can be abnormal, respectively. Usually, it is easy to identify an oral carcinoma by the invasion of the epithelial cell through the basement membrane border into connective tissues [57]. However, histopathological classification of oral carcinoma can be challenging due to heterogeneous structure and textures, and the presence of any inflammatory tissue reaction. With the help of artificial intelligence-aided tools, the automatic classification of histopathological images can not only improve objective diagnostic results for the clinician but also provide detailed texture analysis in order to get an accurate diagnosis [58].

This research proposes an AI-based system for multiclass grading and semantic segmentation of OSCC. The proposed system is compared with similar approaches presented in the literature [31,59] to validate feasibility.

The training process, in the case of multiclass grading, was performed in two stages as described in results section. The purpose of the two-stage learning process is to achieve better performance and at the same time increase the robustness of trained models. Since two additional layers were added at the end of the original Xception architecture, their weights were randomly initialized; therefore, training with a learning rate value of 0.001 was performed. However, in the second stage of the training process, the learning rate was set to a lower value of 0.0001. This way, the weights of layers pre-trained on ImageNet were only adapted to the new problem of multiclass grading.

From the presented results, in the case of multiclass grading with no preprocessing, it can be concluded that high values of 0.929  $AUC_{macro}$  and 0.942  $AUC_{micro}$  are achieved with a combination of Xception architecture and RMSprop optimizer. Furthermore, ResNet50 in a combination with the Adam optimizer showed  $AUC_{macro}$  and  $AUC_{micro}$  values of 0.871 and 0.864, respectively, which is slightly lower than ResNet101 performance (0.882  $AUC_{macro}$  and 0.890  $AUC_{micro}$ ). However, ResNet101—Adam was worst-performing in terms of standard deviation with values of  $\pm 0.125$ , and  $\pm 0.112$ . Lowest values of standard deviation were obtained in the case of MobileNetv2 architecture in a combination with Adam optimizer.

Since important features for distinguishing the differences between OSCC gradings are mostly contained in high-frequency components of the image, the wavelet coefficient mapping function was proposed. With the help of SWT, the image was decomposed on the LL, LH, HL, and HH subbands which allowed coefficient weighting. Mean values of the high-frequency components were located around zero; therefore, constants of coefficient mapping function were determined in a way that features with high- and low-coefficient values were enhanced. SWT reconstruction process with LL, LH', HL', and HH' subbands, resulted in images with enhanced high-frequency features.

If the performances, in the case of multiclass grading with preprocessing, are compared, it can be seen that all of the presented configurations achieved  $AUC_{macro}$  value of 0.947 and  $AUC_{micro}$  value of 0.954 or higher. Moreover, when all results are summed up, it can be noticed that the highest values of performance measure are achieved using the proposed methodology with coefficient mapping function constants a, b, c, and d with values of 0.091, 0.0301, 0.0086, and 0.3444, respectively, and db2 as wavelet function. Performance of the proposed model in terms of  $AUC_{macro}$  and  $AUC_{micro}$  values is  $0.963 \pm 0.042$  and  $0.966 \pm 0.027$ , respectively; therefore, it can be concluded that not only performance measure was increased, but also the values of standard deviation were decreased. A decrease in standard deviation value resulted in increased robustness of the model.

The foremost task for OSCC diagnosis and treatment plan is accurate histopathology examination of samples to detect morphological changes of the tumor cells. Histopathological grading is focused on the tumor cells features like nuclear pleomorphism, mitotic activity, depth of invasion, tumor thickness and degree of differentiation [60]. However, recent research showed the importance of the microenvironment in tumor progression and poor prognosis. The stroma-rich tumors showed association with unfavourable prognosis compared to stroma poor tumors, due to this the tumor-stroma ratio (TSR) could be useful prognostic outcomes factor [61]. Heikkinen et al. suggested that the degree of

stromal tumor-infiltrating lymphocytes (TILs) can be used like prognostic features, while other studies based on immunohistochemistry emphasize the prognostic value of different subtypes of immune cells and lymphocytes [62–64].

The conventional practice, histopathological assessment on light microscopy is time-consuming, tedious, and relies on the experience of pathologists. The characteristic of the microscope (resolution, light source, and the lens) and preparation of the tissue samples may hamper diagnosis and affect manual judgment. The histological parameters can be investigated under different magnification (4×, 10×, 20×, and 40×). Literature reveals that most of the segmentation research has been performed on 40× magnified images [65–68]. The 40× magnification allows the pathologist to visual features of the nucleus of cells and cell structure in the tissue as opposed to 10× magnification but provides more incomprehensive picture of the tumors.

Recently, many computer-aided tools for medical image analysis show significant progress towards better healthcare services. The second stage of the proposed AI-based system is semantic segmentation which is a mandatory step towards tumor microenvironment analysis. The semantic segmentation of epithelial and stromal tissue is performed on 10× magnification histopathology images.

In contrast to the preprocessing in the case of multiclass grading, preprocessing for semantic segmentation implies using only low-frequency coefficients, i.e., LL subband, while the high-frequency subbands are discarded. Therefore, by using extracted low-frequency features, DeepLabv3+ in a combination with Xception\_65 as backbone achieved satisfactory results in the case of multiclass semantic segmentation of tumor tissue. From the obtained results it can be noticed that the segmentation model in a combination with low-frequency features outperformed the model which used original, non-preprocessed data as input. In terms of performance measures, the highest mIOU (0.879), F1 (0.955), and Accuracy (0.941) values are achieved using low-frequency features obtained at SWT decomposition level-one using Haar wavelet. However, in terms of Precision and Specificity performance measures, data preprocessed with db2 wavelet function outperforms others with values of 0.952 (Precision), and 0.911 (Specificity). When Sensitivity measure is observed, non-preprocessed data achieved the highest value of 0.967.

Presented results indicate that segmenting of epithelial and stromal tissue has a great potential in the quantification of qualitative clinic-pathological features in order to predict tumor aggressiveness and metastasis. Segmented areas will be used in future work for analysing the tumor microenvironment, where the stroma is essential for the maintenance of epithelial tissue. Stromal and epithelial regions of the OSCC have a different impact on the disease progression which is important for the histopathology image analysis.

AI computer-aided systems for analysing the tumor microenvironment might contribute to modified treatment planning, improving prognosis and survival rates, and maintaining a high-quality life of patients.

## 5. Conclusions

Histology grading is a way to classify cancer cells based on tissue abnormality. It relies on the subjective component of the clinician and as such may adversely affect appropriate treatment methods and outcomes of the patient. This research highlights the huge potential of the application of image processing techniques with the aid of Artificial Intelligence algorithms in order to achieve an effective prognosis of OSCC and increase the chance of survival among people.

In the first stage of the research, the authors demonstrate the integration of deep convolutional neural networks with stationary wavelet transform along with wavelet coefficient mapping function for multiclass grading of OSCC. From obtained results, it can be concluded that integration of Xception and SWT resulted in the highest classification values of 0.963  $AUC_{macro}$  and 0.966  $AUC_{micro}$  with the lowest standard deviation of  $\pm 0.042$  and  $\pm 0.027$ , respectively.

In the second stage, semantic segmentation was performed. Xception\_65 as DeepLabv3+ backbone, integrated with low-frequency subband resulted in  $0.879 \pm 0.027$  mIOU,  $0.955 \pm 0.014$  F1 score and  $0.941 \pm 0.015$  accuracy. Segmentation of the tumor on the epithelial and stromal regions is the initial step in the study of the tumor microenvironment and its impact on the disease progression.

Based on the results of the proposed methodology, a two-stage AI-based system has been proven successful in terms of multiclass grading as well as segmenting of epithelial and stromal tissue and has a great potential in the clinical application in prediction of tumor invasion and outcomes of a patient with OSCC. Since the limitation of this research was data availability, future work should use a dataset with more histopathology images, to achieve a more robust system. The presented approach will be the first step in analysing the tumor microenvironment, i.e., tumor-stroma ratio and segmentation of the microenvironment cells. The main idea is to develop an advanced automatic prognostic system that could analyse parameters such as cell shape and size, pathological mitoses, tumor-stroma ratio and distinction between early- and advanced-stage OSCCs. Moreover, try to create predictive algorithms that would improve prognostic indicators in clinical practice and thus improve therapeutic treatment for the patient.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/cancers13081784/s1>, Table S1: Legend of abbreviations and acronyms used in the research.

**Author Contributions:** Conceptualization, J.M., D.Š., A.Z., T.Ć., A.D. and Z.C.; methodology, J.M., D.Š. and A.Z.; software, J.M. and D.Š.; validation, A.Z. and Z.C.; formal analysis, T.Ć., A.D. and Z.C.; investigation, J.M., D.Š. and A.Z.; resources, T.Ć. and Z.C.; data curation, A.Z. and A.D.; writing—original draft preparation, J.M., D.Š., A.Z.; writing—review and editing, T.Ć., A.D. and Z.C.; visualization, J.M.; supervision, T.Ć. and Z.C.; project administration, T.Ć. and Z.C.; funding acquisition, T.Ć. and Z.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The research has been approved by Clinical Hospital Center Rijeka, Ethics Board (Krešimirova 42, 51000 Rijeka); under the number 2170-29-02/1-19-2, on 24 September 2019.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author if data sharing is approved by ethics committee. The data are not publicly available due to data protection laws and conditions stated by the ethics committee.

**Acknowledgments:** This research has been (partly) supported by the CEEPUS network CIII-HR-0108, European Regional Development Fund under the grant KK.01.1.1.01.0009 (DATACROSS), project CEKOM under the grant KK.01.2.2.03.0004, CEI project “COVIDAi” (305.6019-20) and University of Rijeka scientific grant uniri-tehnic-18-275-1447.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Torre, L.A.; Siegel, R.L.; Ward, E.M.; Jemal, A. Global Cancer Incidence and Mortality Rates and Trends—An Update. *Cancer Epidemiol. Biomark. Prev.* **2015**, *25*, 16–27. [[CrossRef](#)] [[PubMed](#)]
2. Marur, S.; Forastiere, A.A. Head and Neck Cancer: Changing Epidemiology, Diagnosis, and Treatment. *Mayo Clin. Proc.* **2008**, *83*, 489–501. [[CrossRef](#)] [[PubMed](#)]
3. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Cancer J. Clin.* **2021**. [[CrossRef](#)] [[PubMed](#)]
4. Bagan, J.; Sarrion, G.; Jimenez, Y. Oral cancer: Clinical features. *Oral Oncol.* **2010**, *46*, 414–417. [[CrossRef](#)]
5. Ganesh, D.; Sreenivasan, P.; Öhman, J.; Wallström, M.; Braz-Silva, P.H.; Giglio, D.; Kjeller, G.; Hasséus, B. Potentially Malignant Oral Disorders and Cancer Transformation. *Anticancer Res.* **2018**, *38*, 3223–3229. [[CrossRef](#)]
6. Ettinger, K.S.; Ganry, L.; Fernandes, R.P. Oral Cavity Cancer. *Oral Maxillofac. Surg. Clin. N. Am.* **2019**, *31*, 13–29. [[CrossRef](#)]
7. Milas, Z.L.; Shellenberger, T.D. The Head and Neck Cancer Patient: Neoplasm Management. *Oral Maxillofac. Surg. Clin. N. Am.* **2019**, *31*. [[CrossRef](#)]

8. Warnakulasuriya, S.; Reibel, J.; Bouquot, J.; Dabelsteen, E. Oral epithelial dysplasia classification systems: Predictive value, utility, weaknesses and scope for improvement. *J. Oral Pathol. Med.* **2008**, *37*, 127–133. [[CrossRef](#)]
9. Mehlum, C.S.; Larsen, S.R.; Kiss, K.; Groentved, A.M.; Kjaergaard, T.; Möller, S.; Godballe, C. Laryngeal precursor lesions: Interrater and intrarater reliability of histopathological assessment. *Laryngoscope* **2018**, *128*, 2375–2379. [[CrossRef](#)]
10. Chen, H.; Sung, J.J.Y. Potentials of AI in medical image analysis in Gastroenterology and Hepatology. *J. Gastroenterol. Hepatol.* **2021**, *36*, 31–38. [[CrossRef](#)]
11. Stolte, S.; Fang, R. A survey on medical image analysis in diabetic retinopathy. *Med Image Anal.* **2020**, *64*, 101742. [[CrossRef](#)]
12. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciampi, F.; Ghafoorian, M.; van der Laak, J.A.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)]
13. Singh, A.; Sengupta, S.; Lakshminarayanan, V. Explainable Deep Learning Models in Medical Image Analysis. *J. Imaging* **2020**, *6*, 52. [[CrossRef](#)]
14. Haefner, N.; Wincent, J.; Parida, V.; Gassmann, O. Artificial intelligence and innovation management: A review, framework, and research agenda. *Technol. Forecast. Soc. Chang.* **2021**, *162*, 120392. [[CrossRef](#)]
15. Kaba, K.; Sarigül, M.; Avci, M.; Kandirmaz, H.M. Estimation of daily global solar radiation using deep learning model. *Energy* **2018**, *162*, 126–135. [[CrossRef](#)]
16. Lorencin, I.; Anđelić, N.; Mrzljak, V.; Car, Z. Genetic Algorithm Approach to Design of Multi-Layer Perceptron for Combined Cycle Power Plant Electrical Power Output Estimation. *Energies* **2019**, *12*, 4352. [[CrossRef](#)]
17. Gurcan, M.N.; Boucheron, L.E.; Can, A.; Madabhushi, A.; Rajpoot, N.M.; Yener, B. Histopathological Image Analysis: A Review. *IEEE Rev. Biomed. Eng.* **2009**, *2*, 147–171. [[CrossRef](#)]
18. Sharma, S.; Mehra, R. Conventional Machine Learning and Deep Learning Approach for Multi-Classification of Breast Cancer Histopathology Images—A Comparative Insight. *J. Digit. Imaging* **2020**, *33*, 632–654. [[CrossRef](#)]
19. Wu, Z.; Wang, L.; Li, C.; Cai, Y.; Liang, Y.; Mo, X.; Lu, Q.; Dong, L.; Liu, Y. DeepLRHE: A Deep Convolutional Neural Network Framework to Evaluate the Risk of Lung Cancer Recurrence and Metastasis from Histopathology Images. *Front. Genet.* **2020**, *11*, 768. [[CrossRef](#)]
20. Tabibu, S.; Vinod, P.K.; Jawahar, C.V. Pan-Renal Cell Carcinoma classification and survival prediction from histopathology images using deep learning. *Sci. Rep.* **2019**, *9*, 1–9. [[CrossRef](#)]
21. Arijji, Y.; Fukuda, M.; Kise, Y.; Nozawa, M.; Yanashita, Y.; Fujita, H.; Katsumata, A.; Arijji, E. Contrast-enhanced computed tomography image assessment of cervical lymph node metastasis in patients with oral cancer by using a deep learning system of artificial intelligence. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **2019**, *127*, 458–463. [[CrossRef](#)]
22. Halicek, M.; Dormer, J.D.; Little, J.V.; Chen, A.Y.; Myers, L.; Sumer, B.D.; Fei, B. Hyperspectral Imaging of Head and Neck Squamous Cell Carcinoma for Cancer Margin Detection in Surgical Specimens from 102 Patients Using Deep Learning. *Cancers* **2019**, *11*, 1367. [[CrossRef](#)]
23. Horie, Y.; Yoshio, T.; Aoyama, K.; Yoshimizu, S.; Horiuchi, Y.; Ishiyama, A.; Hirasawa, T.; Tsuchida, T.; Ozawa, T.; Ishihara, S.; et al. Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. *Gastrointest. Endosc.* **2019**, *89*, 25–32. [[CrossRef](#)]
24. Tamashiro, A.; Yoshio, T.; Ishiyama, A.; Tsuchida, T.; Hijikata, K.; Yoshimizu, S.; Horiuchi, Y.; Hirasawa, T.; Seto, A.; Sasaki, T.; et al. Artificial intelligence-based detection of pharyngeal cancer using convolutional neural networks. *Dig. Endosc.* **2020**, *32*, 1057–1065. [[CrossRef](#)]
25. Jeyaraj, P.R.; Nadar, E.R.S. Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm. *J. Cancer Res. Clin. Oncol.* **2019**, *145*, 829–837. [[CrossRef](#)]
26. Bhandari, B.; Alsadoon, A.; Prasad, P.W.C.; Abdullah, S.; Haddad, S. Deep learning neural network for texture feature extraction in oral cancer: Enhanced loss function. *Multimed. Tools Appl.* **2020**, *79*, 1–24. [[CrossRef](#)]
27. Xu, S.; Liu, Y.; Hu, W.; Zhang, C.; Liu, C.; Zong, Y.; Chen, S.; Lu, Y.; Yang, L.; Ng, E.Y.K.; et al. An Early Diagnosis of Oral Cancer based on Three-Dimensional Convolutional Neural Networks. *IEEE Access* **2019**, *7*, 158603–158611. [[CrossRef](#)]
28. Welikala, R.A.; Remagnino, P.; Lim, J.H.; Chan, C.S.; Rajendran, S.; Kallarakkal, T.G.; Zain, R.B.; Jayasinghe, R.D.; Rimal, J.; Kerr, A.R.; et al. Automated Detection and Classification of Oral Lesions Using Deep Learning for Early Detection of Oral Cancer. *IEEE Access* **2020**, *8*, 132677–132693. [[CrossRef](#)]
29. Chan, C.-H.; Huang, T.-T.; Chen, C.-Y.; Lee, C.-C.; Chan, M.-Y.; Chung, P.-C. Texture-Map-Based Branch-Collaborative Network for Oral Cancer Detection. *IEEE Trans. Biomed. Circuits Syst.* **2019**, *13*, 766–780. [[CrossRef](#)] [[PubMed](#)]
30. Fraz, M.M.; Khurram, S.A.; Graham, S.; Shaban, M.; Hassan, M.; Loya, A.; Rajpoot, N.M. FABnet: Feature attention-based network for simultaneous segmentation of microvessels and nerves in routine histology images of oral cancer. *Neural Comput. Appl.* **2020**, *32*, 9915–9928. [[CrossRef](#)]
31. Das, N.; Hussain, E.; Mahanta, L.B. Automated classification of cells into multiple classes in epithelial tissue of oral squamous cell carcinoma using transfer learning and convolutional neural network. *Neural Netw.* **2020**, *128*, 47–60. [[CrossRef](#)]
32. El-Naggar, A.K.; Chan, J.K.; Takata, T.; Grandis, J.R.; Slootweg, P.J. WHO classification of head and neck tumours. *Int. Agency Res. Cancer* **2017**. [[CrossRef](#)]
33. Amin, M.B.; Edge, S.B.; Greene, F.L.; Byrd, D.R.; Brookland, R.K.; Washington, M.K.; Gershenwald, J.E.; Compton, C.C.; Hess, K.R.; Sullivan, D.C.; et al. *AJCC Cancer Staging Manual*, 8th ed.; Springer: Cham, Switzerland, 2017. [[CrossRef](#)]

34. Jakovac, H.; Stašić, N.; Krašević, M.; Jonjić, N.; Radošević-Stašić, B. Expression profiles of metallothionein-I/II and megalin/LRP-2 in uterine cervical squamous lesions. *Virchows Archiv* **2021**, *478*, 735–746. [[CrossRef](#)]
35. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [[CrossRef](#)]
36. Han, J.; Kamber, M.; Pei, J. Classification. In *Data Mining*; Elsevier: Amsterdam, The Netherlands, 2012; pp. 327–391.
37. Addison, P.S. *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance*; CRC Press: Boca Raton, FL, USA, 2017. [[CrossRef](#)]
38. Štifanić, D.; Musulin, J.; Miočević, A.; Šegota, S.B.; Šubić, R.; Car, Z. Impact of COVID-19 on Forecasting Stock Prices: An Integration of Stationary Wavelet Transform and Bidirectional Long Short-Term Memory. *Complexity* **2020**, *2020*, 1–12. [[CrossRef](#)]
39. Zhang, D. Wavelet transform. In *Fundamentals of Image Data Mining*; Springer: Cham, Switzerland, 2019; pp. 35–44. [[CrossRef](#)]
40. Qayyum, H.; Majid, M.; Anwar, S.M.; Khan, B. Facial Expression Recognition Using Stationary Wavelet Transform Features. *Math. Probl. Eng.* **2017**, *2017*, 1–9. [[CrossRef](#)]
41. Janani, S.; Marisuganya, R.; Nivedha, R. MRI image segmentation using Stationary Wavelet Transform and FCM algorithm. *Int. J. Comput. Appl.* **2013**. [[CrossRef](#)]
42. Feurer, M.; Hutter, F. Hyperparameter optimization. In *Automated Machine Learning*; Springer: Cham, Switzerland, 2019; pp. 3–33. [[CrossRef](#)]
43. Swersky, K.; Snoek, J.; Adams, R.P. Multi-task Bayesian optimization. In *NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems 2004–2012*; Curran Associates Inc.: Red Hook, NY, USA, 2013.
44. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
45. 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, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
46. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
47. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [[CrossRef](#)]
48. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
49. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818. [[CrossRef](#)]
50. Choudhury, A.R.; Vanguri, R.; Jambawalikar, S.R.; Kumar, P. Segmentation of brain tumors using DeepLabv3+. In *International MICCAI Brainlesion Workshop*; Springer: Cham, Switzerland, 2018; pp. 154–167. [[CrossRef](#)]
51. Tharwat, A. Classification assessment methods. *Appl. Comput. Inform.* **2020**, *17*, 168–192. [[CrossRef](#)]
52. Leonard, L. Web-Based Behavioral Modeling for Continuous User Authentication (CUA). In *Advances in Computers*; Elsevier: Amsterdam, The Netherlands, 2017; Volume 105, pp. 1–44.
53. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 658–666.
54. Chicco, D.; Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genom.* **2020**, *21*, 6. [[CrossRef](#)]
55. Gunawardana, A.; Shani, G. A survey of accuracy evaluation metrics of recommendation tasks. *J. Mach. Learn. Res.* **2009**, *10*. [[CrossRef](#)]
56. Mumtaz, W.; Ali, S.S.A.; Yasin, M.A.M.; Malik, A.S. A machine learning framework involving EEG-based functional connectivity to diagnose major depressive disorder (MDD). *Med Biol. Eng. Comput.* **2018**, *56*, 233–246. [[CrossRef](#)]
57. Speight, P.M.; Farthing, P.M. The pathology of oral cancer. *Br. Dent. J.* **2018**, *225*, 841–847. [[CrossRef](#)]
58. Li, L.; Pan, X.; Yang, H.; Liu, Z.; He, Y.; Li, Z.; Fan, Y.; Cao, Z.; Zhang, L. Multi-task deep learning for fine-grained classification and grading in breast cancer histopathological images. *Multimed. Tools Appl.* **2018**, *79*, 14509–14528. [[CrossRef](#)]
59. Al-Milaji, Z.; Ersoy, I.; Hafiane, A.; Palaniappan, K.; Bunyak, F. Integrating segmentation with deep learning for enhanced classification of epithelial and stromal tissues in H&E images. *Pattern Recognit. Lett.* **2019**, *119*, 214–221. [[CrossRef](#)]
60. Almangush, A.; Mäkitie, A.A.; Triantafyllou, A.; de Bree, R.; Strojan, P.; Rinaldo, A.; Hernandez-Prera, J.C.; Suárez, C.; Kowalski, L.P.; Ferlito, A.; et al. Staging and grading of oral squamous cell carcinoma: An update. *Oral Oncol.* **2020**, *107*, 104799. [[CrossRef](#)]
61. Mascitti, M.; Zhurakivska, K.; Togni, L.; Caponio, V.C.A.; Almangush, A.; Balercia, P.; Balercia, A.; Rubini, C.; Muzio, L.L.; Santarelli, A.; et al. Addition of the tumour–stroma ratio to the 8th edition American Joint Committee on Cancer staging system improves survival prediction for patients with oral tongue squamous cell carcinoma. *Histopathology* **2020**, *77*, 810–822. [[CrossRef](#)]
62. Heikkinen, I.; Bello, I.O.; Wahab, A.; Hagström, J.; Haglund, C.; Coletta, R.D.; Nieminen, P.; Mäkitie, A.A.; Salo, T.; Leivo, I.; et al. Assessment of Tumor-infiltrating Lymphocytes Predicts the Behavior of Early-stage Oral Tongue Cancer. *Am. J. Surg. Pathol.* **2019**, *43*, 1392–1396. [[CrossRef](#)]

63. Agarwal, R.; Chaudhary, M.; Bohra, S.; Bajaj, S. Evaluation of natural killer cell (CD57) as a prognostic marker in oral squamous cell carcinoma: An immunohistochemistry study. *J. Oral Maxillofac. Pathol.* **2016**, *20*, 173–177. [[CrossRef](#)]
64. Fang, J.; Li, X.; Ma, D.; Liu, X.; Chen, Y.; Wang, Y.; Lui, V.W.Y.; Xia, J.; Cheng, B.; Wang, Z. Prognostic significance of tumor infiltrating immune cells in oral squamous cell carcinoma. *BMC Cancer* **2017**, *17*, 375. [[CrossRef](#)] [[PubMed](#)]
65. Jonnalagedda, P.; Schmolze, D.; Bhanu, B. mvpnets: Multi-viewing path deep learning neural networks for magnification invariant diagnosis in breast cancer. In Proceedings of the 2018 IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE), Taichung, Taiwan, 29–31 October 2018; pp. 189–194. [[CrossRef](#)]
66. Silva, A.B.; Martins, A.S.; Neves, L.A.; Faria, P.R.; Tosta, T.A.; do Nascimento, M.Z. Automated nuclei segmentation in dysplastic histopathological oral tissues using deep neural networks. In *Iberoamerican Congress on Pattern Recognition*; Springer: Cham, Switzerland, 2019; pp. 365–374. [[CrossRef](#)]
67. Fauzi, M.F.A.; Chen, W.; Knight, D.; Hampel, H.; Frankel, W.L.; Gurcan, M.N. Tumor Budding Detection System in Whole Slide Pathology Images. *J. Med. Syst.* **2019**, *44*, 38. [[CrossRef](#)] [[PubMed](#)]
68. Rashmi, R.; Prasad, K.; Udupa CB, K.; Shwetha, V. A Comparative Evaluation of Texture Features for Semantic Segmentation of Breast Histopathological Images. *IEEE Access* **2020**, *8*, 64331–64346. [[CrossRef](#)]