UNIVERSITY OF RIJEKA
FACULTY OF ENGINEERING

Jelena Štifanić

AN AUTOMATED COMPUTER-AIDED SYSTEM FOR ORAL SQUAMOUS CELL CARCINOMA ANALYSIS

DOCTORAL THESIS

Rijeka, 2025.

UNIVERSITY OF RIJEKA
FACULTY OF ENGINEERING

Jelena Štifanić

AN AUTOMATED COMPUTER-AIDED SYSTEM FOR ORAL SQUAMOUS CELL CARCINOMA ANALYSIS

DOCTORAL THESIS

Supervisor: Prof. D. Sc. Zlatan Car
Co-supervisor: Doc. D. Sc. Nikola Anđelić

Rijeka, 2025.

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET

Jelena Štifanić

AUTOMATIZIRANI RAČUNALNO POTPOMOGNUTI SUSTAV ZA ANALIZU KARCINOMA PLOČASTIH STANICA USNE ŠUPLJINE

DOKTORSKI RAD

Mentor: prof. dr. sc. Zlatan Car
Komentor: doc. dr. sc. Nikola Anđelić

Rijeka, 2025.

Thesis Supervisor: Prof. D. Sc. Zlatan Car, Catholic University of Croatia, Faculty of Engineering

Thesis Co-supervisor: Doc. D. Sc. Nikola Anđelić, University of Rijeka, Faculty of Engineering

This doctoral thesis was discussed on _____ at the University of Rijeka, Croatia, Faculty of Engineering in front of the following Evaluation Committee:

1.
2.
3.

Words of appreciation

First of all, dedicating several years of your life to this kind of effort requires you to occasionally rely on the kindness and endurance of others, particularly during difficult periods that can arise when you least expect or need them. Therefore, I would like to express my deepest gratitude to all whose support and presence made completing this doctoral thesis possible. I am particularly grateful to prof. dr. sc. Zlatan Car for the opportunity to pursue doctoral studies.

.

To my daughter, my greatest motivation, thank you for teaching me that I can do hard things.
You encourage me to be creative and passionate about my dreams.

To my husband, thank you for your love, support, and patience.
I could not have wished for a better companion by my side in reaching this milestone.

Because of you two, this journey has even greater meaning.

SAŽETAK

Karcinom pločastih stanica usne šupljine jedan je od najčešćih karcinoma glave i vrata. Standardni postupak za dijagnozu karcinoma pločastih stanica temelji se na histopatološkom pregledu, međutim, glavni izazov kod ove vrste pregleda je heterogenost tumora gdje subjektivna komponenta pregleda može izravno utjecati na način liječenja specifičnog za pacijenta. Iz tog razloga, u ovom doktorskom radu koristili su se algoritmi umjetne inteligencije za razvoj naprednog dijagnostičkog sustava temeljenog na histopatološkim slikama kao pomoć u analizi tumora. Takav sustav objedinio je višeklasnu klasifikaciju gradusa, Grad-CAM vizualizaciju, semantičku segmentaciju tumora na epitelne i stromalne regije te automatsku kvantifikaciju omjera tumora i strome zajedno s analizom preživljenja pacijenta u svrhu smanjenja varijabilnosti između promatrača te ubrzanja vremena potrebnog za postavljanje dijagnoze.

Ključne riječi: Umjetna inteligencija, karcinom pločastih stanica usne šupljine, višeklasna klasifikacija, semantička segmentacija, Grad-CAM vizualizacija, omjer tumor-stroma

ABSTRACT

The most common malignant epithelial tumor that affects the oral cavity is oral squamous cell carcinoma. The histopathological examination of biopsy slides is currently the most reliable method for diagnosing oral cancer. However, tumor heterogeneity is the primary issue with this type of procedure, since a subjective aspect of the examination may have a direct impact on tumor diagnosis. For this reason, in this doctoral thesis, Artificial Intelligence algorithms are used in order to develop an advanced diagnostic system based on histopathological images as computational aid in tumor diagnosis. Such a system is composed of multiclass classification of grades, Grad-CAM visualization, semantic segmentation of tumor into epithelial vs. stromal regions, and automatic quantification of tumor-stroma ratio along with overall survival analysis to reduce inter-and intra-observer variability and speed up the diagnosis process.

Keywords: Artificial Intelligence, Oral Squamous Cell Carcinoma, Multiclass classification, Semantic segmentation, Grad-Cam, Tumor-stroma Ratio

PROŠIRENI SAŽETAK

Karcinom usne šupljine je među deset najčešćih karcinoma u Europi i SAD-u gdje više od 90% spada u skupinu karcinoma pločastih stanica usne šupljine. Standardne metode za otkrivanje karcinoma usne šupljine su inspekcija i palpacija uz detaljnu anamnezu dok se histološki potvrđuje biopsijom tkiva. Biopsijom se definiraju karakteristike tumora te se na temelju toga određuju terapija, prognoza ishoda bolesti i preživljenje pacijenta. Ključni izazov histopatološkog pregleda proizlazi iz subjektivne komponente kliničke dijagnostike, odnosno u varijabilnosti opažanja među različitim stručnjacima. Stoga, cilj ovog istraživanja je razviti dijagnostički sustav uz pomoć algoritama umjetne inteligencije za analizu kancerogenih lezija koji može poboljšati objektivnosti i ponovljivost histopatološkog pregleda, odnosno smanjiti varijabilnost opažanja između stručnjaka. Nadalje, takav sustav pridonio bi minimalnom invazivnom liječenju/kirurškoj terapiji, poboljšanju ishoda te stopi preživljavanja i održavanja visoke kvalitete života pacijenata. Nadalje, mogao bi pomoći stručnjaku odnosno patologu u smanjenju opterećenja ručnih pregleda te ubrzati vrijeme potrebno za dijagnozu.

Prvi korak ovog istraživanja je uspostava jedinstvenog skupa podataka. Prikupljeni histopatološki uzorci su se klasificirali u skladu sa Svjetskom zdravstvenom organizacijom, koje su potom dodatno pregledala i recenzirala dva neovisna patologa. Prema ranije spomenutoj klasifikaciji uzorci su se podijelili u tri klase: Gradus I (dobro diferencirani tumor), Gradus II (umjereno diferencirani tumor) i Gradus III (slabo diferencirani tumor). Dodatno, generirale su se pripadajuće maske na kojima je jasno određena granica između epitelnog i stromalnog tkiva. Prikupljeni slikovni skup podataka koristio se kao ulaz u algoritme umjetne inteligencije kako bi se razvio personalizirani dijagnostički sustav za analizu karcinoma pločastih stanica usne šupljine.

Iz perspektive umjetne inteligencije veliki raspon boja na slikama može uzrokovati da algoritmi teže prepoznaju ključne značajke jer nisu svi dijelovi slike jednako vidljivi ili značajni za analizu, stoga su se koristile relevantne tehnike za predobradu slika u svrhu izlučivanja značajki koje sadržavaju informacije od interesa. Nadalje, ispitalo se više modela temeljenih na umjetnoj inteligenciji za višeklasnu klasifikaciju gradusa OSCC-a. Nakon konačnog odabira modela i tehnike predobrade, ispitala će se mogućnost njihovog daljnjeg razvoja u svrhu poboljšanja performansi. Ovakvim pristupom cilj je poboljšati objektivnost i ponovljivost kako bi se smanjila varijabilnost opažanja među patolozima u klasificiranju gradusa pločastog karcinoma usne šupljine.

Modeli temeljeni na umjetnoj inteligenciji postali su izuzetno moćan alat za otkrivanje i klasifikaciju karcinoma, međutim, mnogi modeli dubokog učenja i dalje se smatraju 'crnim kutijama' što se tiče razumijevanja njihovih mehanizama za donošenje odluka, posebice u ključnim primjenama kao je dijagnoza karcinoma. Stoga se u drugom koraku ovog istraživanja primijenila metoda objašnjive umjetne inteligencije zvana Grad-CAM kako bi se vizualizirale odluke modela dubokog učenja. Takvim pristupom poboljšalo se povjerenje i transparentnost u dijagnostičkom procesu temeljenom na umjetnoj inteligenciji.

U idućem koraku izvršila se semantička segmentacija, gdje je svaki piksel slike označen odgovarajućom klasom onoga što predstavlja. Na taj se način točno odredilo područje interesa, tj. lezije tumora na slici, zajedno s točnom granicom između epitela i strome. U istraživanju se ispitalo više modela temeljenih na umjetnoj inteligenciji za semantičku segmentaciju epitelnog i stormalnog tkiva u svrhu odabira modela s optimalnim performansama. Kao i u prethodnom koraku, ispitala se mogućnost daljnjeg unapređenja segmentacijskog modela u svrhu poboljšanja performansi. Ovakav pristup automatizirane segmentacije stromalne regije može pomoći patolozima u otkrivanju novih informativnih značajki.

Semantičkom segmentacijom epitelnog i stormalnog tkiva dolazi se do završnog koraka istraživanja, a to je kvantifikacija omjera tumora i strome (TSR). Omjer tumora i strome u pokazao se kao obećavajuća metoda za predviđanje ishoda bolesti i preživljenja pacijenta. Usprkos potencijalnoj prognozirajućoj vrijednosti određivanje TSR-a ponekad je izazovno, stoga se u ovom istraživanju razvio protokol za analizu omjera tumora i strome na histopatološkim uzorcima. Automatiziranom kvantifikacijom omjera tumora i strome doprinijelo bi se poboljšanju objektivnosti i ponovljivosti histopatološkog pregleda. Zaključno, omjer tumora i strome karcinoma pločastih stanica koristio se kao podatak za procjenu ukupnog preživljenja pacijenta.

Za postupak validacije prikupio se i označio novi skup histopatoloških podataka po istom principu kao i inicijalni skup podataka. Predikciju UI sustava temeljenu na novo prikupljenim podacima vrednovali su stručnjaci iz KBC Rijeka kako bi se utvrdile performanse sustava i potvrdio koncept.

Contents

1. Introduction

This chapter introduces the doctoral theses along with scientific motivation for research. Additionally, research goals and hypotheses are outlined. Furthermore, the significance and contribution of the research are explained. Finally, the structure of the thesis is provided.

1.1. Scientific Motivation

More than 90% of cases of oral cancer (OC) are squamous cell carcinoma, making it one of the top ten most prevalent cancers in both Europe and the United States [53, 91]. However, with advancements in diagnosis and treatment for OC patients, mortality and morbidity rates have not decreased over the past 50 years [7]. Oral squamous cell carcinoma (OSCC) frequently develops from pre-existing oral mucosal lesions that have a higher chance of malignant transformation into cancer. Early detection, diagnosis, and therapy at the precancerous stage improves the survival rates and morbidity related to OSCC treatment [31]. Surgical resection, with or without adjuvant radiation, is typically the main treatment for OSCC, and it has a substantial effect on the patient's quality of life [27]. Even with significant progress in comprehending the intricate processes of carcinogenesis, a trustworthy prognostic prediction tool is still lacking. When determining the prognosis, treatment strategy, and predicting outcomes for patients with OSCC tumor-node-metastasis (TNM) staging is frequently utilized. However, the limits of TNM staging in prognostic prediction are evident in its ability of assessing the individual characteristics of the patient, such as lifestyle choices and clinical features [58].

The current gold standard for detecting oral cancer is clinical examination, conventional oral examination (COE), and histological evaluation following biopsy. These approaches can identify cancer in the stage of established lesions with notable malignant changes [105].

However, the main drawback of employing histological examination for tumor classification and prognostic evaluation is inter- and intra-observer variability [55]. The most recent advances in artificial intelligence (AI)-based medical imaging contributed to reducing variability among observers as well as reducing repetitive tasks and enabling quick accurate diagnosis [11].

This research aims to create an advanced automated AI prognostic system that may directly influence patient-specific interventions by determining patient's outcome, while also increasing inter- and intra-observer variability.

1.2. Research Objectives and Hypotheses

The main research objective is to develop an automated system based on artificial intelligence algorithms for the analysis of oral squamous cell carcinoma. Such a system would enable the objective grading of the carcinoma, precise separation of epithelial and stromal tissues, which would be used for automatic quantification of the tumor and stroma, as well as patient survival analysis.

Based on the defined objective, the following hypotheses are proposed:

- Through advanced data preprocessing combined with an artificial intelligence-based model, it is possible to achieve high performance in the multiclass classification of oral squamous cell carcinoma grades.

- With a hybrid artificial intelligence-based model, it is possible to achieve high performance in the semantic segmentation of OSCC and to perform automatic quantification of the tumor-to-stroma ratio, along with patient survival analysis.

1.3. Research Contibution and Significance

This research makes a significant contribution to the field of medical image analysis and the diagnosis of oral cancer using artificial intelligence. A novel AI-based system has been proposed, and it consists of:

Stage 1: A novel preprocessing method based on the Stationary Wavelet Transform (SWT) is intended to:

- increase classification performance by enhancing high-frequency components and
- extract low-level features for more precise semantic segmentation.

Stage 2: Automated multiclass grading of oral squamous cell carcinoma (OSCC), which attempts to decrease the time needed for manual pathological inspections while increasing the objectivity of histological evaluations.

Stage 3: Providing interpretable explanations to establish confidence and guarantee transparency in AI-based diagnostic process utilizing explainable AI techniques.

Stage 4: Semantic segmentation of tumor into epithelial vs. stromal tissue regions in histopathological images, enabling the identification of features that are clinically informative and may assist in predicting tumor invasion and metastasis.

Stage 5: Establishing a procedure for the automated quantification of the tumor-stroma ratio and the analysis of patient survival would help to increase the objectivity and repeatability of histopathological analysis.

The scientific contributions of the research are:

- Development of data preprocessing methodology and implementation of a model for multiclass grading of oral squamous cell carcinoma.
- Development of a customized hybrid model for semantic segmentation of the tumor into epithelial vs. stromal regions.
- Creation of a protocol for automatic quantification of the tumor-to-stroma ratio, along with patient survival analysis.

1.4. Structure of the Thesis

The first step of the research is the establishment of a unique dataset. The collected histopathology image dataset will be used as input for artificial intelligence algorithms in order to develop a personalized diagnostic system for analyzing oral squamous cell carcinoma. In the next step, image preprocessing techniques will be applied to extract features containing information of interest. In the third step, multiple AI-based models will be evaluated for multiclass classification of OSCC grades. After selecting the final model and preprocessing technique, the possibility of further development will be explored in order to improve performance. In the fourth step, visualization tools such as Grad-CAM will be utilized to enhance transparency in the AI-based diagnostic process. In the next step, semantic segmentation will be performed. The research will evaluate several AI-based models for the semantic segmentation of tumor on epithelial vs. stromal tissue in order to select the model with optimal performance. As in the third step, the potential for further enhancement of the segmentation model will also be examined to improve its performance.

Semantic segmentation of tumor on epithelial vs. stromal tissue leads to the final step of the research, which is the quantification of the tumor-stroma ratio (TSR). Despite its potential prognostic value, determining TSR can be challenging. Therefore, this research will develop a protocol for the analysis of the tumor-stroma ratio in histopathological samples. Finally, the TSR will be used as a parameter for evaluating overall patient survival analysis. The framework of the proposed AI-based system is shown in Figure 1.1.

For the validation process, a new set of histopathological images will be collected and annotated using the same principles as those applied to the initial dataset. The AI system's predictions based on this newly collected data will be evaluated by experts from the Clinical Hospital Center Rijeka (KBC Rijeka) to assess the performance of the proposed AI-based system and validate the concept.

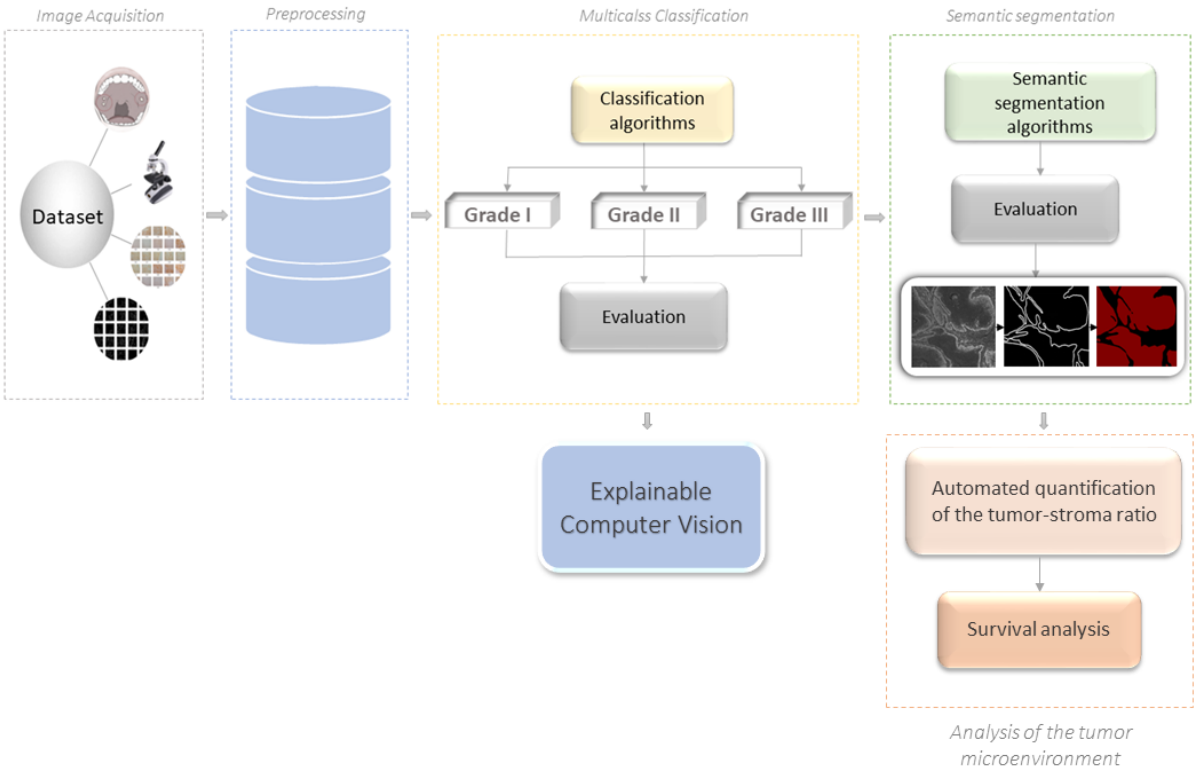

Figure 1.1. Framework of the proposed AI-based system: It incorporates image acquisition, preprocessing, tumor grading, explainable AI, semantic segmentation, quantification of tumor-stroma ratio and overall survival analysis.

2. Literature Review

This chapter aims to provide an overview of the existing AI – solutions in medical image analysis. It briefly describes the various models, techniques, and methodologies used in different solutions in similar areas of study. First, in order to ensure that proper studies are chosen, inclusion and exclusion criteria are established. Then, an overview of AI solutions for OSCC classification is presented. After that, an overview of AI solutions for segmentation of tumor on epithelial vs. stromal tissue is given. Furthermore, research related to the automatic quantification of TSR is presented. A literature overview regarding explainable computer vision for OSCC is provided at the end of this section.

2.1. Inclusion and Exclusion Criteria

Establishing inclusion and exclusion criteria is crucial when performing a literature review in order to guarantee the selection of high-quality literature. These standards aid in streamlining the search and preserving the review's reliability and focus.

Inclusion criteria:

- Research should focus on oral cancer classification (grading) based on histopathological images using AI techniques.
- Research should focus on the segmentation of oral cancer based on histopathological images using AI techniques.
- Research should focus on digital image processing to enhance input histopathological images or extract helpful information.
- Research should focus on interpretability and explainability of AI techniques using histopathological images as input.

Exclusion Criteria:

- ❖ Research articles that do not primarily address the detection of oral cancer through AI techniques.
- ❖ Research that involves animals.
- ❖ Research articles with results lower than 80% accuracy in detecting oral cancer.

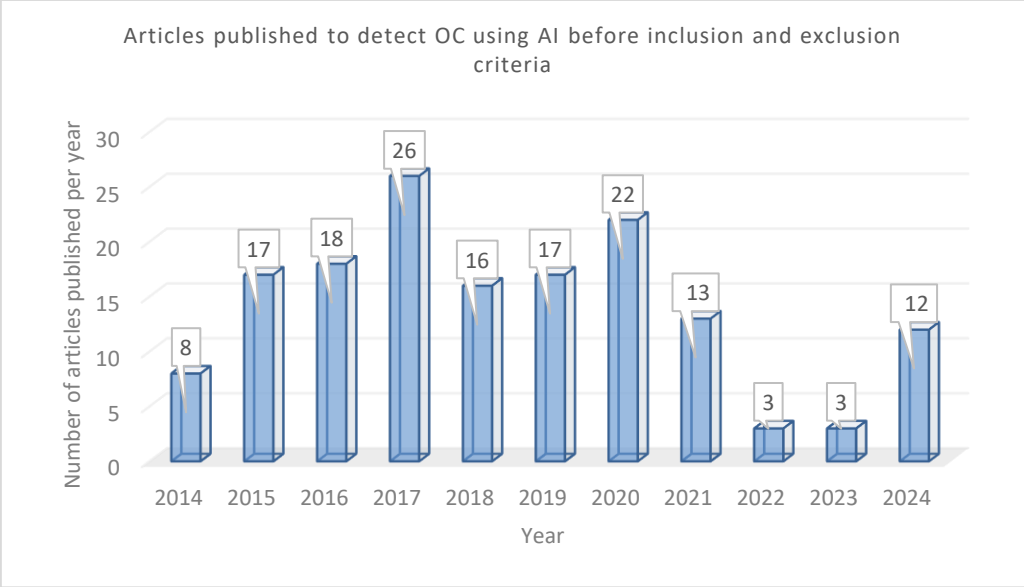

Figure 2.1. Graphical representation of studies published to detect oral cancer using AI techniques – before inclusion and exclusion criteria.

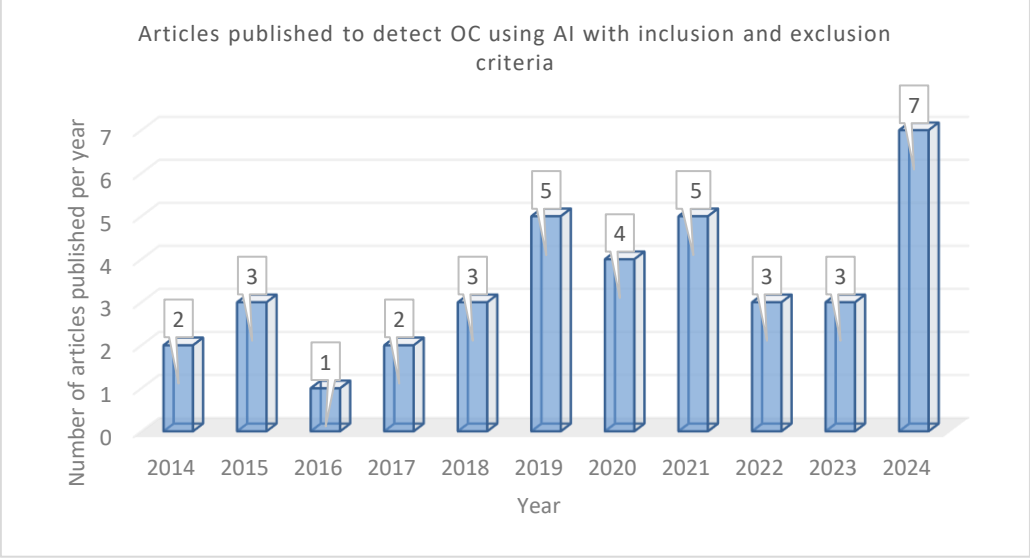

Figure 2.2. Graphical representation of studies published to detect oral cancer using AI techniques – with inclusion and exclusion criteria.

Figure 2.1. shows the amount of research published in the period from 2014. to 2024. to detect oral cancer. The articles that were collected are from various international journals and conferences (Elsevier, IEEE Xplore, Springer, etc.). As can be seen from Figure 2.1., research interest in this area increased in 2017 and is still growing as the field continues to improve.

After reviewing the inclusion and exclusion criteria, 38 articles covering numerous methods for detecting oral cancer based on histopathological images were used to compare the results, as shown in Figure 2.2. However, some of them discuss the nucleus's classification and segmentation, which is outside the scope of this study; thus, these findings cannot be compared. Furthermore, some researchers have included both classification and segmentation in their research, even though, as previously mentioned, studies that include both classification and segmentation of nuclei will not be taken into consideration, they will be described since the techniques are noteworthy.

**Since no single study that has addressed the automated quantification of the tumor-stroma ratio using AI algorithms and histopathological images as input data, the inclusion and exclusion criteria do not refer to the literature on the subject.*

2.2. Application of AI algorithms for OSCC classification

Image classification is the process of dividing data into distinct classes. OC classification determines whether the provided data is malignant or not, and this categorization is known as binary. Furthermore, it is also possible to use multiclass classification to determine the various stages (grades) of cancer.

By employing Random Forests, a tree-based ensemble classifier, Baik et al. (2014) examined a novel, semi-automated technique to separate OPLs at high risk of developing into invasive SCC from those at low risk. For the test set, the novel method demonstrated an 80% right classification rate at the cellular level (80.6% sensitivity, 79.3% specificity) and a 75% correct classification rate at the tissue level (77.8% sensitivity, 71.4% specificity) [8]. Banerjee et al. (2016) assessed the function of morphometric, intensity, and textural features extracted from

liquid-based exfoliative cytology (LBEC), intensity and textural features extracted from ex vivo optical coherence tomography (OCT) images, and spectral features from the difference between mean spectra (DBMS) for classification. The results show that oral leukoplakia and OSCC could be distinguished utilizing cellular characteristics of LBEC data with 100% sensitivity and specificity at 10-fold cross-validation. Effective spectral biomarkers that could identify the disorders with 81.3% sensitivity and 91.3% specificity were also retrieved, illustrating chemical molecules' role in pathological change [9].

Lu et al. (2017) examine computer-extracted features such as texture and nuclear shape on digital H&E-stained images compared to standard clinical and pathologic parameters. To create the oral cavity histomorphometric-based image classifier, a machine learning classifier was used to combine the five best predictive quantitative histomorphometric features from the modelling set. When identifying disease-specific outcomes on the test set, the classifier achieved an AUC of 0.72 [49]. In order to train convolutional neural networks (CNNs) for tissue categorization efficiently, Folmsbee et al. (2018) investigated the use of Active Learning (AL) as opposed to the more popular Random Learning (RL).In the challenge of employing a CNN to detect seven tissue classes (stroma, lymphocytes, tumor, mucosa, keratin pearls, blood, and background/adipose), they compared AL and RL training. For a given training set size, they discover that the AL method outperforms RL by an average of 3.26% [30].

A study by Rahman et al. (2018) attempts to categorize microscopic images of OSCC from histology slides into two groups: abnormal (malignant) and normal (benign). The classification process takes into account the texture characteristics of the images. For feature extraction, GLCM and histogram algorithms are employed. Linear SVM has been utilized for classification, and the outcomes are quite pleasing since 100% accuracy is attained [73]. In their research, Gupta et al. (2019) suggested a deep Convolutional Neural Network (CNN) framework for the classification of images of dysplastic tissue. Normal tissue, mild dysplastic tissue, moderate dysplastic tissue, and severe dysplastic tissue are the four categories into which CNN has divided the provided images. It has been noted that the testing data attains 89.3% accuracy, whereas the training data displays 91.65% accuracy [34].

The automatic diagnosis of oral cancer utilizing histopathology images of oral squamous cell carcinoma is more accurate when features are chosen precisely. Therefore, Nawandhar et al. (2019) have employed the neighbourhood component analysis (NCA) feature selection

technique with a feature weight estimator based on stochastic gradient descent. In order to confirm the effectiveness of the feature selection method and the independence of classifier selection, three popular classifiers are employed. A few chosen features are used to categorize the oral mucosa histopathological images. It has been noted that using feature selection significantly lowers the misclassification rate and increases classification accuracy by 49% to 65% [63]. Creating a CNN model that can classify oral histopathology images as either malignant or non-cancerous is the primary goal of the study by Panigrahi et al. (2019). In order to extract features and classify images of oral cancer, they recommend using convolutional neural networks with four layers (5X5X3). With 10-fold cross-validation, the suggested model's accuracy of 96.77% is comparable to that of pathologists and cytotechnologists [68].

A cutting-edge Inception-V4-based CNN architecture is used for automated SCC detection in the Halicek et al. (2019) study, which details a new and sizable histological SCC dataset of primary head and neck SCC. The training group consists of patches from the tumor and normal tissue samples, while the validation group consists of patches from the tumor-normal margin sample. The testing and validation groups' AUCs for the suggested approach are 0.92 and 0.91, respectively [37]. In their research, Wetzer et al. (2020) examine and assess the following three strategies: (1) special-purpose CNNs that concentrate on texture information extraction; (2) general-purpose CNNs that benefit from pretraining; and (3) data augmentation, which enhances the performance of OC detection. They demonstrate that even with considerable data augmentation and pretraining, texture-focused methods perform better on OC classification than general networks [107].

Rahman et al.'s (2020) aim is to create an exact algorithm that might be applied as an OSCC screening tool. As a result, the binary classification method was adopted to filter out cancerous cases automatically. Using morphological and textural data, they classified OSCC with a decision tree classifier and achieved 99.78% accuracy [74]. The study by Das et al. (2020) aims to categorize OSCC into four classes using Broder's histological grading system. Their research uses two approaches to examine oral biopsy images. First, the best model for their classification problem was identified by applying pre-trained deep convolutional neural networks. Second, a CNN model has been proposed. The experimental results show that the proposed CNN model performed better than the traditional transfer learning algorithms, with

an accuracy of 97.5%, even though the Resnet-50 model attained the best classification accuracy of 92.15% [20].

Wang et al. (2021) used machine learning techniques in conjunction with transmission FTIR imaging to accurately distinguish OSCC biopsy samples from HK samples. Their current study had the following specific goals: 1. produce representative epithelial FTIR spectra from formalin-fixed paraffin-embedded biopsy samples in an efficient and useful manner; 2. characterize HK, OED, and OSCC samples according to their representative spectra; 3. create machine learning models to distinguish OSCC from HK samples, and 4. create a novel approach to categorize OED samples for possible risk stratification applications. Despite the study's limitations, their findings demonstrate that an FTIR-machine learning strategy can accurately distinguish OSCC from HK oral biopsy samples [103].

Panigrahi et al. (2022) introduced a novel method for classifying oral cancer by utilizing the capsule network. The capsule network is more resilient to rotation and affine modification of the augmented oral dataset when it uses dynamic routing and routing by agreement. With 97.78% sensitivity, 96.92% specificity, and 96.77% accuracy, the proposed approach can effectively classify the histological (cancerous and non-cancerous) images of OSCC, according to cross-validation results [69]. Rahman et al. (2022) used biopsy images of oral squamous cell carcinoma to predict malignant and normal mouth tissue using a modified CNN AlexNet. Thus, the suggested model's prediction accuracy and loss rate were 90.06% and 9.08%, respectively [76].

The study by Mohan et al. (2023) suggests OralNet, a framework for detecting oral cancer from histopathology images. The study has four stages: The initial stage involves gathering and preprocessing histopathological pictures in order to get them ready for analysis. Both conventional and deep learning techniques are being used to extract relevant features from images in the second phase, which involves feature extraction utilizing a deep and handcrafted strategy. Concatenation and feature reduction with the artificial hummingbird algorithm (AHA) are part of the third step. Binary classification and three-fold cross-validation performance validation are included in the final step. These involve classifying images as either healthy or OSCC tumor while evaluating the framework’s efficacy using 3-fold cross-validation. According to OralNet's test findings, it could detect oral cancer with more than 99.5% accuracy [60].

The primary goal of the Meyyappan et al. (2024) study is to find a solution to the challenge of distinguishing between benign and malignant histology images. Although the images can be accurately identified by Transfer Learning (TL) models, their research indicates that weighted ensemble learning can improve the model's accuracy to 93.16%, which is higher than the 90% accuracy that individual TL models could reach [57]. In their study, Deo et al. (2024) extracted features from the images using a 2D empirical wavelet transform. The images were then classified into normal and OSCC classes using an ensemble of two pretrained models, ResNet50 and DenseNet201. The model's effectiveness is evaluated and compared in terms of accuracy, sensitivity, and specificity; the suggested model has a maximum classification accuracy of 0.92, according to the simulation results [24].

Das et al. (2024) presented a deep ensemble learning and transfer learning-based classification model for binary oral cancer classification using histopathology images. The advantages of the DL technique can be increased via ensemble learning, which improves accuracy and generalization. In this work, an ensemble model is constructed using the stacking method, outperforming base models with an accuracy of 97.88% [21]. In their study, Maia et al. (2024) examined the use of multiple deep learning architectures to classify histological images of epithelial dysplasia and oral cancer. According to experimental results, there is no statistically significant difference between CNN and transformer models overall. The only model that outperforms transformers is DenseNet-121, which has a balanced accuracy (BCC) of 91.91%, Recall, and Precision of 91.93% [51].

Squeeze-excitation with Hybrid Deep Learning for Oral Squamous Cell Carcinoma Recognition (SEHDL-OSCCR) on HIs was presented by Ragab and Asar (2024) in their paper. Hybrid DL models are the primary tool used in the presented SEHDL-OSCCR technique for the detection of oral cancer. First, the noise is eliminated using the bilateral filtering (BF) approach. After that, the SE-CapsNet model is used by the SEHDL-OSCCR approach to identify the feature extractors. The SE-CapsNet model's performance is enhanced using an enhanced crayfish optimization algorithm (ICOA) approach. Lastly, a CNN with a bidirectional long short-term memory (CNN-BiLSTM) model is used for binary classification. In comparison to more contemporary methods, the experimental validation of the SEHDL-OSCCR technique showed a higher accuracy result of 98.75% [72].

2.3. Application of AI algorithms for semantic segmentation on epithelial and stromal region

Image segmentation is the process of splitting an image into several parts, known as segments. These sections are useful for a straightforward analysis of the digital image. This aids the medical field by enabling more rapid and effective diagnosis.

The computational imaging method for automatic mitotic cell segmentation in OSCC diagnosis is demonstrated in the study by Das et al. (2014). When it came to screening mitotic cells from in vitro histology images, their suggested methodology worked noticeably with Precision of 83,8%, Recall of 73.5% and F-score of 78.3% [15]. The approach proposed by Albasri et al. (2015) shows that it is possible to segment individual cells in a tissue image using a robust algorithm, PCA, and Local Adaptive Thresholding to identify the contour of b-catenin expression found by immunohistochemistry staining of oral cancer [4].
Das et al.'s (2017) paper aims to provide an automated method for counting mitotic cells from relevant histopathology pictures. Regarding this, a novel machine learning approach has been presented that uses a random forest tree classifier that learns across four entropy measures, fractal dimensions, and seven Hu's moments-based descriptors. According to the performance validation, the suggested methodology has an 89% precision, 95% recall or sensitivity, 97.35% specificity, 96.92% accuracy, 96.45% AUC, and 92% F-score measure for effectively detecting mitotic cells from OSCC histological pictures [17].
Wu et al. (2022) created a computerized segmentation model for automatic epithelial segmentation from diagnostic OSCC H&E-stained histology images. They then independently assessed the trained model using images from three separate institutions. Moreover, they demonstrated that the DL model that was trained on tissue microarray (TMA) images can be used to whole-slide images from various locations and pre-analytic variation sources. They also showed that the extraction of morphological features from manually annotated and automatically segmented epithelial sections was equivalent [109].

2.4. Application of AI algorithms for OSCC classification and semantic segmentation

Kumar et al. (2015) described an automated detection and classification process that uses clinically meaningful and biologically interpretable features to detect cancer from microscopic biopsy images. A contrast-limited adaptive histogram equalization technique was employed to improve microscopic biopsy images. Then, the k-means clustering algorithm was applied to image segmentation. Moreover, K-nearest neighborhood (KNN), fuzzy KNN, Support Vector Machines (SVM), and Random forest-based classifiers were used for classification. The average accuracy, specificity, sensitivity, BCR, F-measure, MCC, and specificity for the connective tissues dataset are 0.921909, 0.940164, 0.819922, 0.880263, 0.759395, and 0.717455, respectively [46].

Das et al. (2015) aim to develop a computer-assisted quantitative microscopic methodology for automatic keratinization and keratin pearl region detection using in situ oral histology images. The Chan-Vese approach uses the proposed model to segment the keratinized area. Comparing the model to ground truths based on (manually) experts, the segmentation accuracy is 95.08%. Additionally, a keratinization area grading index is investigated for OSCC cases (poorly, moderately, and well-differentiated) [16].

Moreover, in 2018. Das et al. proposed a two-stage method for computing oral histology images. In the first stage, a 12-layered ($7 \times 7 \times 3$ channel patches) deep convolution neural network is used to segment the constituent layers. In the second stage, texture-based feature (Gabor filter) trained random forests are used to detect keratin pearls from the segmented keratin regions. For epithelial layer segmentation, their approach achieved an average of 98.42% segmentation accuracy, 97.76% sensitivity, 90.63% Jaccard index, and 95.03% dice coefficient. Furthermore, the proposed approach achieved an average segmentation accuracy of 98.05%, a Jaccard index of 71.87%, and a dice coefficient of 75.19% for the keratin region. The keratin pearl recognition accuracy of the suggested texture-based random forest classifier is 96.88% [18].

Das et al. (2019) developed a two-stage computational pipeline for automatic nucleus recognition and segmentation from oral histology images with the aim of assisting healthcare

professionals in diagnosing OSCC. The nucleus is efficiently detected (88.87% recall and 82.03% precision) in the first stage using a 12-layer CNN driven by wavelet downsampled patches, and in the next phase, the AC-NSCT-based nucleus segmentation technique achieves comprehensive accuracy (Dice coefficient of 94.22%, Jaccard index of 89.38%, Precision of 97.56%, and Recall of 91.58%) for its automatic delineation [19]. An automated, effective computer-aided system for diagnosing the normal and malignant (OSCC) categories has been proposed by Rahman et al. (2020). The images' color, texture, and shape have all been retrieved. Various classifiers were used to achieve classification. For form, textural, and color features, respectively, accuracy of 99.4% was obtained using the Decision Tree Classifier, accuracy of 100% using SVM and Logistic Regression, and accuracy of 100% using SVM, Logistic Regression, and Linear Discriminant [74].

Segmentation, object recognition, and image classification are the three deep learning (DL) techniques compared in the Matias et al. (2021) research. Their findings demonstrate that the most effective method for detecting and localizing nuclei is detection using Faster R-CNN (0.76 IoU. ResNet 34 performed well in classifying abnormal nuclei (0.86 scores). Therefore, they deduced that these two models could be combined to create a dependable pipeline for localization and classification [54].

The study by Hameed et al. (2021) uses a blue color component feature-based SVM classifier to build an automatic IHC scoring technique. Entropy thresholding is used to partition the tissue images, and the watershed transform is applied selectively to resolve clustered cells. Using a SVM, each cell nucleus in tissue pictures is categorized as positive or negative based on the staining intensity. The J-scoring technique is then used to obtain the tissue score. The feature that was taken from the blue component achieved the maximum classification accuracy of 98.01%, with sensitivity and specificity of 98.86% and 94.74%, respectively, according to the testing results [38].

In their study, Sujatha et al. (2021) improve the image by removing noise. The preprocessed image is then sent to the segmentation process, using the Patch-based Fuzzy Local Similarity CMeans (PFLSCM) scheme. They used feature extraction techniques to extract the feature from the image. Ultimately, a Hybrid Hopfield Neural Network with an Ant Colony Optimization (ACO) algorithm is used to accurately identify retrieved features images. The accuracy of the suggested model was 98.98% [90].

Using oral histopathology images, Musulin et al. (2021) propose a two-stage AI-based system for automatic multiclass grading (the first stage) and segmentation of the tumor on epithelial and stromal tissue (the second stage) to aid the clinician in diagnosing oral squamous cell carcinoma. Semantic segmentation prediction using DeepLabv3+ and Xception_65 as backbone and data preprocessing produced mIOU of 0.878 and F1 of 0.955 score, while the combination of Xception and SWT produced the highest classification value of 0.963 AUCmacro and 0.966 AUCmicro [62]. The study by Shetty et al. (2023) creates a design for the detection of oral cancer in a scattered cloud environment. Following initial preprocessing, images were segmented using a region-growing algorithm. Graph, textural, and morphological aspects are also retrieved. The characteristics in this study were selected using the suggested Linear Discriminant Analysis. Ensemble classifiers are used for the chosen features in order to classify cancer. Additionally, stage 1 incorporates the Multi-layer Perceptron (MLP) and Support Vector Machine (SVM) models for disease categorization. The optimal CNN, which determines whether oral cancer is present, is part of the stage 2 phase [83].

In their study, Dharani and Danesh (2024) suggested two novel deep-learning techniques for OSCC segmentation and identification: MaskMeanShiftCNN and SV-OnionNet. While SV-OnionNet is appropriate for classifying oral cancer and normal oral tissues, MaskMeanShiftCNN segments OSCC regions from input images using color, texture, and shape. The suggested techniques achieved a classification accuracy of 98.94%, sensitivity of 98.96%, specificity of 97.18%, and error rate of 1.05%, outperforming current methods for OSCC detection [22].

In an effort to improve diagnostic precision, Shukla et al. (2024) presented a unique method for cancer diagnosis that uses machine vision. Unlike conventional deep learning or supervised algorithms, they use an unsupervised approach for cancer identification due to the complexity of histopathology images. Because of its essential features and shape, the nucleus is recognized as the region of interest (ROI) in a biopsy image of malignant tissue. For the last step of cancer identification, they use a unique binary classification method and K-means clustering enhanced with a thresholding strategy to extract the ROI. The suggested model is more effective and dependable at detecting cancer since it achieved an accuracy of nearly 97.28% with a closely followed validation accuracy of roughly 96.34% [85].

The literature indicates that the majority of researchers have used AI algorithms in retrospective studies to detect and classify oral cancer. It is evident that binary classification, which uses the image's color, form, and texture, constituted most of the classification tasks. Deep CNN architectures were used to complete most segmentation tasks using histopathology images. A shortcoming of the aforementioned studies is that they were trained to determine miotic cells from relevant data.

The only deep learning model for classifying cells into multiple classes in OSCC epithelial tissue was proposed by Das et al. (2020), based on a literature review. The dataset consisted of image patches derived from whole slide biopsy images. The proposed CNN model resulted in accuracy of 97.5% [20].

According to a thorough literature review, at the time when this research was performed, no studies had been done on multiclass grading along with segmenting of OSCC using histopathology images obtained by biopsy and stained with marker protein.

2.5. Automatic quantification of tumor-stroma ratio

The predictive value of the tumor-stroma ratio in various cancer types has been validated by multiple investigations using manually examined histopathology images. However, the subjective nature of pathologists and the variability of observers render manual visual evaluation techniques inappropriate for extensive implementation in clinical practice. Recent advancements in artificial intelligence and digital pathology have made it possible to perform additional quantitative analysis on numerous histopathological images.
Hong et al. (2021) introduced a DL-based TSR measuring tool for advanced gastric cancer [40], while Zhao et al. (2020) used whole-slide HE-stained images to demonstrate a deep-learning (DL) model for completely automated TSR quantification of colorectal cancer [115]. Furthermore, Millar et al. (2020) used machine learning algorithms and digital image analysis to determine the clinical importance of tumor stroma ratio in luminal and triple negative breast cancer (TNBC) [59]. Using H&E-stained images of bladder cancer, Zheng et al. (2023) created a machine learning technique for the quantitative evaluation of TSR [117].

Smit et al. (2023) examined whether completely or semi-automated uses of artificial intelligence (more precisely, deep learning algorithms) could produce comparable outcomes in automated analysis, particularly for hard-to-score cases. The study found that the TSR evaluated by deep learning algorithms and using a microscope had good relationships [87].
Their fully automated techniques allow for objective and consistent application while reducing the workload of pathologists. Most of these papers looked at the TSR quantification of various cancer forms, although OSCC is the primary focus of this study.

2.6. Explainable computer vision for OSCC classification

Explainable deep learning (XDL) has drawn a lot of interest in the field of artificial intelligence, particularly in domains such as medical imaging, where accurate and understandable machine learning models are crucial for effective diagnosis and treatment planning [89]. In order to enhance reliability and confidence in results, Grad-CAM is a baseline that determines the key image regions used in a deep learning model's decision-making. There are several computer vision (CV) uses for it, such as classification and explanation [89].
To improve diagnostic reliability and interpretability, Grad-CAM has been used in a variety of studies to classify cancer images with higher performance.
Oya et al. (2022) aimed to investigate ability of AI to evaluate OSCC by employing a novel training approach that considers cellular and structural atypia and their applicability. The convolutional neural network model that was employed was EfficientNet B0. The use of gradient-weighted class activation mapping provided insight into its validity. The proposed method achieved an accuracy of 99.65% using images with 512 × 512 pixels as input. Grad-CAM results showed that the AI model covered both the cellular and structural atypia of SCC, focusing on the region around the basal layer [66]. The study by Afify et al. (2023) proposes a novel model that employs Grad-CAM and deep transfer learning to identify the lesion area in the image in order to predict oral squamous cell carcinoma. The results of the proposed method are noteworthy since they demonstrate the clinical community's crucial role in the prompt and accurate detection of oral cancer [3]. The performance of two DL models that are renowned for their high accuracy in oral cancer classification was thoroughly evaluated by

Da Silva et al. (2024) in order to better understand the potential and constraints of DL methods in the context of oral cancer diagnosis. Beyond just analyzing standard accuracy measures, they additionally examined subclass accuracy rates and the distribution of prediction confidences, furthermore, they used Grad-CAM to visualize the models' decisions. [14].

3. Oral Squamous Cell Carcinoma

This chapter aims to provide an overview of oral squamous cell carcinoma, first focusing on its clinical features and diagnostic procedures, and then discussing the latest advancements in computer-aided diagnosis.

3.1. Clinical Presentation and Diagnosis

Oral cancer makes up 2% to 4% of all cancer cases worldwide. The most prevalent malignant epithelial neoplasm that affects the oral cavity is oral squamous cell carcinoma [53]. The GLOBOCAN database estimates that 377 713 new cases were diagnosed in 2020, and 177 757 deaths were reported [91]. The morbidity and mortality rates for OSCC have not changed much over the past 30 years, despite improvements in therapy techniques [7]. OSCC frequently develops from pre-existing oral mucosal lesions that have a higher chance of developing into cancer. Even though the oral cavity is easily accessible for clinical inspection, OSCC is typically detected in advanced stages. However, early detection and care at the precancerous stage increases OSCC survival rates and the morbidity associated with treatment [31]. The primary therapy for OSCC is usually surgical resection, either with or without adjuvant radiotherapy, which significantly affects the patient's quality of life [27].

In the Western world, smoking tobacco and drinking alcohol are the most significant risk factors for oral cancer. Although the risk factors are independent, they appear to work together. Smoking tobacco is associated with 75% of all cases of oral cancer, and the risk of developing oral cancer is six times higher for smokers than for non-smokers. Additionally, alcohol drinkers are six times more likely to develop oral cancer than non-drinkers [52]. Even though alcohol and tobacco use are typically the most significant risk factors, it is crucial to consider other known risk factors, like chewing betel quid in some ethnic groups.

Other factors also contribute, such as immune defects, deficiencies in vitamins A, E, C, or trace elements, and an impaired capacity to metabolize carcinogens and repair DNA damaged by mutagens [52]. Risk factors are demonstrated in Figure 3.1.

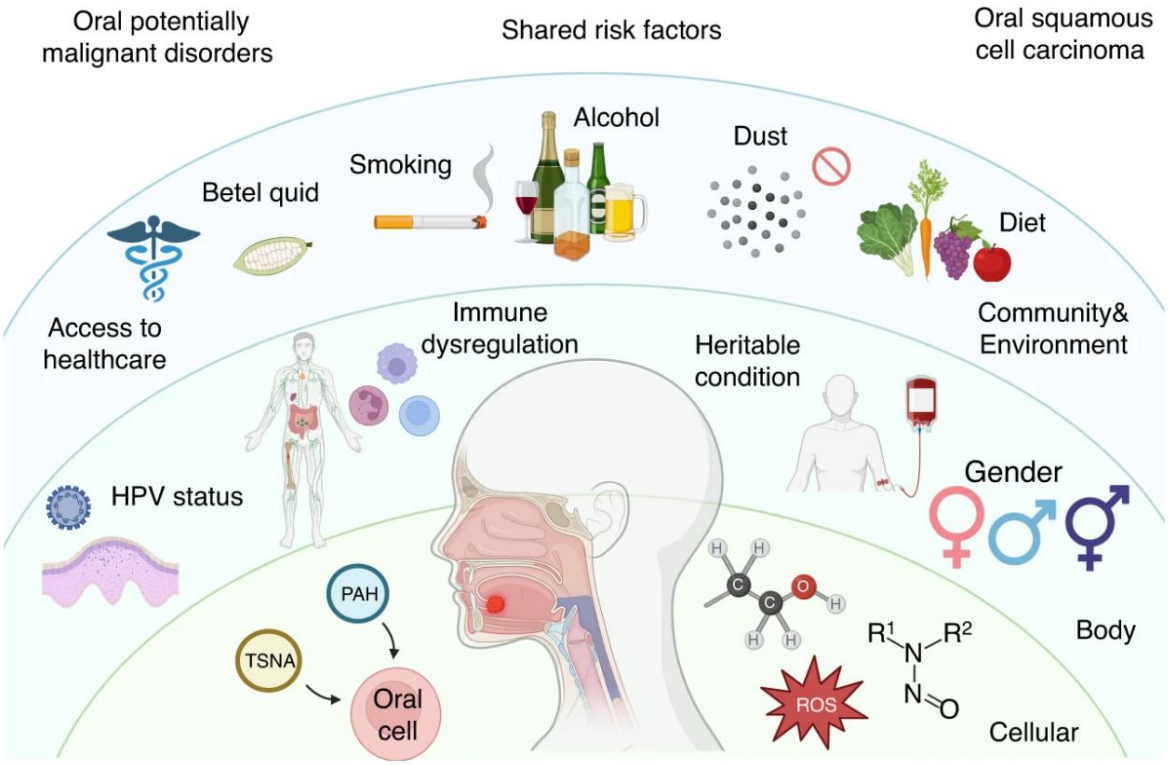

Figure 3.1. Risk factors, such as malnutrition, immunological deficiencies, smoking, alcohol misuse, chewing betel quid (BQ), human papillomavirus (HPV) infection, and genetic disorders [97].

In the USA, the median age of diagnosis for OSCC is 62 years, but the incidence of OSCC in people under 45 is rising. The reason why OSCC affects men more often than women (M:F = 1.5:1) is that more men engage in high-risk behaviors than women. The likelihood of developing OSCC rises with the length of time that a person is exposed to risk factors, and growing older adds the additional dimension of age-related mutagenic and epigenetic changes [28]. The most common locations for the malignant neoplasm are the oral cavity floor, the tongue's lateral borders, and the lip.

OSCC may appear as one of the following [7]:

- an area of redness (erythroplakia),
- a white lesion (erythroleukoplakia),
- higher exophytic borders or fissuring in a granular ulcer,
- a unilateral lesion on the buccal mucosa or tongue's lateral edge that is red and white,
- an ulcer or indurated lump, which is a solid infiltration beneath the mucosa,
- and an ulcer or crust that has been present for more than three weeks on the vermilion edge of the lip (rule out herpes simplex).

An example of OSCC in patients is presented in Figure 3.2.

Figure 3.2. The tongue is where 30% of oral cancers originate, followed by the lip (17%) and the floor of the mouth (14%). HPV-related oropharyngeal cancer primarily affects the tonsil and tonsillar pillars, the base of the tongue, and the oropharynx [25].

Despite significant progress in comprehending the intricate process of carcinogenesis, no trustworthy predictive tool has been discovered. For the prognosis, treatment strategy, and outcome prediction of oral cancer in patients with OSCC, tumor-node-metastasis (TNM) staging is commonly utilized. The limitation of TNM staging in prognostic prediction is reflected in its deficiency to include clinical features as well as personal traits of the patient, such as lifestyle choices [58]. The current gold standard for detecting oral cancer is:

- clinical examination,
- conventional oral examination (COE),
- and histological evaluation following biopsy.

These approaches can identify cancer in the stage of established lesions with notable malignant changes [105]. The International Histological Classification of Tumors classifies the lesions based on the degree of tumor differentiation [78]:

- Grade I - well differentiated,
- Grade II - moderately differentiated,
- Grade III - poorly differentiated.

Most medical centers base their decisions upon clinical and pathological medical data. The main determinants of the therapeutic approach are the TNM stage, the degree of tumor differentiation, and the patient's health status [78]. An example of tumor differentiation is shown in Figure 3.3.

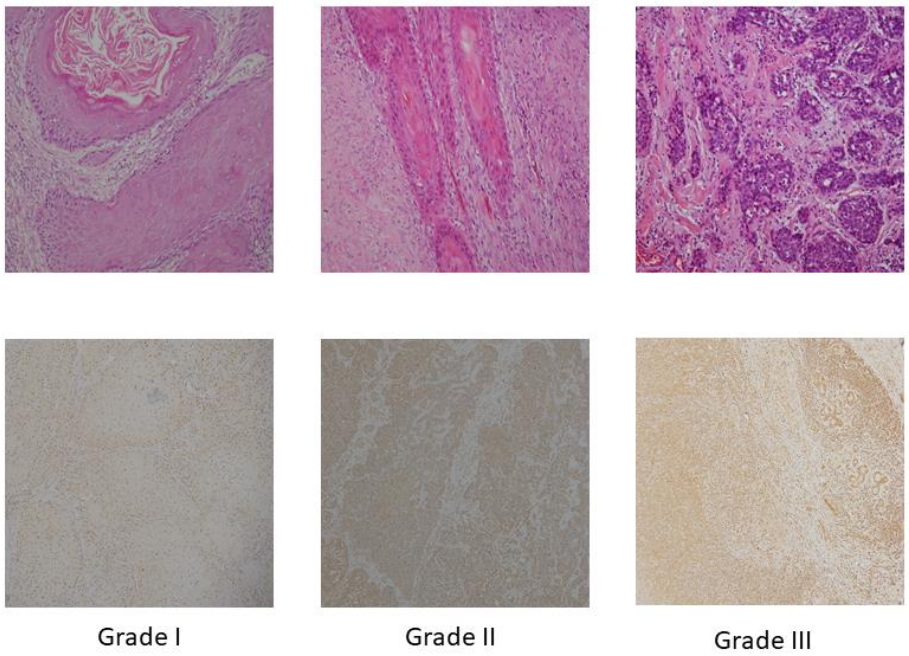

Figure 3.3. OSCC group of Grade I, Grade II and Grade III. First row represents H&E-stained images while the second row represents IHC-stained images.

3.2. Advances in Computer-aided OSCC Diagnosis

The primary issue with histological examination for tumor differentiation is the subjective nature of the examination, specifically the intra- and inter-observer variability [55]. By determining patient outcomes, computer-aided diagnosis systems (CAD) that increase objectivity and accuracy while decreasing inter- and intra-observer variability could immediately impact patient-specific therapeutic interventions. Additionally, such an approach could help the pathologist make quicker, more accurate conclusions and reduce the workload associated with manual inspections [55]. Due to recent AI and image processing developments, CAD systems can now recognize and classify OSCC with near-human or even better performance.

3.2.1. Role of Artificial Intelligence algorithms in OC analysis

The development of artificial intelligence may enhance the screening process for OC. AI can accurately analyze an enormous dataset from multiple imaging modalities and help in healthcare, primarily in the field of oncology [110]. Fundamentally, AI aims to enable computers to perform operations that usually require human intelligence, such as learning, problem-solving, applying logic, and making rational choices [101]. This covers a wide range of techniques and strategies, including robotics, computer vision (CV)-image analysis, natural language processing (NLP), machine learning (ML), and deep learning (DL) [101]. Within science, AI enables the development of personalized treatment strategies by incorporating patient-specific data, such as genetic profiles and medical histories, to create personalized interventions based on unique traits, maximizing effectiveness and reducing side effects [45]. Ability of AI models to recognize molecular signs and biomarkers reinforces the idea of personalized healthcare by making it easier to formulate treatments that specifically target the mechanisms causing cancer to advance. Figure 3.4. lists the use of AI in OC detection and treatment.

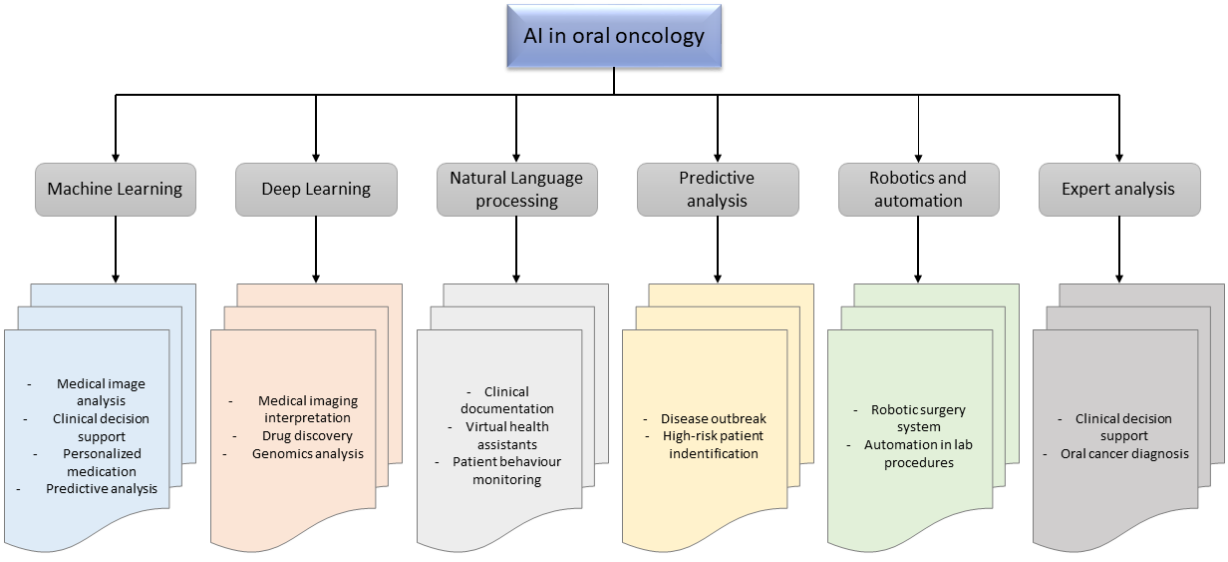

Figure 3.4. A visual representation of AI in oral oncology; it facilitates the use of various cutting-edge technologies for imaging, diagnosis, prediction, patient monitoring, and therapy automation.

3.2.2. Application of Artificial Intelligence algorithms in OC analysis

By using various data sources to increase accuracy and efficiency, machine learning techniques are proven to be extremely useful tools in OC detection and diagnosis [6]. ML algorithms absorb data, identify trends, and make predictions without the help of humans. Deep learning uses multilayer artificial neural networks to analyze and interpret complex medical data in the healthcare industry [6]. This technology has the potential to completely transform several aspects of healthcare, including patient care management, personalized medicine, treatment planning, and diagnostics. In clinical practice, the application of DL to oral cancer data may assist healthcare professionals diagnose, identify, and forecast prognoses for oral cancer. This allows for early diagnosis and therapy selection, which increases the survival rate of patients with oral cancer [23].

Around the world, hospitals are quickly switching from paper-based to electronic medical data. In the healthcare industry, natural language processing (NLP) is essential for gathering and interpreting data from medical records. By enabling improved clinical documentation, analysis, and decision-making, this technology has the potential to completely transform the way OC is identified, treated, and managed. Through the extraction of pertinent data from pathology reports, radiological findings, and medical notes of OC, NLP can automate the clinical documentation process [44].

Predictive analytics is one of the big data analytics that is becoming increasingly significant in clinical care. Risk stratification, differential diagnosis, illness occurrence prediction, and intervention efficacy prediction are just a few of the clinical medicine applications of predictive analytics. In order to create predictive models that can aid health professionals with early detection, individualized treatment planning, and disease progression monitoring, this approach uses a range of data types, such as patient demographics, medical records, genetic traits, and clinical pathological results. Predictive methods like these can improve prior treatment planning, enable more individualized therapy approaches, and improve patient outcomes managing OSCC [48]. Figure 3.5. gives information about the role of ML, DL and NPL in OC treatment and diagnosis.

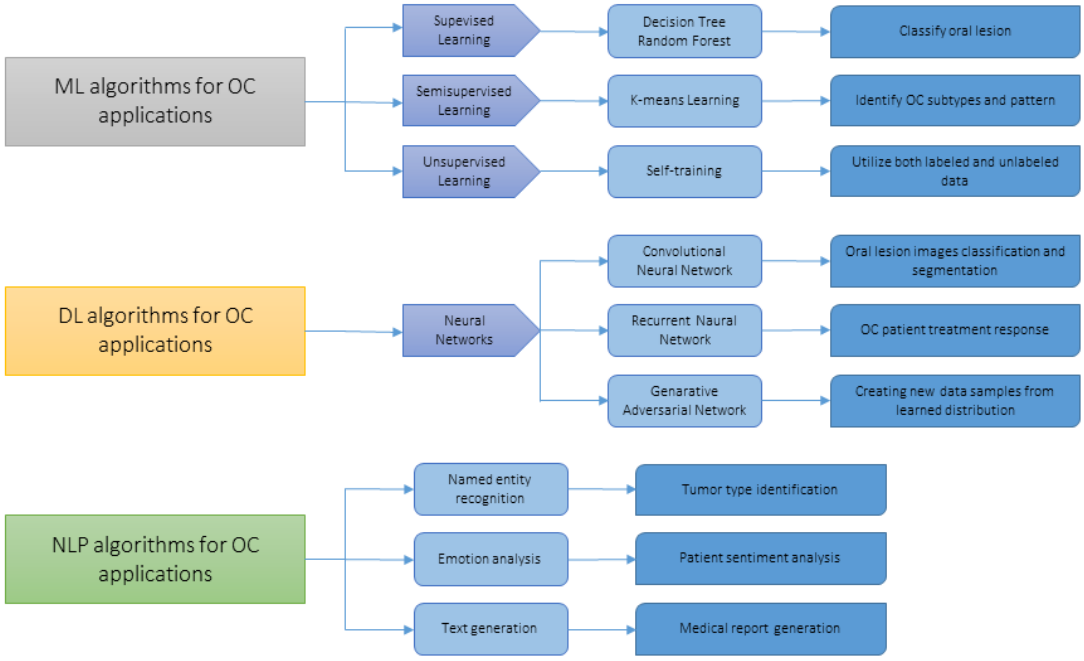

Figure 3.5. An illustration of machine learning, deep learning, and natural language processing algorithms used in oral cancer, including their particular methods and associated clinical tasks like data generation, clinical text analysis, lesion classification, subtype identification, and treatment response prediction.

Due to improvement in robotics and automation, OC therapy and surgical procedures are facing tremendous advancements, which provide creative solutions that increase accuracy, and shorten recuperation periods. OC surgeons can now perform complex procedures with more control and precision with the development of robotic-assisted technologies. Automation technologies can improve surgical results' consistency by optimizing several OC treatment procedures, including tissue sampling, suturing, and organ retraction. This would lower the possibility of human error [56].

Another kind of AI is expert systems, created to imitate experts' judgment in particular fields. Expert networks in OC management may play a significant role in giving healthcare professionals information by evaluating patient data, making suggestions, and supporting treatment planning and monitoring [2]. Figure 3.6. gives information about the role of predictive analysis along with robotics and automation in OC treatment and diagnosis.

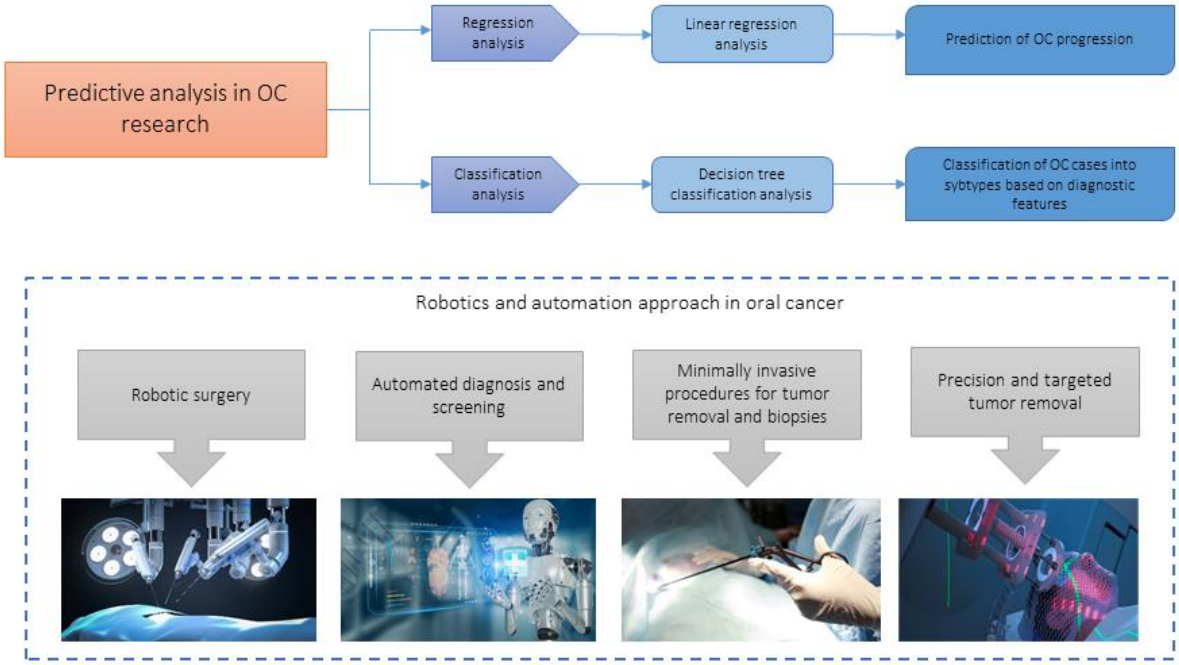

Figure 3.6. An outline of predictive analysis approaches and robotics in oral cancer, demonstrating how robotic technologies improve diagnosis, screening, and precise tumor removal while regression and classification methods contribute to disease progression prediction and subtype identification.

4. Dataset Description

The aim of this chapter is to provide a thorough overview of the dataset, including patient demographics and metadata, data collection, and the procedures used for dataset preparation and splitting.

4.1. Data Collection and Sources

A dataset consisting of 322 histology images with 768 x 768-pixel size was created for this research. The Clinical Department of Pathology and Cytology's archives of the Clinical Hospital Center in Rijeka provided the formalin-fixed, paraffin-embedded oral mucosa tissue blocks of instances of oral squamous cell carcinoma that were histopathologically documented. Two independent pathologists examined the sample slides, and they were categorized in accordance with the World Health Organization's (WHO) [26]. The Kappa coefficient was used to evaluate the pathologists' degree of agreement. Kappa coefficient score was 0.94.

Briefly, a range of marker proteins were used to stain paraffin-embedded tissue slices that were 4 μm in size using the conventional IHC methodology. DAB and hematoxylin were employed to stain the IHC images. The light microscope (Olympus BX51, Olympus, Japan) with a digital camera (DP50, Olympus, Japan) was used to capture the images, and CellF software (Olympus, Japan) transferred the images to a computer. Moreover, images were captured with 10x objective lenses.

As illustrated in Figure 4.1., images have been categorized into three classes based on the previously established classification.

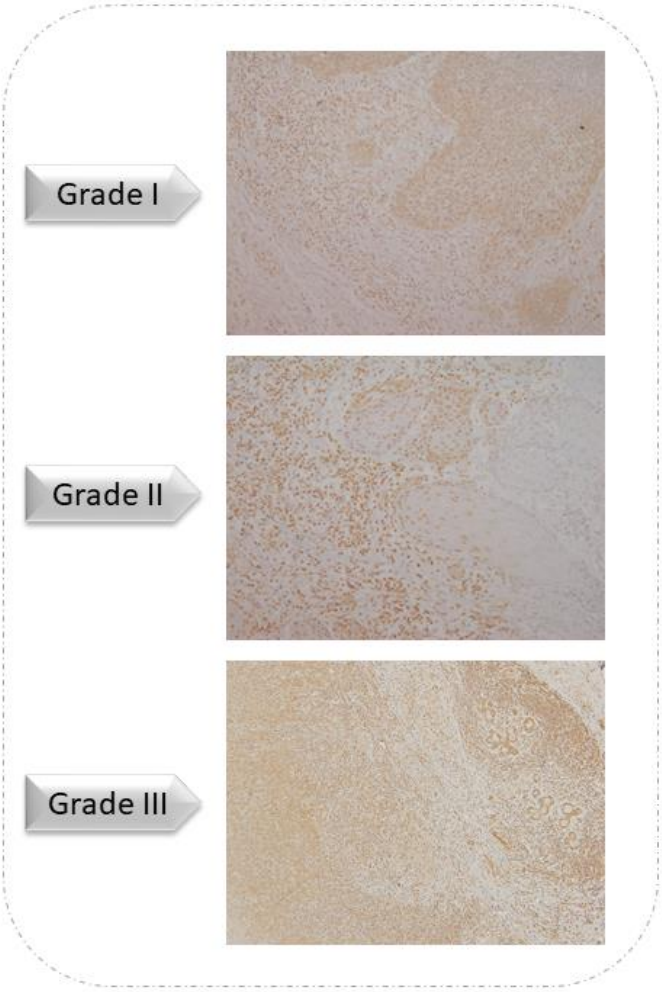

Figure 4.1. The OSCC group of Grade I, Grade II and Grade III under x10 magnification.

An additional dataset of 101 histopathological images was collected for experimental proof of concept in order to ensure the proposed AI-based system's robustness. The protocol for collecting additional images was the same as for the original data set.

This research guarantees data quality, representativeness, and generalizability for AI-driven analysis in medical research by carefully choosing data sources and upholding strict ethical norms.

4.2. Patients Demographics and Metadata

Medical dataset analysis is extended by the contextual information provided by patient demographics and metadata. Table 4.1. shows a comparable clinic-pathological report for the patients. Demographic information included the patient's age at diagnosis, sex, smoking status, and alcohol use.

70% of the patients were men and 30% were women. The median age among adult patients was 64. Of the patients, 55% smoked, and 38% consumed alcohol. 45% of patients were diagnosed with a Grade I, while only 15% were diagnosed with a Grade III. More patients (52%) had lymph nodes metastases.

Table 4.1. Characteristics of the patients include sex, age, smoking and alcohol habits, presence of metastases in the lymph nodes, and grade of OSCC.

Characteristics of the patients		**n = 40 (100%)**
Sex	F	30
	M	70
Age	To 49	5
	50 – 59	13
	60 – 69	55
	+70	27
Smoking	Y	55
	N	45
Alcohol	Y	38
	N	62
Lymph Node Metastases	Y	52
	N	48
Grading	I	45
	II	40
	III	15

4.3. Data preparation

4.3.1. Segmentation mask construction

In medical imaging, segmentation mask construction is an essential task that a medical expert can manually perform. Mask is an array or matrix that highlights areas of interest in an image. It marks which pixels are part of a specific item, class, or region. Each pixel in the mask corresponds to a pixel in the original image.

Standard annotation software and tools include Labelbox, GIMP, ImageJ, ITK-SNAP, 3D Slicer, and CVAT [65]. These tools offer features like region-growing methods, brush tools, and polygonal annotation.

Figure 4.2. shows OSCC images with corresponding segmentation masks which are created using GIMP.

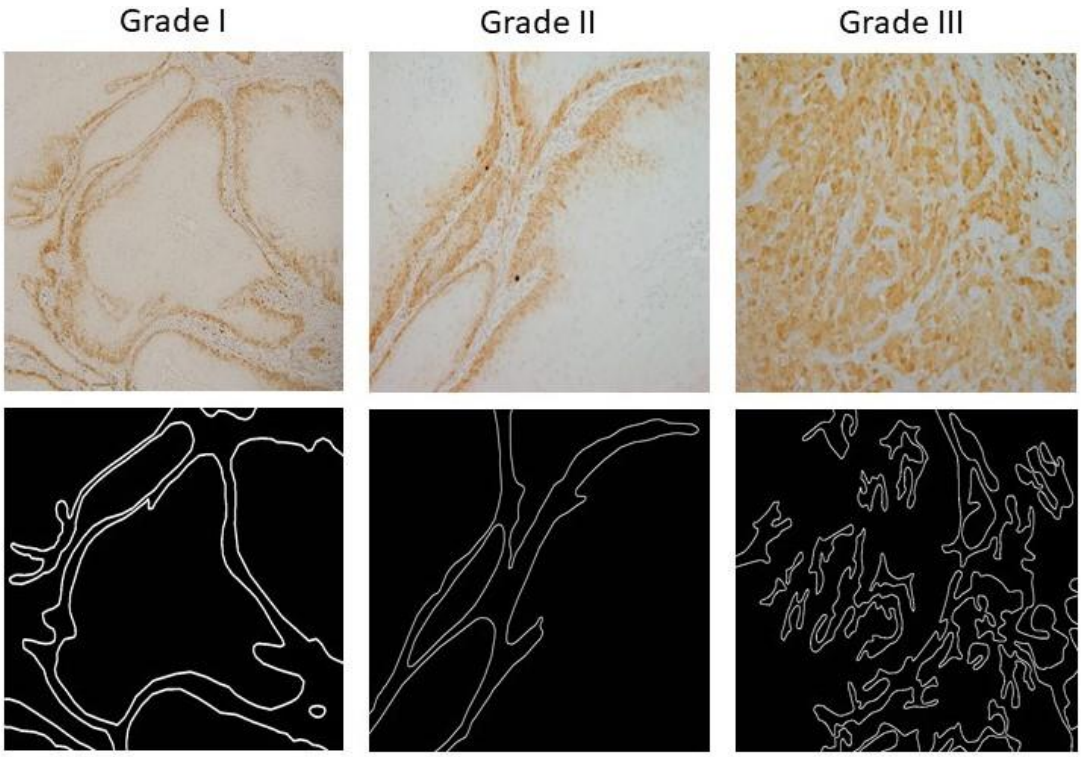

Figure 4.2. Group of well-differentiated, moderately differentiated and poorly differentiated OSCC along with segmentation masks.

4.3.2. Image Augmentation

Deep convolutional neural networks are strongly dependent on many samples to achieve good performance and prevent overfitting. However, since fields like medical image analysis sometimes lack access to a large number of samples augmentation techniques are required. Image augmentation is the process of applying different modifications to increase the size and diversity of a dataset. It prevents overfitting and improves model generalization. Due to previously mentioned neural network demand and the restricted availability of data, in this research, augmentation techniques such as geometric transformations are used to artificially increase the quantity of samples.

Geometrical transformations used for the augmentation procedure are shown in Figure 4.3.

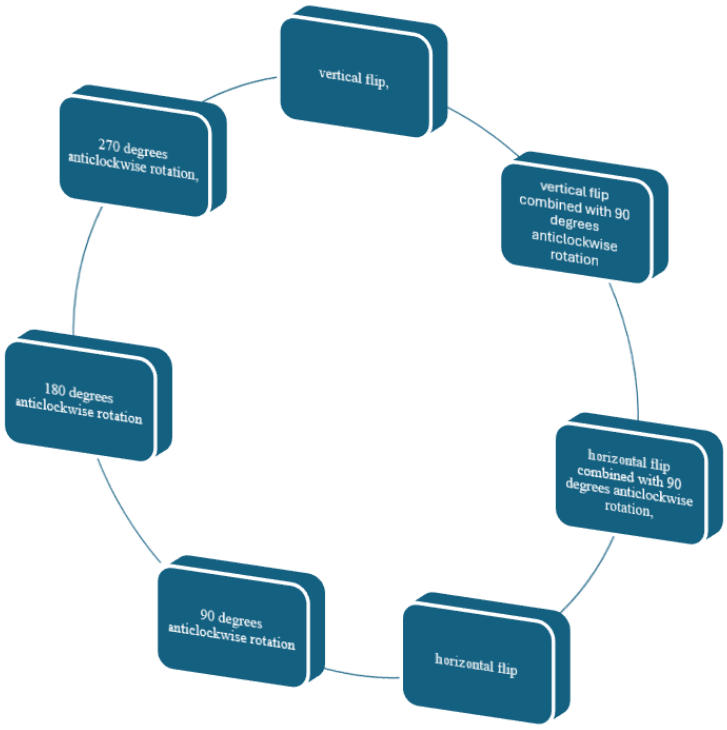

Figure 4.3. Geometrical transformations for augmentation procedure.

Testing samples are not augmented. The augmentation method is only utilized to create training samples since newly created data are variations of the original data.

As seen in Figure 4.4., a new training set including an additional 1799 images has been created after applying geometrical transformations, resulting in a total of 2056 images.

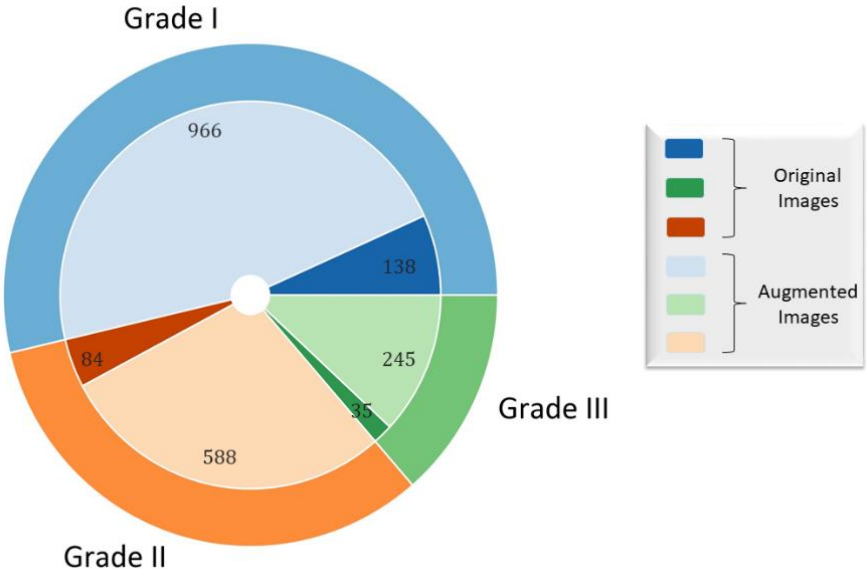

Figure 4.4. Visual representation of the original and augmented dataset.

4.4. Dataset Splitting

A reasonable data splitting approach is essential for model validation and developing a model with strong generalization performance. Although there are other data splitting techniques that have been described and utilized in the literature, cross-validation (CV) is the most used. CV divides the data into *k* different parts, often known as k-folds. The validation set represents a single fold. The model is trained on the remaining k-1 folds and then used in the validation set to record its predictive performance [112]. In order for each part to be used as a validation set once, this process was repeated *k* times. After averaging the recorded predictive performances, the model parameter with the best average predictive performance is identified as optimal.

In this research, due to the high imbalance among OSCC classes the performance of AI-based models is estimated using stratified 5-fold cross-validation. In this manner, each class is roughly represented throughout all test folds.

5. Image Preprocessing

Significant increases in processing capacity and developments in image analysis techniques over the past decade have made it possible to create robust computer-aided analytical tools for medical data. With the development of whole-slide digital scanners, tissue histopathological slides can now be scanned and stored digitally. Whole slide imaging (WSI) is frequently used to examine tissue samples and diagnose cancerous diseases. However, some scanning equipment, staining techniques, and tissue reactivity can cause color variations in histopathology images, making it difficult to analyze them. This chapter gives an overview of preprocessing techniques used in this research in order to aid computers comprehend histopathology images for diagnostic purposes.

5.1. Normalization of Histopathological Images

Digital histopathology is a field of study that uses techniques such as color normalization and feature extraction that aid computers comprehend histopathology images for diagnostic purposes [108]. However, variations in color in histopathological images might lead to issues. The stain or dye used to prepare histopathological images typically gives the image a different hue. The results of analyzing images without preprocessing could lead to an inaccurate diagnosis [35].

One tissue staining method that pathologists are particularly interested in is hematoxylin and eosin (H&E) staining. Pathologists can quickly identify and analyze tissue sections according to the H stain, which highlights nuclei in blue against the pink background of the cytoplasm and surrounding structures [118].

Figure 5.1. shows OSCC H&E histopathological images of well- and moderately differentiated OSCC with magnification x10.

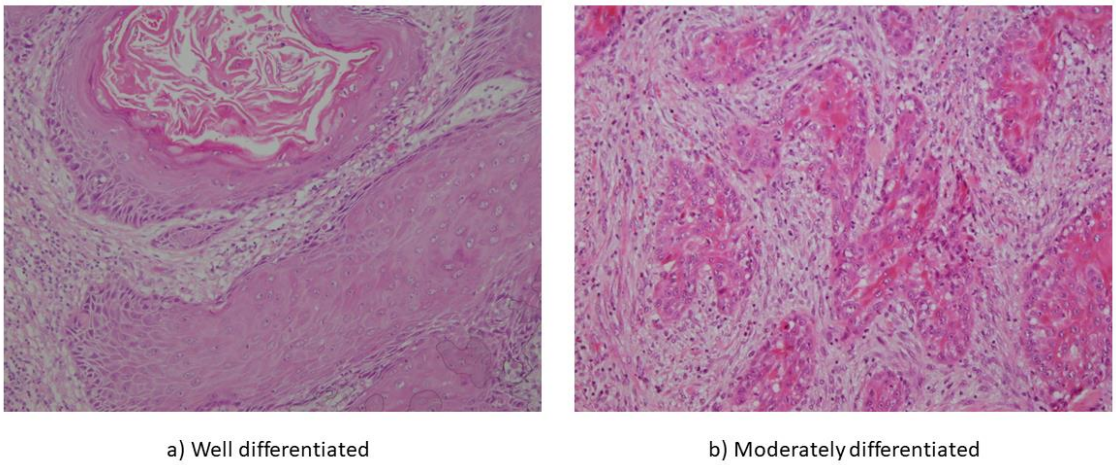

Figure 5.1. Tissue slides of well- and moderately differentiated oral carcinoma.

The reasons for color diversity in histopathology images are heterogeneous stain coloring, chemicals from different manufacturers, and the use of various scanners and equipment during slide preparation [42]. Therefore, to ensure visual consistency in histopathology images, color normalization is necessary.

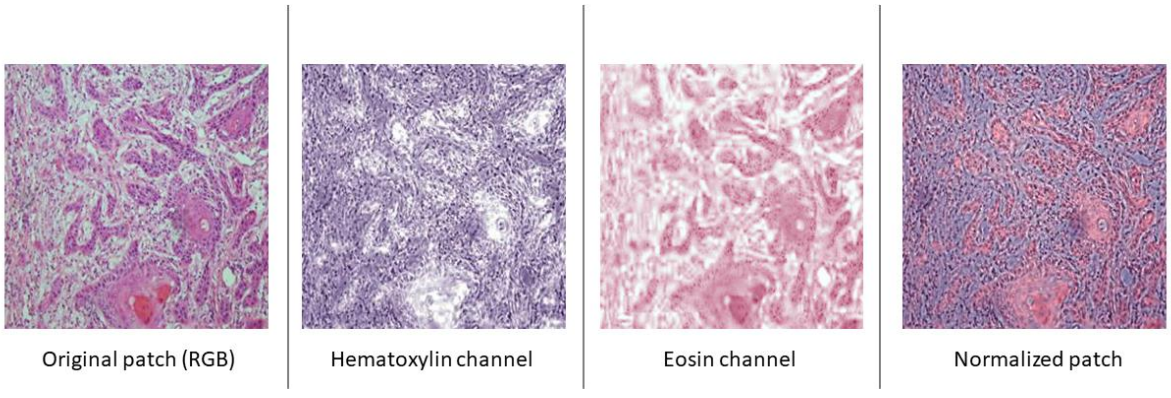

Figure 5.2. An illustration of H&E stain normalization shows the initial RGB patch, the separated hematoxylin and eosin channels, and the final normalized patch for a uniform histopathological image presentation.

Histopathology images can be color-normalized using a variety of algorithms, including the Reinhard method, Macenko method, stain color descriptor (SCD), histogram specification, complete color normalization, and structure preserving color normalization (SPCN). However, Macenko method is the most used color normalizing technique when utilizing H&E stained images.

Figure 5.3. shows H&E histopathological images of OSCC before and after color (Macenko) normalization.

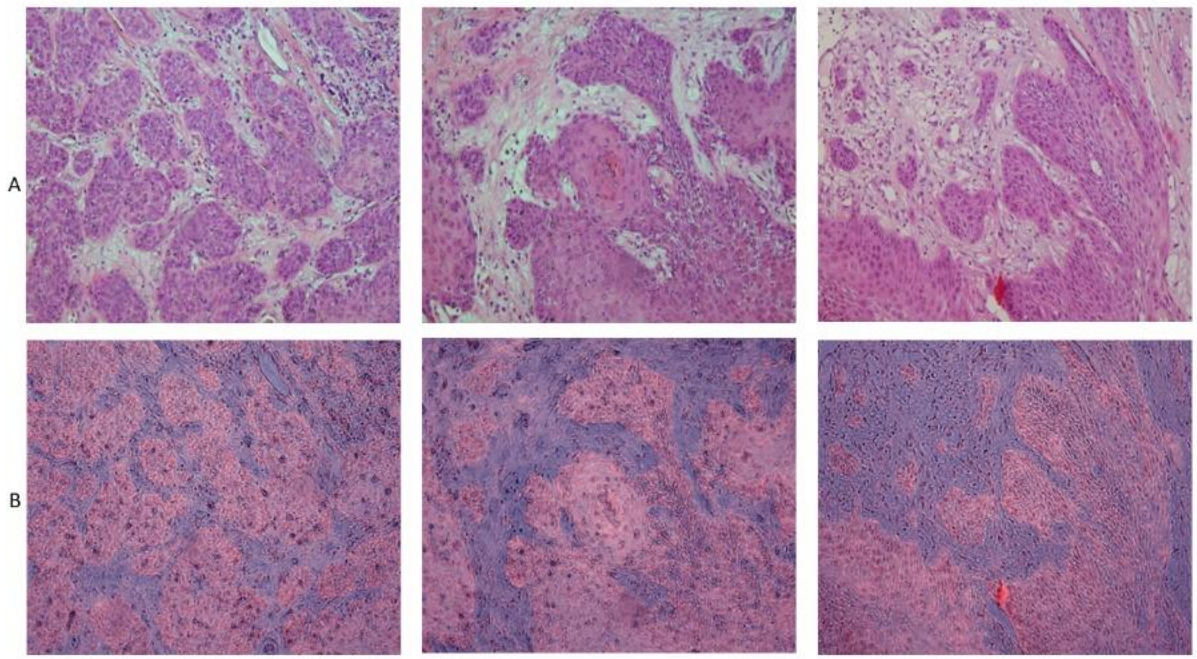

Figure 5.3. Visual representation of A) H&E-stained images and B) normalized H&E-stained images.

Obtained results reveal that the application of preprocessing method, such as Macenko image normalization for image analysis, has great potential as the first step in the diagnosis of OSCC. However, in our research, the histopathological sections were treated with two different antibodies. A polyclonal rabbit anti-megalin antibody (Santa Cruz Biotechnology, USA; also diluted 1:100 in the same buffer) and monoclonal mouse anti-MT I+II antibody (clone E; DAKO, USA) was employed. A standard immunohistochemistry methodology was followed throughout the process. Diaminobenzidine (DAB) was added to a peroxidase substrate in order to visualize the immunological response.

After visualization, the slides were dehydrated, stained with hematoxylin (Sigma, Germany), and then mounted in Entelean (Sigma).

The aforementioned examples show how to use the Macenko approach to preprocess H&E-stained histopathology images. However, this study employs IHC histopathology images, whereas all the previous examples are centered around H&E images. An example of an IHC image using the Macenko approach is presented in Figure 5.4.

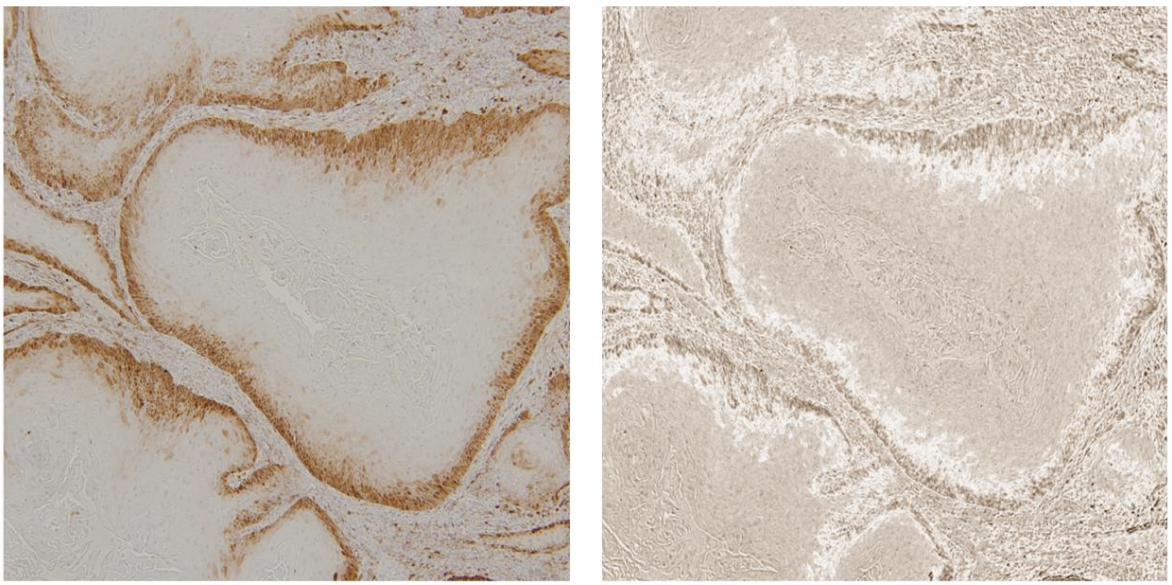

Figure 5.4. Visual representation of IHC stain normalization.

Figure 5.4. shows that even though Macenko method is a widely used color normalization technique for H&E images, it is not well-suited for immunohistochemical staining. Using two dominating stain vectors, usually representing hematoxylin (blue) and eosin (pink), Macenko is based on optical density (OD) deconvolution. Assuming that there are only two stains in the color space, it conducts stain separation and calculates stain vectors using singular value decomposition (SVD). Depending on the antibody and detection method, IHC slides are stained with chromogens such as hematoxylin (counterstain) and DAB (brown), or different combinations. This research staining process uses various antibodies, which produce chromogen patterns that deviate from the Macenko method's H&E presumptions.

Therefore, the normalizing process may result in images that seem faded, with areas that are strongly stained losing detail and contrast. When one of the components, the hematoxylin or the DAB is insufficient, Macenko normalization may not be able to distinguish them correctly. That will result in a merging of tissue features and an incorrect representation of color.

Moreover, the Macenko approach dismisses significant spatial information, such as microstructural texture and local architectural patterns, which are important for later deep learning tasks, especially for transformer-based segmentation models that rely on multi-scale contextual reasoning. To solve these limitations, this research employs a preprocessing pipeline based on the Stationary Wavelet Transform (SWT) for classification task and Luminance Wavelet Enhancement (LWE) for semantic segmentation task, described in chapter 5.2. and 5.3.

5.2. Preprocessing Method Based on SWT

Wavelet Transform (WT) is a powerful method frequently employed in data preprocessing [1]. Wavelet transformation examines spatial frequency components at various scales rather than depending on color deconvolution. This facilitates the maintenance of fine and global structural characteristics of the tissue, making it resistant to changes in scanner lighting and staining intensity.

Wavelet transform of signal x(t) can be defined as [95]

$$X(\tau, a) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t)\psi^* \left(\frac{t-\tau}{a}\right) dt, \tag{5.1}$$

where [114]:

- ⇒ a is the dilation,
- ⇒ ψ is the analyzing wavelet, and
- ⇒ τ is the translation parameter.

The Discrete Wavelet Transform (DWT) of signal x[m] can be calculated as follows [95]

$$X[k,l] = 2^{-\frac{k}{l}} \sum_{m=-\infty}^{\infty} x[m]\psi[2^{-k}m - l]\,. \tag{5.2}$$

The Discrete wavelet transform (DWT) can be applied independently along each dimension during image processing. As a result, the image is divided into four subbands: LL, LH, HL, and HH. While the detail coefficients are represented by LH, HL, and HH, the approximation coefficients can be identified as the LL subband [71]. Although DWT is simple to implement and reduces computing time, it has drawbacks in terms of decimation and shift-invariance. In order to overcome the aforementioned issues, this research utilizes Stationary Wavelet Transform (SWT), which enables the decomposition of histopathological images.

The advantages of SWT are as follows [43];

- improved time-frequency localization,
- no decimation step, which provides duplicate information and
- invariance of translation.

Following the SWT decomposition process, a mapping function is used to weigh the derived coefficients. This enables the further enhancement of important features of an image. The mapping function is determined by incorporating the following factors:

- only detail coefficients undergo coefficient mapping and
- both details with high and low coefficient values are heavily weighted, as they preserve important information.

Wavelet coefficient mapping function can be mathematically defined as follows:

$$y_{i,j} = aw_{i,j}^3 + bw_{i,j}^2 + cw_{i,j} + d, \tag{5.3}$$

where;

⇒ a, b, c, and d represent constants,

⇒ $w_{i,j}$ is an input coefficient, and

⇒ $y_{i,j}$ is a coefficient after mapping.

An improved image is obtained by performing the SWT reconstruction using weighted and approximate SWT coefficients after the coefficient mapping procedure. Figure 5.5. illustrates the SWT decomposition, coefficient mapping, and SWT reconstruction procedure.

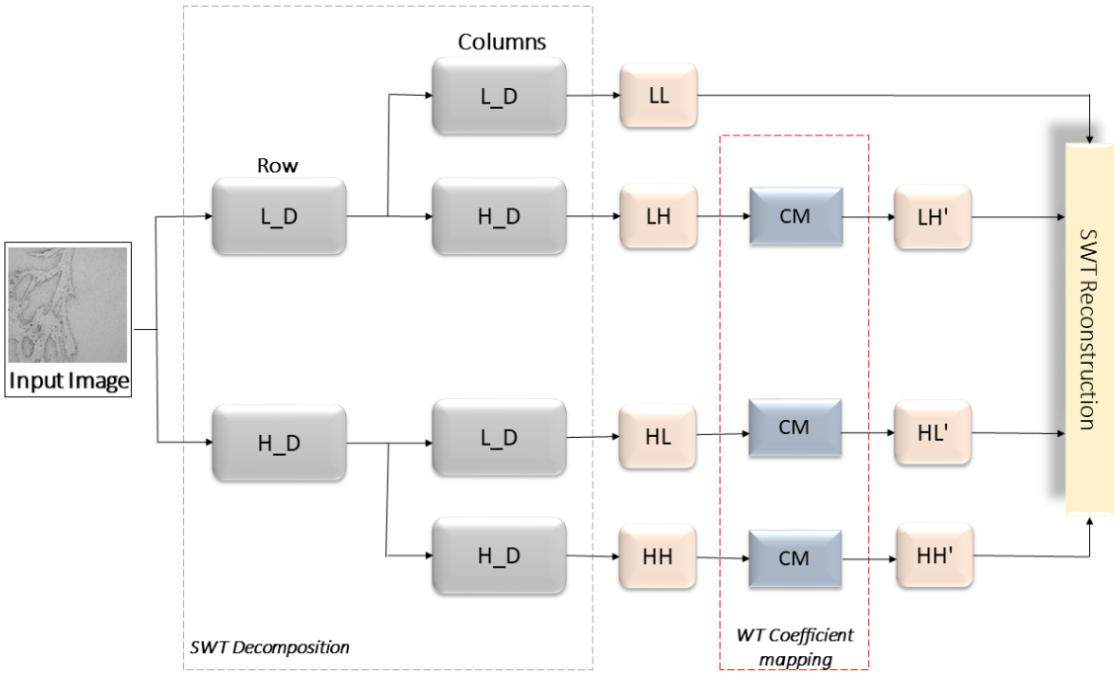

Figure 5.5. The following symbols are used to represent wavelet coefficient mapping, SWT reconstruction, and SWT decomposition: LL for approximation coefficients, LH for horizontal coefficients, HL for vertical coefficients, HH for diagonal coefficients, CM for coefficient mapping function, and L_D for low pass filter and H_D for high pass filter.

The wavelet function and mapping function have a direct impact on the quality of weighted coefficients, therefore careful selection of these values is essential. Since it is highly computationally costly to evaluate each value in huge search-spaces, conventional methods for determining parameters, like random search or grid search, may not always be practical [29].

These approaches choose the next parameter configuration without considering the assessed performance of previous iterations, which often results in time spent assessing the function with suboptimal parameter selection. The Bayesian technique, on the other hand, chooses the subsequent parameter configuration for the mapping function based on the outcomes of previous iterations [92]. This approach outperforms more conventional approaches by achieving convergence to the optimal solution in reduced iterations. In order to determine the most suitable values for the wavelet function and wavelet coefficient mapping function constants (a, b, c, and d), Bayesian optimization has been utilized.

The domain of mapping function constants over which to search is defined and shown in Table 5.1.

Table 5.1. Combination of the 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

Wavelet transform decomposes signals or images into their component parts at various frequency and spatial scales using mathematical bases called wavelet functions, such as Haar, sym2, db2, and bior1.3. Due to its distinct characteristics, each wavelet can be used for specific types of signal or images.

5.3. Luminance Wavelet Enhancement (LWE)

In order to improve the structural representation of immunohistochemistry images before they are transmitted to the segmentation model, the proposed preprocessing technique introduces Luminance Wavelet Enhancement (LWE). Unlike global color normalization techniques which largely focus on modifying stain intensity distributions, LWE directly tackles the spatial aspects of the image by increasing texture and border information included within the luminance channel of the LAB color space [100]. Figure 5.6. shows pipeline of the proposed LWE preprocessing approach.

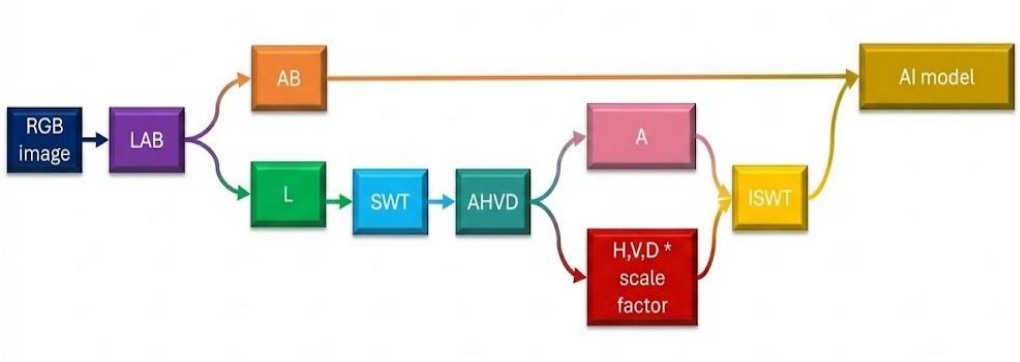

Figure 5.6. Illustration of the Luminance Wavelet Enhancement (LWE) preprocessing pipeline, displaying the transition from RGB to LAB color space and subsequent processing steps: L (luminance channel), AB (chromatic channels), SWT (Stationary Wavelet Transform), AHVD (approximation and horizontal, vertical, and diagonal detail coefficients), ISWT (Inverse Stationary Wavelet Transform).

The RGB image is translated into the LAB color space as follows:

$$I_{LAB} = f_{RGB \to LAB}(I_{RGB}). \tag{5.4.}$$

After converting the input image to LAB space, the luminance component L is retrieved and decomposed using the SWT as shown:

$$(L) \overset{SWT}{\rightarrow} \{A, H, V, D\} \tag{5.5.}$$

where SWT offers three high-frequency detail sub-bands (horizontal, vertical, and diagonal), each of which captures structural information at a distinct orientation, in addition to an approximation sub-band. Since SWT is shift-invariant and does not include downsampling, its coefficients preserve complete spatial resolution, making this approach suitable for downstream pixel-accurate segmentation.

Structural detail is increased by scaling each high-frequency component with a scale factor as seen in Eq. 5.6.:

$$H' = kH, V' = kV, D' = kD. \tag{5.6.}$$

The high-frequency detail coefficients are intentionally increased to highlight delicate morphological details and reinforce boundary cues. This regulated enhancement corrects for challenges such as poor staining, inconsistent illumination, and low contrast, which are frequently observed in IHC slides and may hide diagnostically relevant patterns.

The enhanced coefficients are then combined again using the inverse SWT to reconstruct an enhanced luminance channel L and can be described as follows:

$$L' = ISWT(A, H', V', D'). \tag{5.7.}$$

Color differences that are biologically significant are preserved because chromatic channels A and B do not change. The increased luminance channel L is combined with the chromatic channels as seen in Eq 5.8.:

$$I'_{LAB} = [L', A, B]. \tag{5.8.}$$

6. Artificial Intelligence Algorithms

This chapter presents the artificial intelligence algorithms used in this research, emphasizing techniques for multiclass classification and semantic segmentation.

6.1. AI algorithms for multiclass classification

By combining AI algorithms with medical image analysis, large and complex datasets can be analyzed in real time and provide insights that can improve patient outcomes. This chapter gives a brief description of most used image classification algorithms.

6.1.1. ResNet50 and -101

The well-known vanishing gradient issue enables deep neural networks more challenging to train. To facilitate deep neural network training, He et al. (2016) proposed a residual network (ResNets) [39]. Authors improved the residual block and its pre-activation version, allowing vanishing gradients to move freely to any other earlier layer via shortcut connections. In the ResNet50 architecture, each 2-layer block in the 34-layer network is swapped out for a 3-layer bottleneck block, producing 50 layers. On the other hand, the ResNet101 architecture is built with additional 3-layer blocks, as shown in Table 6.1.

Table 6.1. ResNet50 and ResNet101 architecture representation.

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

He et al. (2016) demonstrated on the ImageNet dataset that ResNets perform better than other topologies on the ILSVRC classification test, with an error of 3.57% [39].

6.1.2. InceptionV3

InceptionV3 was the concept proposed by Szegedy et al. (2015) after InceptionV1 and InceptionV2 [93]. Its main goal is to reduce the amount of computing power by altering earlier Inception designs. To relieve the limitations for simpler model adaptation, InceptionV3 has proposed several network optimization techniques, such as factorized convolutions, regularization, dimension reduction, and parallelized computations.

Setting a new state of the art, their best quality version of Inception-v3 achieves 21.2%, top 1, and 5.6% top-5 error for single crop evaluation on the ILSVR 2012 classification. Figure 6.1. illustrates the Inception-v3 architecture, which had one input block, two grid size reduction blocks, three Inception Modules A, B, and C blocks, one auxiliary classifier block, and one output block.

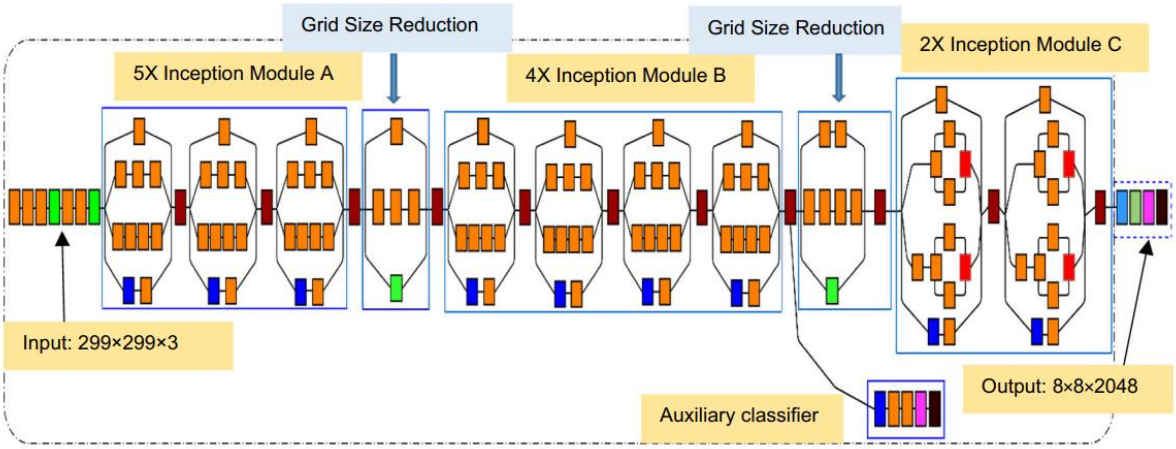

Figure 6.1. Block diagram of InceptionV3 architecture [86].

6.1.3. InceptionResNetV2

Szegedy et al. (2016) present the InceptionResNetv2, which combines the Inception design with residual connections [94]. It increases network efficiency and permits deeper penetration without running into issues like vanishing or gradient explosion. The network gains greater depth, improved processing power, and stronger nonlinearity through the breakdown of the convolution kernel. The presented design shown in Figure 6.2., significantly increases training speed and enhances recognition performance.

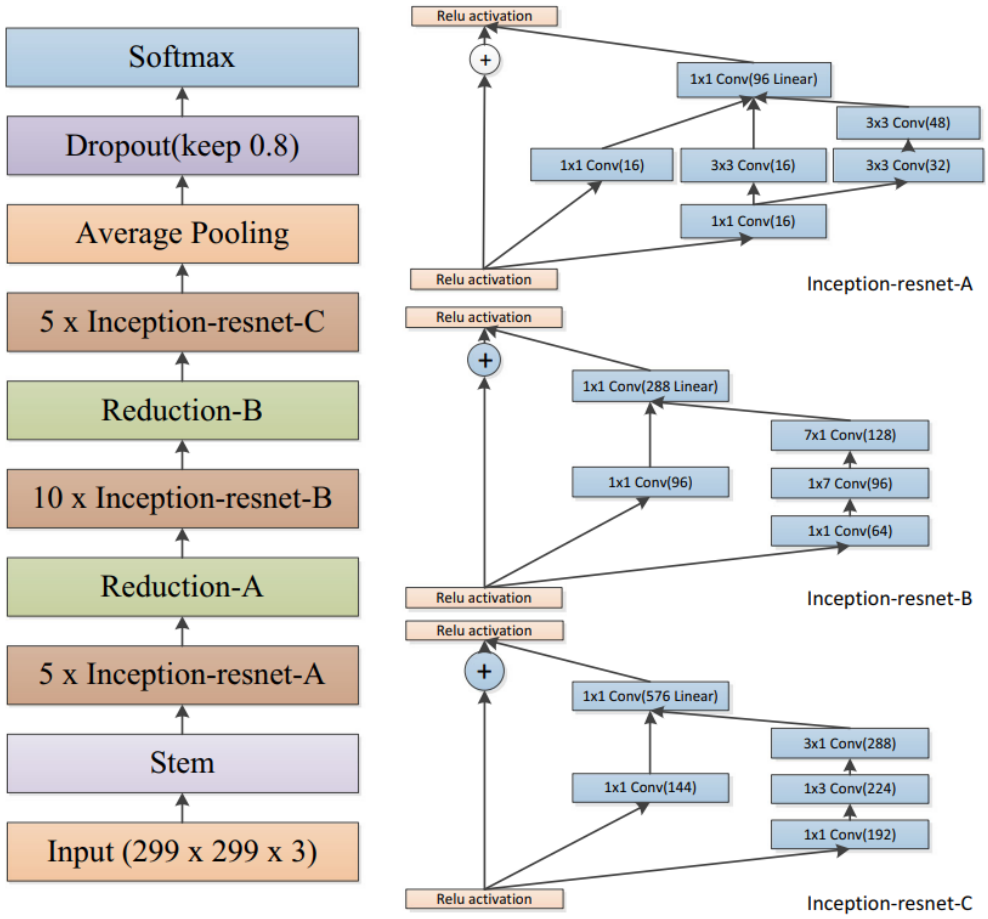

Figure 6.2. Diagrams of the overall network structure and module structure of InceptionResNetV2 [70].

6.1.4. Xception

In 2017. Chollet presented a new architecture called Xception [13]. Convolutional layers in a conventional convolutional neural network seek correlation by navigating over space and depth. Xception goes a step further by independently mapping the spatial correlations for every output channel and capturing cross-channel correlation through 1x1 depth-wise convolution. The 36 convolutional layers that comprise the Xception architecture are organized into 14 modules [13]. Figure 6.3. shows that every module, aside from the first and last modules, has linear residual connections surrounding it.

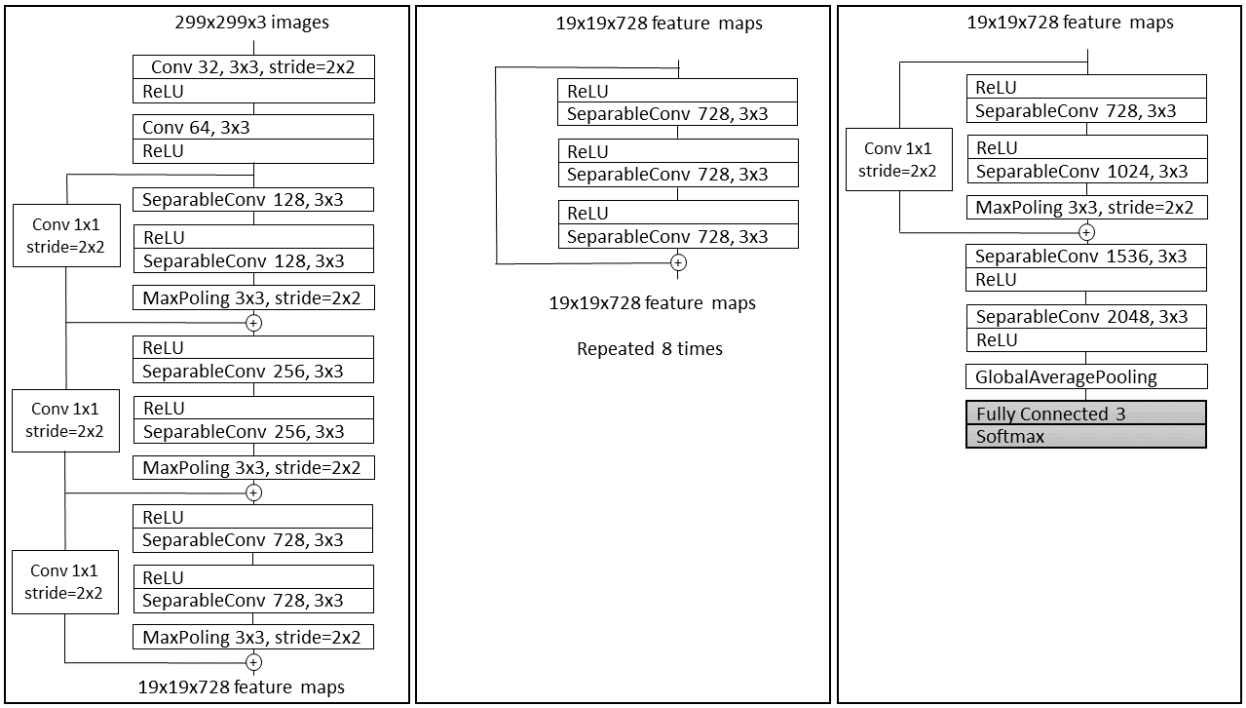

Figure 6.3. Xception architecture; the data propagates eight times, first through the input flow and then through the middle flow. Furthermore, data moves through the third box, representing the exit flow.

6.1.5. MobileNet

Due to its significant memory and computational demands classical CNN is not suitable for use on mobile and embedded systems. For that reason, Howard et al. (2017) proposed MobileNetV1, a lightweight network designed for embedded and mobile applications.

In 2018. in order to enhance the functionality of mobile models, Sandler et al. introduced MobileNetv2 architecture [81]. It expands on the concepts of MobileNetV1 by using depthwise separable convolution as effective building blocks. MobileNetV2 uses small bottleneck layers as input to the residual block, unlike traditional residual models that employ an extended input representation. Table 6.2. demonstrates a detailed architecture structure.

Table 6.2. Each row in the MobileNetV2 architecture represents a set of identical layers that have been repeated n times. Every layer in a sequence has the same number of output channels (*c*). The initial sequence's layer employs a stride of *s,* but the subsequent layers use a stride of 1. The expansion factor (*t)* determines the size of the input.

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

6.1.6. NasNet

By framing the task of determining the optimal CNN architecture as a reinforcement learning problem, Zoph et al. (2018) developed NASNet [117]. The main idea was to find the optimal parameters inside the specified search space, including strides, number of layers, output channels, filter sizes, etc. NASNet proposes identifying two kinds of cells: reduction and normal cells. Reduction cells are primarily utilized to lower spatial resolution, while normal cells are used to extract advanced information while maintaining the exact spatial resolution.

The depth of a network defines the search space, allowing the discovery of effective architectures using a small dataset (e.g., CIFAR-10) and enabling the transfer of the learned architecture to image classification tasks across various data sizes and computational scales.

The generated architectures outperform state-of-the-art performance on both the CIFAR-10 and ImageNet datasets while demanding less computational effort than human-designed architectures. An illustration of a two-cell search space is shown in Figure 6.4.

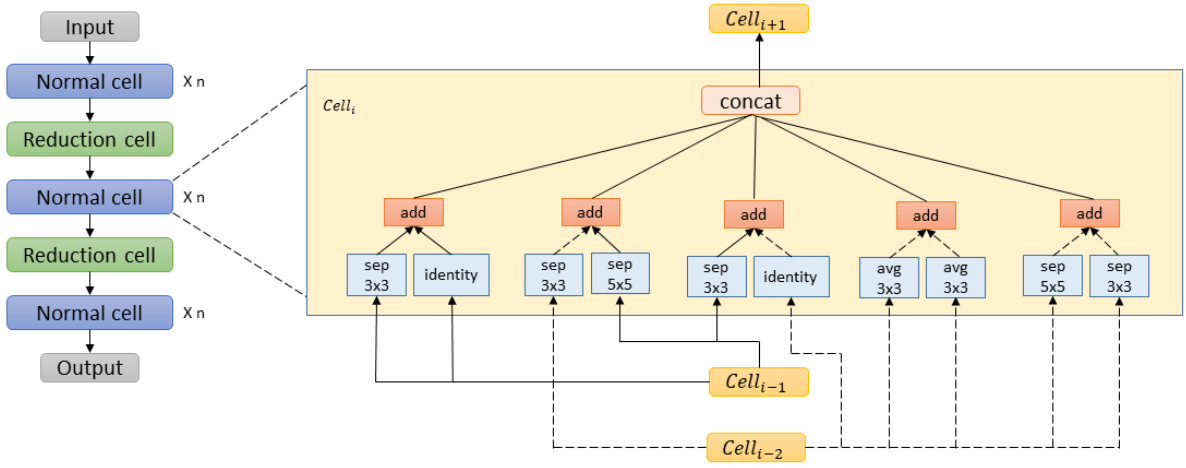

Figure 6.4. Left: An illustration of a two-cell search space. Right: An illustration of the ideal design for a typical cell.

6.1.7. EfficientNetB3

EfficientNet, which Tan and Lee first presented in 2019, quickly gained prominence as the preferred architecture for various demanding applications, such as language processing, image segmentation, and object recognition [96]. The reason for success is its capacity to compromise model performance and computing efficiency, two essential aspects of deep learning.

The EfficientNet family of models, which includes EfficientNetB3, is regarded as a balanced and effective model. Compound scaling is one of EfficientNet-B3 key features. It automatically scales up the model's architecture in terms of width (number of filters per layer) and depth (number of layers), depending on the input image resolution. With this model, it can be analyzed both large and small images more efficiently with better results while using more resources. The architecture of EfficientNetB3 can be summarized as presented in Table 6.3.

Table 6.3. EfficientNetB3 architecture.

Stage	Operator	Resolution	Number of channels	Layers
1	Conv3x3	300x300	32	1
2	MBConv1, k3x3	150x150	16	2
3	MBConv6, k3x3	150x150	24	3
4	MBConv6, k5x5	75x75	40	3
5	MBConv6, k3x3	38x38	80	5
6	MBConv6, k5x5	19x19	112	5
7	MBConv6, k5x5	10x10	192	6
8	MBConv6, k3x3	10x10	320	2
9	Conv1x1 & Pooling & FC	10x10	1280	1

6.2. AI algorithms for semantic segmentation

In addition to multiclass classification, semantic segmentation is an essential AI-driven method in medical image analysis, especially for identifying OSCC. Semantic segmentation provides pixel-by-pixel classification, thus allowing to precisely identify malignant areas in histopathology images. Deep learning architectures like U-Net, DeepLabV3+, and transformer-based models are presented in this chapter.

6.2.1. U-Net

Ronneberger et al. (2015) in their research presented U-Net, a well-known deep learning architecture. Having both contracting and expanding pathways is an advantage of the U-Net architecture. The contracting path gradually lowers the input's spatial resolution by using encoder layers to extract contextual features. The expanding path, on the other hand, uses skip connections from the contracting path to precisely create the segmentation map by including decoder layers that reconstruct the encoded representation [80].

This network was created to efficiently utilize a smaller amount of data while preserving speed and accuracy, with the main goal of addressing the problem of limited annotated data in the medical field [80]. U-Net architecture is shown in Figure 6.5.

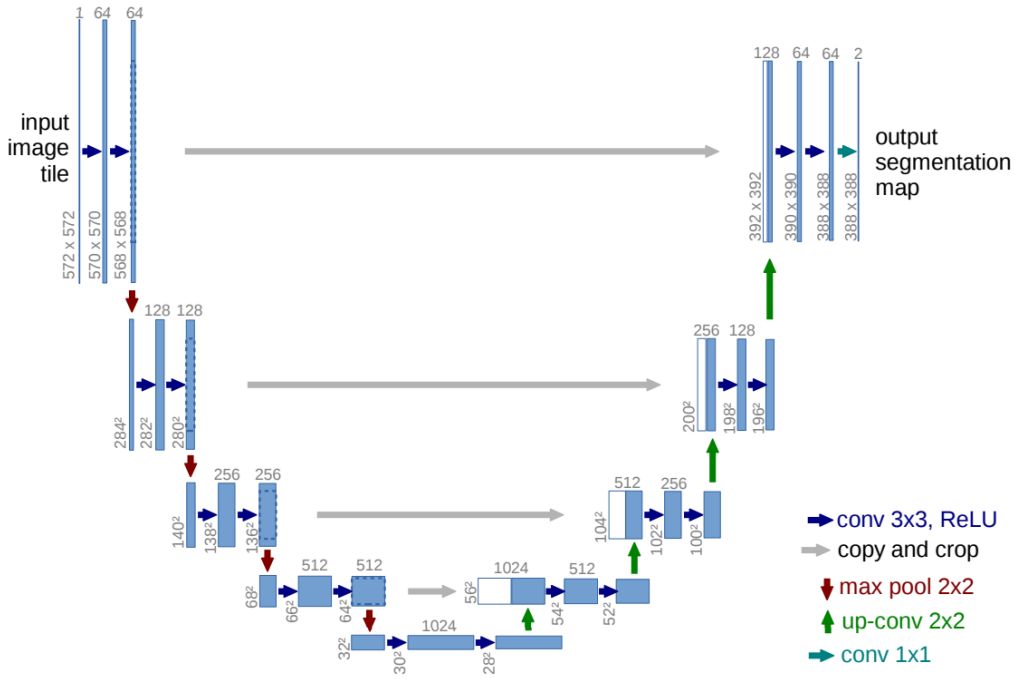

Figure 6.5. Each multi-channel feature map in the U-Net architecture is represented by a blue box with a label on top indicating the number of channels it contains. The box's lower left edge displays the x- and y-sizes. Replicated feature maps are shown by white boxes, and arrows show the operations performed between them [80].

The contracting path in U-Net is responsible for extracting the important features from the input image. In order to capture increasingly abstract representations of the input, the encoder layers use convolutional operations to gradually increase the feature maps' depth while decreasing their spatial resolution. On the contrary, while preserving the input's spatial resolution, the expanding path decodes the encoded data and locates the features. In addition to conducting convolutional operations, the decoder layers in the expanded path upsample the feature maps. Skip connections from the contraction path are employed to preserve the spatial information that would otherwise be lost during the downsampling process, allowing the decoder layers to localize features more precisely [80].

6.2.2. DeepLabV3+

DeepLabV3 is a deep learning model for image semantic segmentation. Chen et al. in 2018 proposed the newest version of DeepLabV3 called DeepLabv3+ [11]. It adds a simple yet effective decoder module to DeepLabV3 to help refine segmentation results, particularly along object boundaries. It controls the feature map and receptive field resolutions using Atrous (Dilated) Convolutions without adding more parameters overall. Atrous Spatial Pyramid Pooling is an additional key characteristic that efficiently obtains multiscale characteristics including valuable segmentation information [11].

DeepLabv3+ achieved remarkable results, 82.1% mIOU on the Cityscapes dataset and 89% mIOU on the PASCAL VOC 2012 test set. These accomplishments demonstrate the series' ongoing development in expanding the possibilities for semantic image segmentation. The framework of DeepLabV3+ architecture is shown in Figure 6.6.

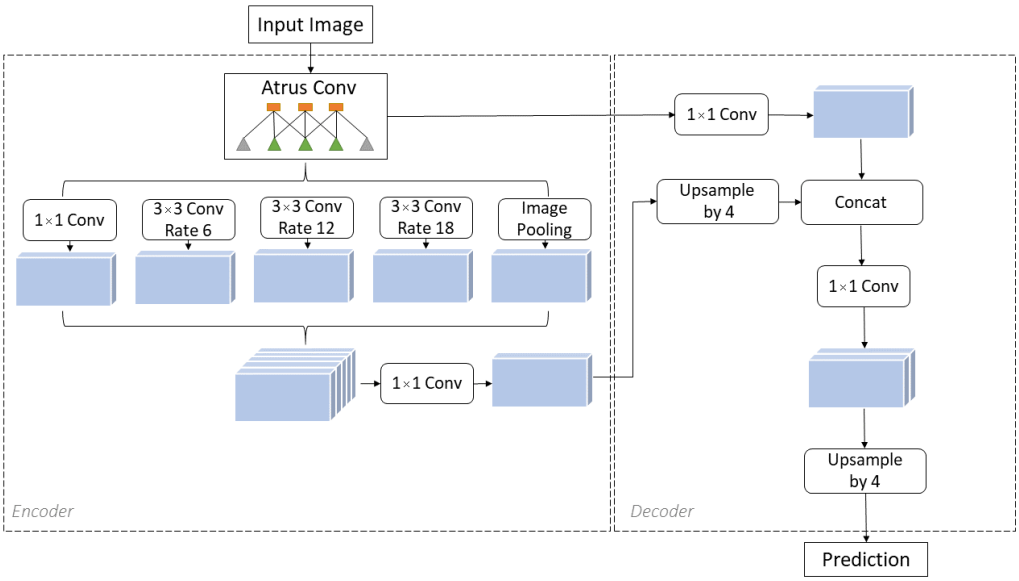

Figure 6.6. The architectures described in subsection 6.1. (Xception, ResNet101, MobileNetv2) can be used as DeepLabv3+ backbones.

6.2.3. SegFormer

SegFormer, a straightforward, effective, and reliable framework for semantic segmentation, was presented by Xie et al. in 2021. It combines lightweight multilayer perception (MLP) decoders with Transformers [111].

SegFormer has two main attributes:

a) Multiscale feature generation from a hierarchically structured Transformer encoder. By eliminating the need for positional encoding, it prevents positional codes from being interpolated, which could otherwise degrade performance in situations when the test resolution differs from the training resolution.

b) SegFormer avoids complicated decoders. In order to produce effective representations, the proposed MLP decoder combines local and global attention by aggregating data from many layers.

Segformers appear in six different configurations, ranging from B0 to B5. The lightest configuration is B0, while the best segmentation quality is achieved with B5 configuration. On the Cityscapes validation set, their top model, SegFormerB5, gets 84.0% mIoU and has exceptional zero-shot resilience on Cityscapes-C [111]. SegFormer framework is shown in Figure 6.7.

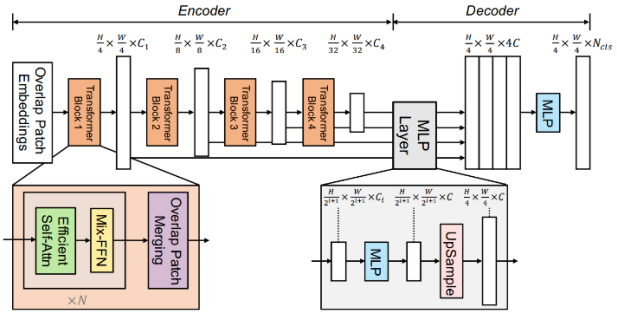

Figure 6.7. Two primary modules comprise the described SegFormer framework: lightweight all-MLP decoder that directly incorporates these multi-level characteristics to produce the semantic segmentation mask and a hierarchical Transformer encoder that records both coarse and fine-grained information [111].

7. Explainable Computer Vision for Interpretable Analysis of OSCC

In recent years, a consistent increasing trend has been observed in the application of AI-based models in the medical field, with numerous studies on automated diagnosis and prognosis. However, many AI models are still considered as a black box and not very interpretable. The issue of interpretability in the medical field significantly exceeds simple intellectual interest. More precisely, it is noted that interpretabilities in the medical domain include elements such as risk and responsibilities that are not considered in other fields. When medical decisions are made, human lives can be at risk. It would be equivalent to completely avoiding responsibility to entrust such major decisions to computers that are incapable of providing accountability. For that reason, in this chapter explainable AI is demonstrated in order to make AI systems more understandable to health professionals.

7.1. Explainability in Medical AI Systems

Explainable AI (XAI) is an emerging field that is extremely important in the medical field [84]. The development of AI is briefly related to data science, computer vision, natural language processing, machine learning, and statistical analysis. Despite these advancements, they were unable to surpass human intellect, which was further enhanced by deep learning, neural networks, and reinforcement learning. These developments were crucial for the improvement of the medical field. However, in order to comprehend specific decisions, outcomes, and the present state of the patient's problems, it is crucial that the medical field incorporate explanations regarding legal and ethical AI [41].

XAI aims to improve performance and explainability, which makes it easier for users to trust, comprehend, accept and manage AI systems.

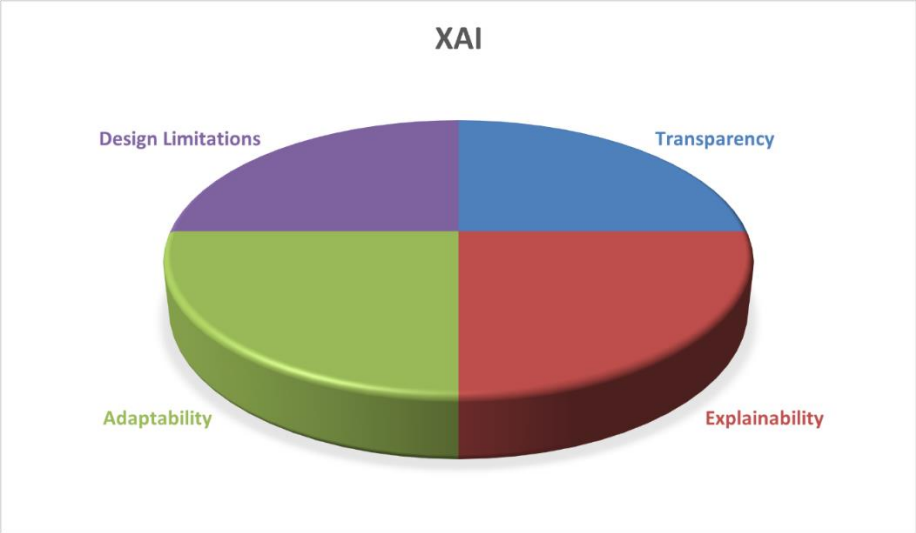

Figure 7.1. Key components of Explainable AI (XAI), such as transparency, explainability, adaptability, and limitations of design.

Figure 7.1. represents the benefits of XAI methods [84];

- **Design Limitations:** XAI facilitates a deeper understanding of data quality, feature distribution, classification and comparison evaluation by improving interpretability at every structural layer.
- **Transparency:** The transparent frameworks of XAI techniques are well-known for offering details on data processing and model creation. With an accurate understanding of the system's fundamental features, transparency allows users to effectively optimize the system.
- **Explainability:** When it comes to model design problems, explainability can assist in identifying the process step where the incorrect choice was taken, allowing for a later correction. For the initial data analysis, decision, and action for the entire XAI model, explainability is crucial.

- **Adaptability:** By utilizing the feedback technique, XAI models are renowned for their great degree of adaptability. The ability of XAI systems to adapt explanations and decision-support tools to various users

7.2. Global and Local Methods for the Preprocessing

Recent literature demonstrates the strategy of offering interpretability and transparency while utilizing the models as [32]:

- Gradient Weighted Class Activation Mapping (Grad-CAM),
- Layer-Wise Relevance Propagation (LRP),
- Statistical Functions for the Feature Analysis and Processing,
- SHapley Additive exPlanations (SHAP),
- Attention Maps and
- Local Interpretable Model-Agnostic Explanations (LIME)

Based on the literature review and aim of this research Grad-CAM will be utilized for visual representation.

7.2.1. Gradient Weighted Class Activation Mapping

As interpretability in deep learning has become more significant, especially in CNN architectures, Selvaraju et al. (2017) proposed Gradient-weighted Class Activation Mapping (Grad-CAM) as a visual explanation technique [82]. Grad-CAM leverages the gradient information of a target concept flowing backward into the last convolutional layer to generate coarse localization maps that emphasize the most discriminative portions of the input image that are most important in the model's prediction. This contribution enabled increased transparency in decision-making and represented a major advancement toward XAI.

To construct the class discriminative localization map $L^c_{Grad-CAM} \in \mathbb{R}^{uxv}$, the authors first calculate the gradient of the class score c, y^c with respect to feature maps A^k. Global average pooled gradients are used to determine the neuron significance weights, α^c_k:

$$\alpha^c_k = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A^k_{ij}} \tag{7.1}$$

This represents the significance of feature map k for a target class c and represents a partial linearization of the deep learning model downstream from A. By using the ReLU activation function, α^c_k gathers the corresponding class discriminative localization map.

$$L^c_{Grad-CAM} = ReLU\left(\sum_k \alpha^c_k A^k\right) \tag{7.2}$$

In general, a CNN that classifies images does not always need to produce y^c as its class score. It might be any differentiable activation, such as a question response or words from a caption. The Global Average Pooling method is used to spatially pool the K feature maps $A^k \in \mathbb{R}^{uxv}$. The pooled feature map and linear transformation are then used to obtain the class c score, S^c.

$$S^c = \sum_k w^c_k \frac{1}{Z}\sum_i \sum_j A^k_{ij} \tag{7.3}$$

It is possible to modify the previous equation by using L^c(CAM).

$$S^c = \frac{1}{Z}\sum_i \sum_j \sum_k w^c A^k_{ij} \tag{7.4}$$

8. Assessment of TSR in Histopathological Samples

In this chapter, the assessment of the tumor-stroma ratio (TSR) in histopathological specimens is described, along with its biological foundation, methodological techniques, and prognostic significance using Kaplan-Meier survival analysis.

8.1. Biological Foundation of TSR Interaction

For a prolonged period, clinicopathological factors, including tumor type, malignancy grade, tumor size, patient age and the existence of local or distant metastases, have determined the optimal plan of treatment for cancer [102]. However, the tumor microenvironment is becoming important feature of current biomarker development research. The tumor-stroma ratio (TSR) is one of the simplest yet effective histopathological metrics that represent the tumor metastatic environment (TME). The stroma interacts with both malignant and nonmalignant cells during all stages of carcinogenesis, from tumor onset to invasion and metastasis, making the tumor-stroma crucial to the growth and progression of cancer [36].

Not all tumors are formed out of cancerous epithelial cells. They instead coexist alongside a dynamic stroma consisting of extracellular matrix (ECM), fibroblasts, immunological cells, and endothelial cells.

Stroma actively contributes to the development of cancer by [99]:

- stimulating angiogenesis,
- modifying the extracellular matrix to make invasion easier,

- ❖ supplying cytokines and growth factors, and
- ❖ causing immunological evasion.

Increased communication between the microenvironment and malignant cells is shown in stroma-rich tumors, which are frequently associated with more aggressive biological behavior [106].

8.2. Method of Assessment

Based on the tissue slide used in normal diagnostic pathology to determine the tumor grade, the tumor-stroma ratio can be calculated. Using a 10× objective, one region within vision site that has both tumor and stromal tissues should be chosen. The chosen image area should show the tumor cells on all four sides. Groups with different stromal ratios were separated into stroma-high and stroma-low groups. According to the histological section, a tumor is classified as stroma-low if its stromal area is less than 50% and as stroma-high if it is more than 50% [88]. The TSR assessment methodology is shown in Figure 8.1.

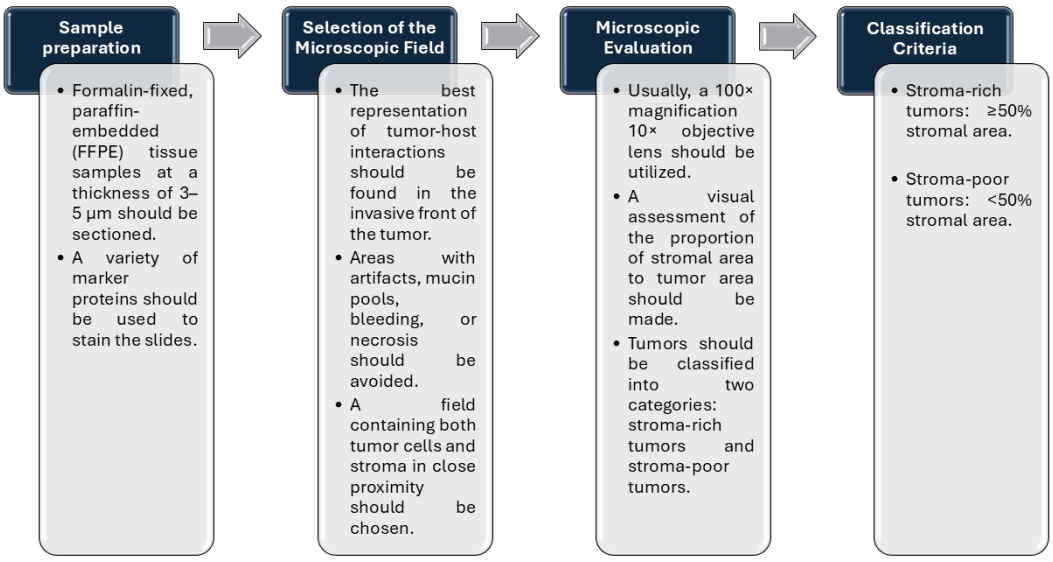

Figure 8.1. Framework for stromal assessment in histopathological samples that describes how to prepare samples, choose fields, examine them under a microscope, and classify tumors according to their stromal proportion.

8.3. Prognostic Significance

TSR has become a powerful independent prognostic indicator for several cancers:

- ❖ Colorectal carcinoma: Disease-free survival (DFS) and overall survival (OS) are negatively correlated with stroma-rich tumors.
- ❖ Breast cancer: Particularly in triple-negative breast cancer, a high stromal content is associated with an elevated risk of recurrence.
- ❖ Gastric and esophageal malignancies: TSR indicates a poor response to treatment and a lower survival rate.
- ❖ Non-small cell lung cancer (NSCLC): Stroma-rich tumors exhibit aggressive characteristics and a poor prognosis.

Table 8.1. Advantages and limitations of TSR assessment [113].

	Point	Description
Advantages of TSR Assessment	Cost-effective	Requires only routine stained slides, no additional tests.
	Reproducible	Standardized methodology allows high interobserver agreement.
	Clinically relevant	Provides prognostic information beyond conventional staging.
	Easily integrable	It can be incorporated into routine pathology workflow.
Challenges and Limitations	Subjectivity	TSR estimation relies on visual assessment, which may lead to inter-observer variability.
	Tumor heterogeneity	Different areas of the tumor may show variable stromal content.
	Cutoff discrepancies	Lack of universal agreement on the cutoff threshold.
	Limited validation in rare cancers	Most evidence is restricted to common carcinomas.

According to the limitations presented in Table 8.1, this research will utilize artificial intelligence (AI)-based image analysis and digital pathology to standardize TSR evaluation and reduce observer bias.

8.4. Kaplan-Meier survival analysis

The significance of tumor-stroma ratios as a prognostic marker will be assessed in this research using the Kaplan–Meier (KM) survival curve. A subfield of statistics called survival analysis examines time-to-event data, where the outcome of interest is the interval between an event—such as death, return of an illness, or failure of a treatment—and the time until it happens. Introduced by Edward L. Kaplan and Paul Meier in 1958, the Kaplan–Meier survival curve is one of the most used methods in this field. Researchers and clinicians can use this method to estimate survival probabilities across time, even if some data are censored (i.e., the event of interest has not occurred for some individuals during the study period [79].

The Kaplan-Meier curve is a step function that decreases when events occur. A death or relapse, for example, is represented by each step that goes down. The process of estimation includes the following steps [79]:

I. Determine unique event times by arranging survival times in ascending order.

II. Determine the survival rate for each incident:

$$\boldsymbol{S(t)} = \prod_{\boldsymbol{i}:\boldsymbol{t_i} \leq \boldsymbol{t}} \left(\mathbf{1} - \frac{\boldsymbol{d_i}}{\boldsymbol{n_i}}\right) \tag{8.1}$$

where $\boldsymbol{t_i}$ is the time of the $\boldsymbol{i^{th}}$ event, $\boldsymbol{d_i}$ is the number of the events at $\boldsymbol{t_i}$ and $\boldsymbol{n_i}$ is the number of individuals at risk just before $\boldsymbol{t_i}$.

III. Plot the curve: A stepwise function that starts at 1.0 (100 percent survival at time zero) and gets smaller with every occurrence. Usually, tick marks along the curve denote censored data points.

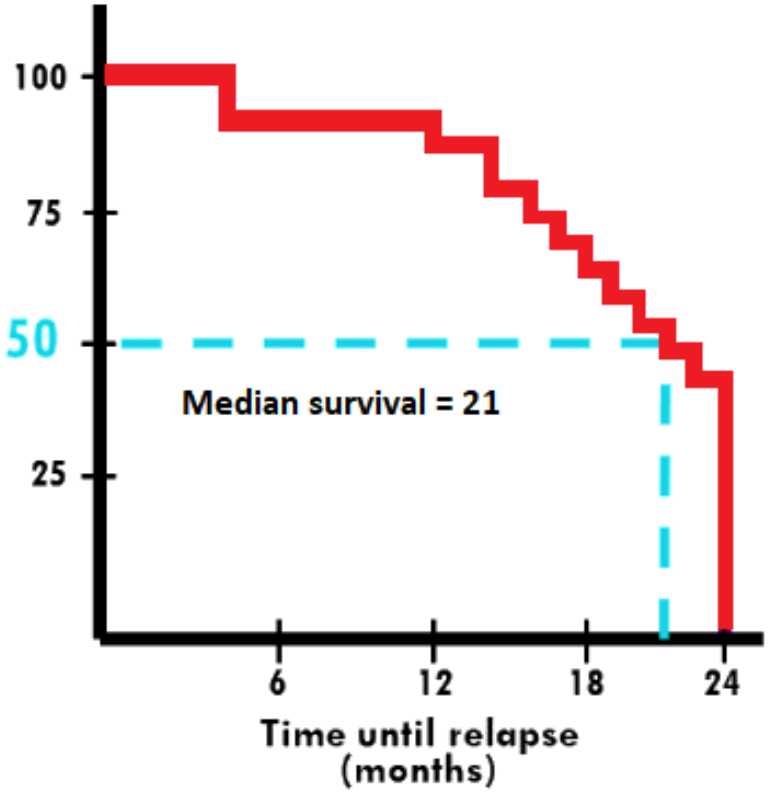

Figure 8.2. A schematic representation of a Kaplan-Meier survival curve which demonstrates the point at which median survival is established as well as the decline in patient survival with time.

The KM survival curve is schematically represented in Figure 8.4.1, with the vertical axis representing the estimated probability of survival. Time is shown on the horizontal axis in months, years, or any other applicable unit. Additionally, the point at which the survival probability drops to 50 is known as the median survival time.

By using Kaplan–Meier analysis, it is possible to demonstrate whether patients with a high stromal component (stroma-high TSR) indeed have lower overall or disease-free survival rates compared to those with a low stromal component.

9. Evaluation Criteria

The ability of a deep learning model to generalize new data is a fundamental component when evaluating its performance [64]. Furthermore, validation methods are crucial for detecting and preventing overfitting of the model, which ensures accurate results on unseen data. The following metrics are commonly used to assess classification and segmentation models.

The accuracy measure (ACC) points out what percentage of the pixels in the image are assigned to the correct class and can be defined as follows [33]:

$$ACC = \frac{TP + TN}{TN + TP + FN + FP}. \tag{9.1}$$

Cases where both the actual and predicted results are positive are referred to as true positives (TP). When the actual and predicted outcomes are both negative, this is referred to as a true negative (TN). When a positive actual outcome is mistakenly assigned a negative predicted by the model, this is known as a false negative (FN). On the other hand, when the model predicts a positive result when the actual result is negative, this is known as a false positive (FP).

Precision, shows the percentage of the results which are relevant and can be defined as [12]:

$$Precision = \frac{TP}{TP + FP}. \tag{9.2}$$

Sensitivity, sometimes referred to as Recall or the True Positive Rate, quantifies the percentage of data points with positive labels that the model correctly classifies and can be calculated as [61]:

$$Sensitivity = \frac{TP}{TP + FN}. \tag{9.3}$$

Specificity, also known as the True Negative Rate, quantifies the percentage of data points with a negative label that the model correctly classifies and can be mathematically expressed as[61]:

$$Specificity = \frac{TN}{TN + FP}. \tag{9.4}$$

Accuracy, Precision, Sensitivity and Specificity can be used as evaluation criteria for both classification and segmentation models. However, multiclass classification requires evaluation criteria considering numerous categories, unlike binary classification, which only has two categories.

Model classification ability can be assessed using statistical metrics like Micro- and Macro-Area Under the Curve (AUC). The AUC is an evaluation metric used to determine the binary classifier's performance. In order to use AUC for multiclass classification, the problem needs to be considered as binary classification problem using the One vs. All technique, in which one class is categorized against every other class. The ratio of correctly identified cases across all classes to the total number of samples is known as the micro-averaged true positive rate, or TPR. The percentage of cases that are incorrectly classified across all classes in relation to the total number of samples is also known as the false positive rate (FPR), or fallout. The mathematical representation of Micro averaging is defined as follows [98]:

$$TPR_{micro} = \frac{\sum_{i=1}^{k} TP_i}{\sum_{i=1}^{k} TP_i + \sum_{i=1}^{k} FN_i} \tag{9.5}$$

and

$$FPR_{micro} = \frac{\sum_{i=1}^{k} FP_i}{\sum_{i=1}^{k} FP_i + \sum_{i=1}^{k} TN_i}, \tag{9.6}$$

by which AUC_{micro} can be calculated. In Macro averaging for k classes, the metrics are calculated separately for each class, and the results are averaged together. Based on the computation of both TPR_{macro} and FPR_{macro}, AUC_{macro} can be computed as follows [47]:

$$TPR_{macro} = \frac{\sum_{i=1}^{k} TPR_i}{k} \tag{9.7}$$

and

$$FPR_{macro} = \frac{\sum_{i=1}^{k} FPR_i}{k}. \tag{9.8}$$

The Jaccard Index, sometimes referred as Intersection-Over-Union (IOU), is one of the most popular metrics for semantic segmentation, and it can be defined as [77]:

$$IOU = \frac{TP}{TP + FP + FN} \tag{9.9}$$

The mIOU has a positive correlation with the Dice coefficient (F1). It is an overall measure of a model's accuracy and can be calculated as follows [12]:

$$F1 = \frac{2TP}{2TP + FP + FN}. \tag{9.10}$$

10. Results and Discussion

This chapter summarizes the main outcomes of this doctoral thesis, which include multiclass classification results, GRAD-CAM visualization for model interpretability, semantic segmentation performance, automatic TSR quantification and experimental proof of concept.

10.1. Multiclass classification

A thorough deep learning pipeline designed for OSCC multiclass classification is demonstrated by the framework in Figure 10.1. Image acquisition is the first step in the pipeline, which is followed by preprocessing and data augmentation. Preprocessing method based on SWT is developed in order to increase classification performance by enhancing high-frequency components. Augmentation techniques such as geometric transformations are used to artificially increase the quantity of training samples. The processed images are then forwarded into pre-trained deep CNN architectures. Each model performs multiclass classification in order to assign histopathological images to one of three classes: Grade I, Grade II, or Grade III. Due to the high imbalance among OSCC classes the performance of AI-based models is estimated utilizing stratified 5-fold cross-validation. In the last step AUC_{micro} and $-_{macro}$ metrics are used to evaluate model performance.

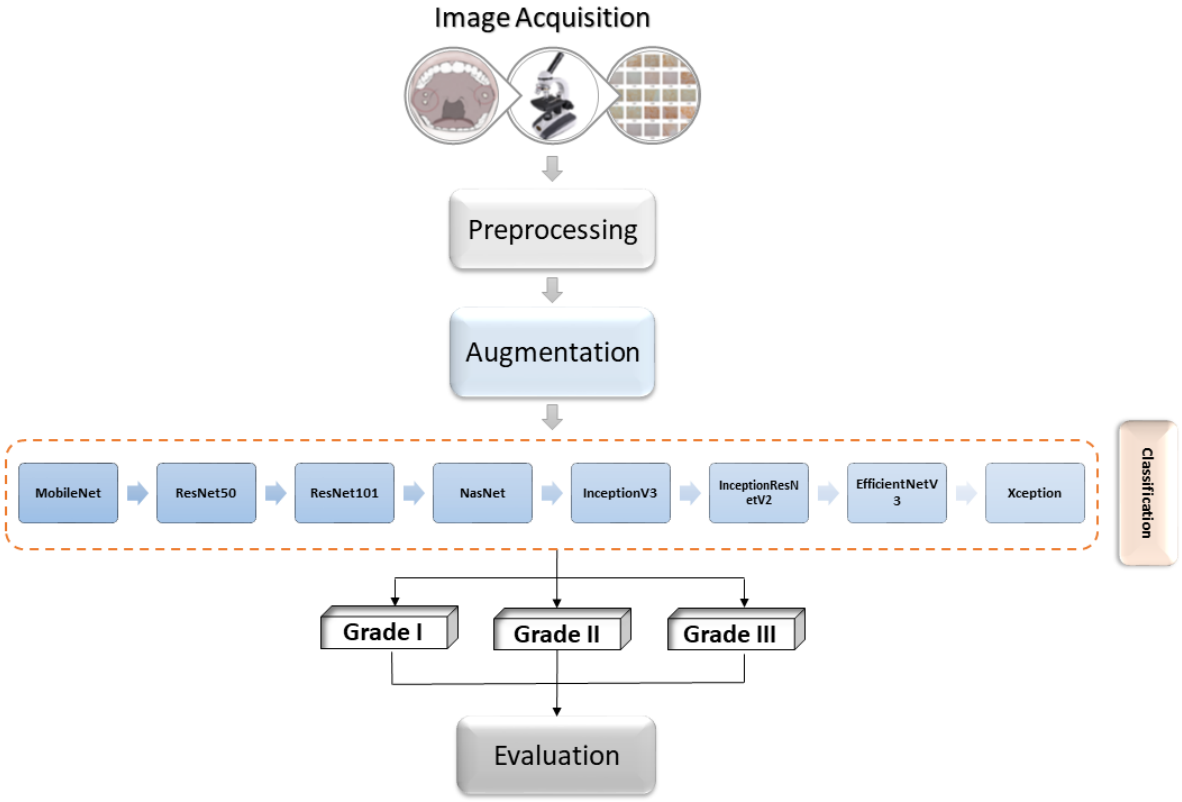

Figure 10.1. Framework for multiclass grading approach.

Initial experimental results are obtained on ImageNet using pretrained MobileNetV2, ResNet50, ResNet101, NASNet, InceptionV3, InceptionResNetV2, EfficientNetB3, and Xception architectures. In order to perform multiclass classification of OSCC, the current research adds two additional layers to the widely used deep CNN architectures. The first layer is the global average pooling layer, which reduces the h×w×c (height, width, channels) tensor to a 1×1×c, which also forces the network to focus on global spatial information Furthermore, the fully connected layer is the second added layer, consisting of three neurons and a Softmax activation function. For training each model architecture, three optimizers are used: Adam, RMSprop, and Stochastic Gradient Descent (SGD).

Additionally, every AI model architecture is trained in two steps:

- the first step involves only the output layer being trainable while the others are frozen, and
- the second step involves the output layer being frozen while the other layers are trainable.

This method provides steady training and gradual adaptation.

The results presented in Figures 10.2. – 10.9. are achieved by utilizing early stopping and modifying optimizer hyperparameters such as learning rate and learning rate decay. In order to offer a robust and unbiased evaluation of model performance, stratified 5-fold cross-validation was used.

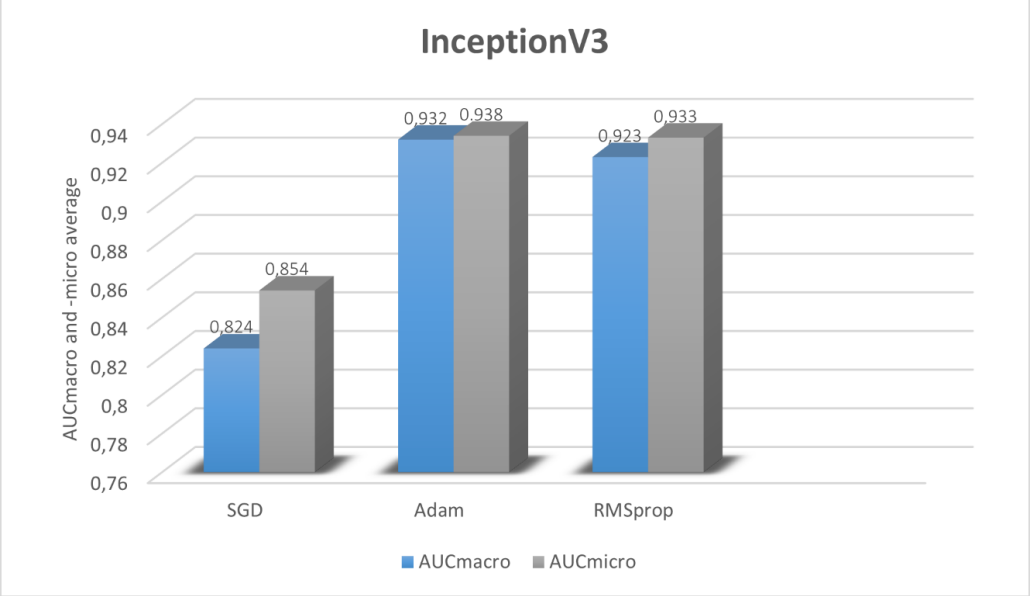

Figure 10.2. InceptionV3; The AUC_{macro} and AUC_{micro} measures are used in the bar graph to compare the performance of three optimization algorithms: SGD, Adam, and RMSprop.

Figure 10.2. compares the performance of InceptionV3 with different optimization algorithms. Based on the SGD optimization algorithm, InceptionV3 achieved an AUC_{macro} of 0.824 and an AUC_{micro} of 0.854. With an AUC_{macro} of 0.932 and an AUC_{micro} of 0.938, Adam, however, achieved superior results. RMSprop also achieved strong performance, with an AUC_{macro} of 0.923 and an AUC_{micro} of 0.933.

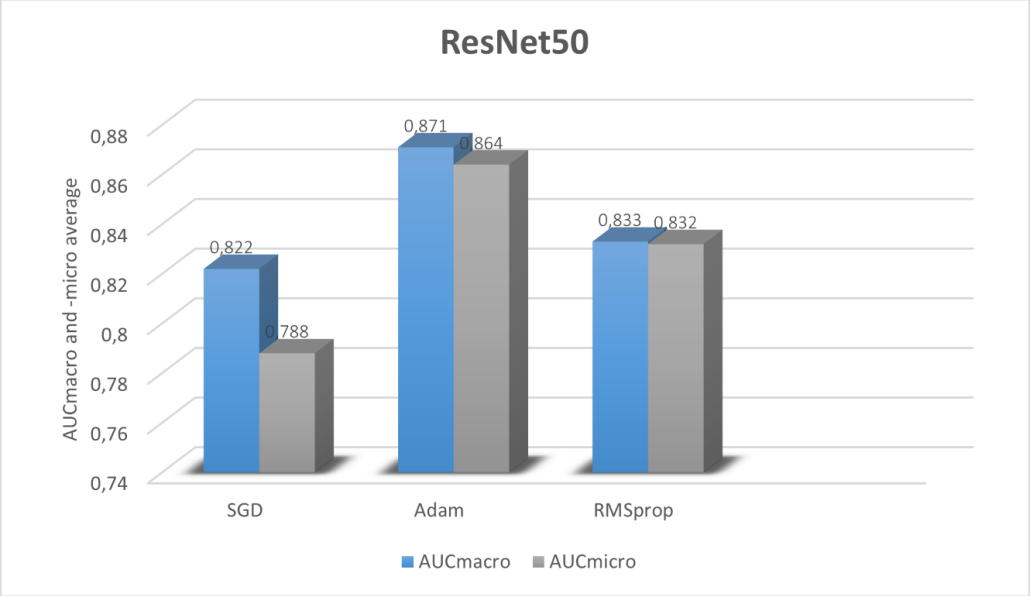

Figure 10.3. ResNet50; The AUC_{macro} and AUC_{micro} measures are used in the bar graph to compare the performance of three optimization algorithms: SGD, Adam, and RMSprop.

ResNet50 results are shown in Figure 10.3. With an AUC_{macro} of 0.871 and an AUC_{micro} of 0.864, the Adam optimizer achieved the highest results. On the other hand, RMSprop produced comparatively lower results, showing consistent performance but not exceeding Adam, with AUC_{macro} of 0.833 and AUC_{micro} of 0.832. With a relatively low AUC_{micro} of 0.788 and a lowest AUC_{macro} of 0.822, SGD showed limited efficiency.

Figure 10.4 shows the performance of ResNet101 when trained with various optimizers, which is comparable to the outcomes of ResNet50. Obtaining the highest values, an AUC_{macro} of 0.882 and AUC_{micro} of 0.890, Adam surpasses the other optimizers. With an AUC_{macro} of 0.860 and an AUC_{micro} of 0.834, SGD demonstrates comparatively strong results. RMSprop performs moderately but less reliably than Adam and SGD, with the lowest macro score (0.829) and a slightly higher micro score (0.836).

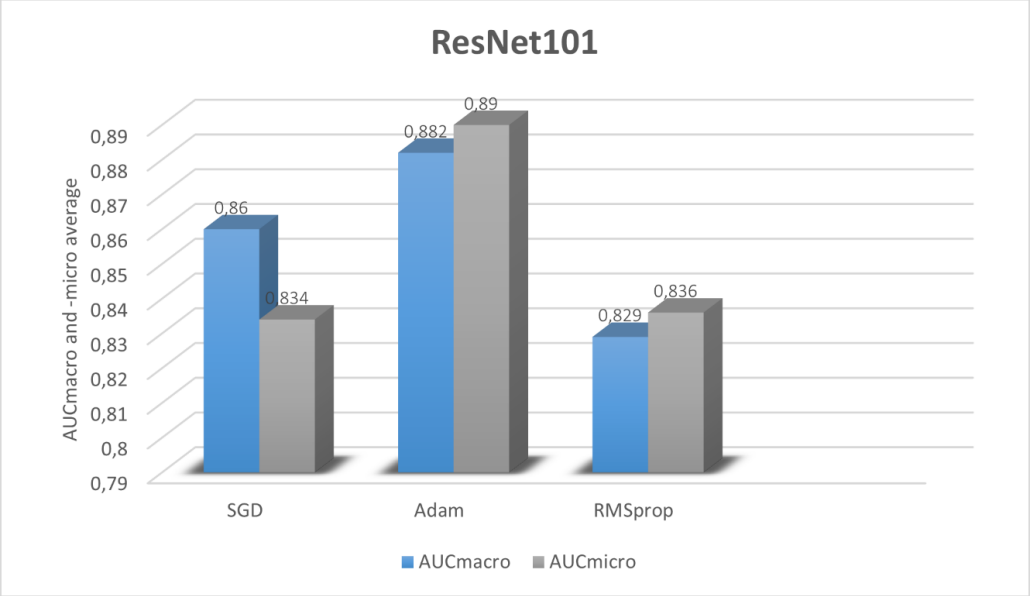

Figure 10.4. ResNet101; The AUC_{macro} and AUC_{micro} measures are used in the bar graph to compare the performance of three optimization algorithms: SGD, Adam, and RMSprop.

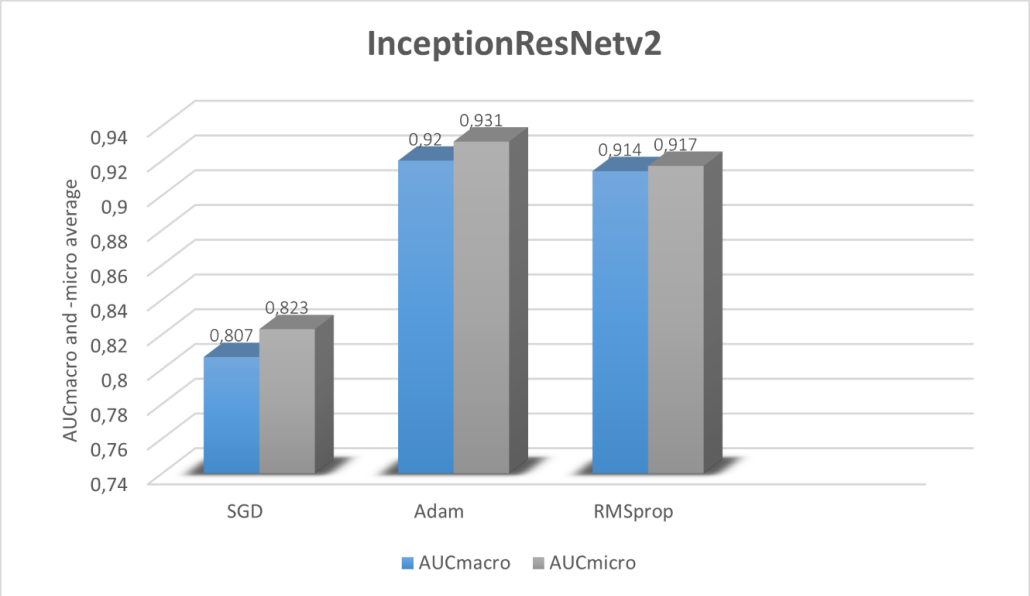

Figure 10.5. InceptionResNetv2; The AUC_{macro} and AUC_{micro} measures are used in the bar graph to compare the performance of three optimization algorithms: SGD, Adam, and RMSprop.

InceptionResNetV2 results are shown in Figure 10.5. The best performance is demonstrated by the Adam optimizer, which has the highest AUC_{macro} (0.920) and AUC_{micro} (0.931). With closely aligned scores (AUC_{macro} 0.914, AUC_{micro} 0.917), RMSprop also performs well.

On the other hand, SGD performs noticeably worse with AUC_{macro} of 0.807 and AUC_{micro} of 0.823.

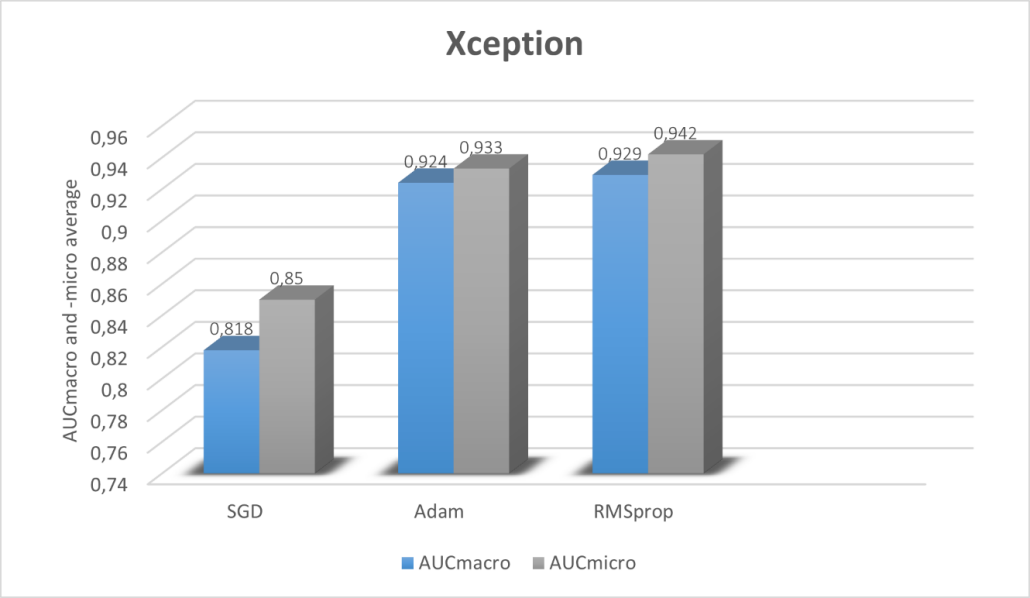

Figure 10.6. Xception; The AUC_{macro} and AUC_{micro} measures are used in the bar graph to compare the performance of three optimization algorithms: SGD, Adam, and RMSprop.

The lowest values (AUC_{macro} = 0.818, AUC_{micro} = 0.850) of the three optimizers in Xception architecture were obtained by SGD as seen in Figure 10.6. In contrast, Adam's performance showed a significant improvement, achieving an AUC_{micro} of 0.933 and an AUC_{macro} of 0.924. The superior results were obtained by RMSprop, which had an AUC_{macro} of 0.929 and an AUC_{micro} of.992.

The comparison outcomes of the three optimizers for the MobileNet architecture are shown in Figure 10.7 With an AUC_{micro} of 0.901 and an AUC_{macro} of 0.877, SGD performed the best. Adam achieved satisfactory performance of AUC_{macro} (0.762), however AUC_{micro} (0.613) performance significantly declined. With an AUC_{micro} of 0.592 and an AUC_{macro} of 0.745, RMSprop achieved the lowest performance.

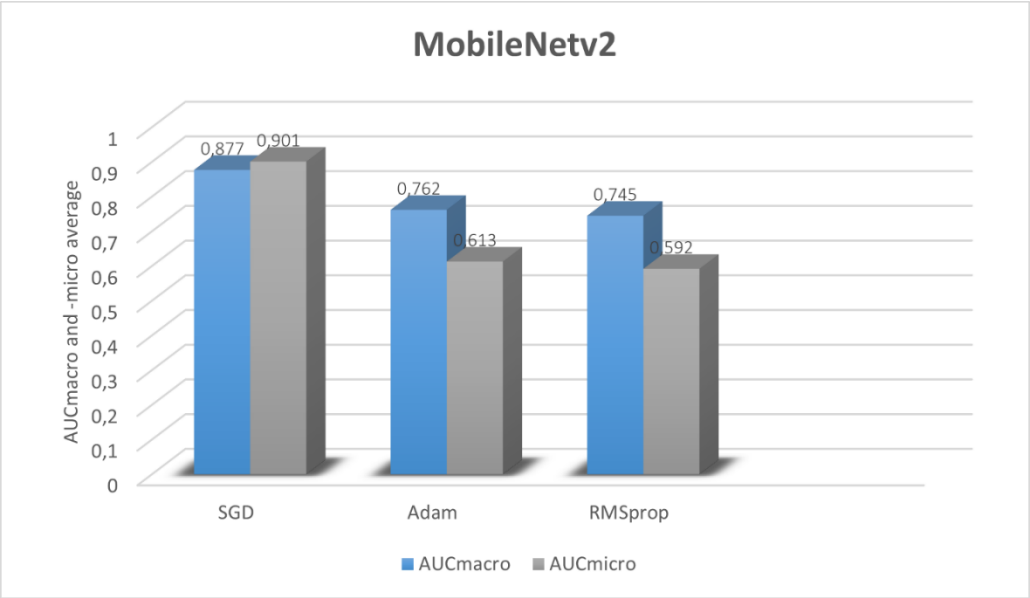

Figure 10.7. MobileNetv2; The AUC_{macro} and AUC_{micro} measures are used in the bar graph to compare the performance of three optimization algorithms: SGD, Adam, and RMSprop.

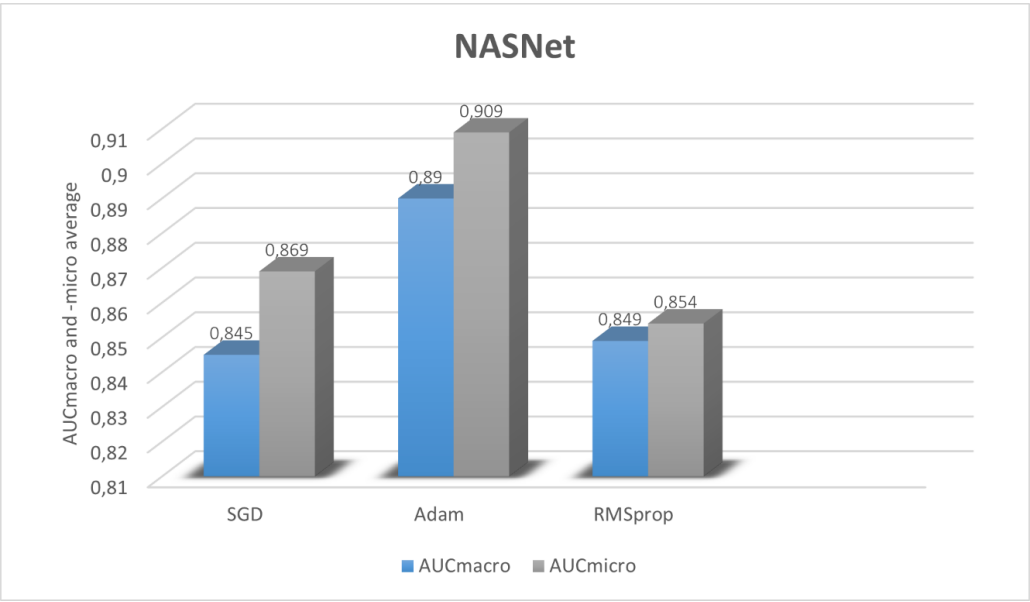

Figure 10.8. NASNet; The AUC_{macro} and AUC_{micro} measures are used in the bar graph to compare the performance of three optimization algorithms: SGD, Adam, and RMSprop.

In Figure 10.8., the NASNet results are shown. Among the optimizers, Adam achieved superior performance with an AUC_{macro} of 0.890 and an AUC_{micro} of 0.909. With an AUC_{macro} of 0.845 and an AUC_{micro} of 0.869, SGD showed solid performance, whereas RMSprop showed a limited improvement, better in terms of macro-average (0.849) but worse in terms of micro-average (0.854).

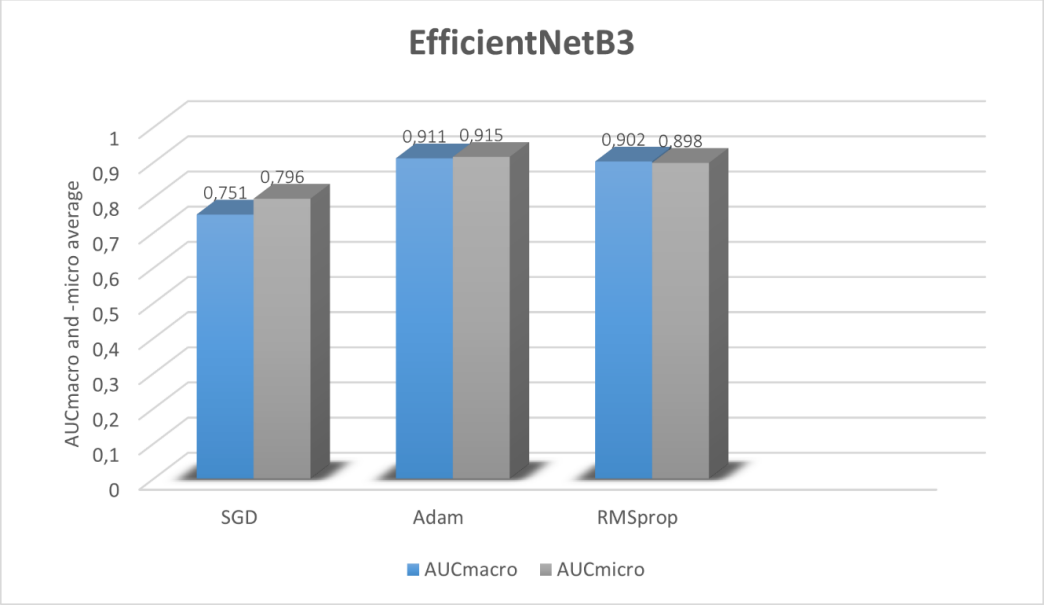

Figure 10.9. EfficientNetB3; The AUC_{macro} and AUC_{micro} measures are used in the bar graph to compare the performance of three optimization algorithms: SGD, Adam, and RMSprop.

The performance of the EfficientNetB3 model trained with SGD, Adam, and RMSprop optimizers is shown in Figure 10.9. With an AUC_{macro} of 0.751 and an AUC_{micro} of 0.796, SGD produced the lowest results. Adam, on the other hand, achieved the highest results with an AUC_{macro} of 0.911 and an AUC_{micro} of 0.915. With AUC_{macro} and AUC_{micro} values of 0.902 and 0.898, respectively, RMSprop also demonstrated strong performance.

A review of the research results revealed that the best overall performance metrics were obtained with a two-step training approach. In the first phase, only the output layer was tuned, using a learning rate decay of 1×10^{-6} and a learning rate of 1×10^{-3}. The next step involved freezing the output layer and continuing training for the remaining network layers with the same decay value of 1×10^{-6} and a reduced learning rate of 1×10^{-4}.

Stratified 5-fold cross-validation outcomes showed that the Adam optimizer consistently performed better than most model architectures, as seen by both AUC $_{micro}$ and $_{macro}$ values.

Adam performed the best within the ResNet50 architecture, achieving 0.871 ± 0.105 for AUC_{macro} and 0.864 ± 0.090 for AUC_{micro}, whereas SGD produced the lowest results. Adam produced better results with AUC_{macro} of 0.882 ± 0.125 and AUC_{micro} of 0.890 ± 0.11 for ResNet101, following a similar pattern. Additionally, the best overall performance was shown by the NASNet architecture trained using Adam, which achieved 0.890 ± 0.054 for AUC_{macro} and 0.909 ± 0.043 for AUC_{micro}. RMSprop, on the other hand, consistently achieved the lowest results, indicating that its optimization approach was less appropriate for the dataset and the feature representations that these architectures were able to extract.

According to the performance evaluation, the Xception architecture with RMSprop optimizer yielded the best overall results, with AUC_{macro} of 0.929 ± 0.087 and AUC_{micro} of 0.942 ± 0.074. In a comparable manner, SGD produced lowest results for InceptionV3 architecture, whereas the Adam optimizer produced the best results for this architecture (AUC_{macro} of 0.932 ± 0.081 and AUC_{micro} of 0.938 ± 0.088). With SGD, MobileNetV2 had the best performance, with AUC_{macro} of 0.877 ± 0.062 and AUC_{micro} of 0.901 ± 0.049. The Adam optimizer repeatedly produced the best results for InceptionResNetV2 (AUC_{macro} of 0.920 ± 0.059 and AUC_{micro} of 0.931 ± 0.064), whereas SGD produced the lowest results. AUC_{macro} of 0.911 ± 0.148 and AUC_{micro} of 0.915 ± 0.148 were achieved using EfficientNetB3 and Adam; nevertheless, this configuration showed more variability among folds, by showing a slight increase in standard deviations.

It can be observed from the model architecture and optimizer performances that the Adam optimizer generally provides superior classification performance on the data used in this research. SGD, on the other hand, showed limited performance across all assessed models, whereas RMSprop produced inconsistent results, demonstrating competitive performance only with the Xception architecture.

The SWT is used in the second step of the proposed approach to preprocess the data. The original histopathological images were decomposed at level 1 using the Haar, sym2, db2, and bior1.3 wavelet functions. After the decomposition process, high-frequency wavelet coefficients LH, HL, HH are weighted using a mapping function, which resulted in new, modified LH', HL', HH' subbands. An input image for the AI model was obtained utilizing SWT reconstruction using modified subbands alongside the unmodified LL subband, as seen in Figure 10.10.

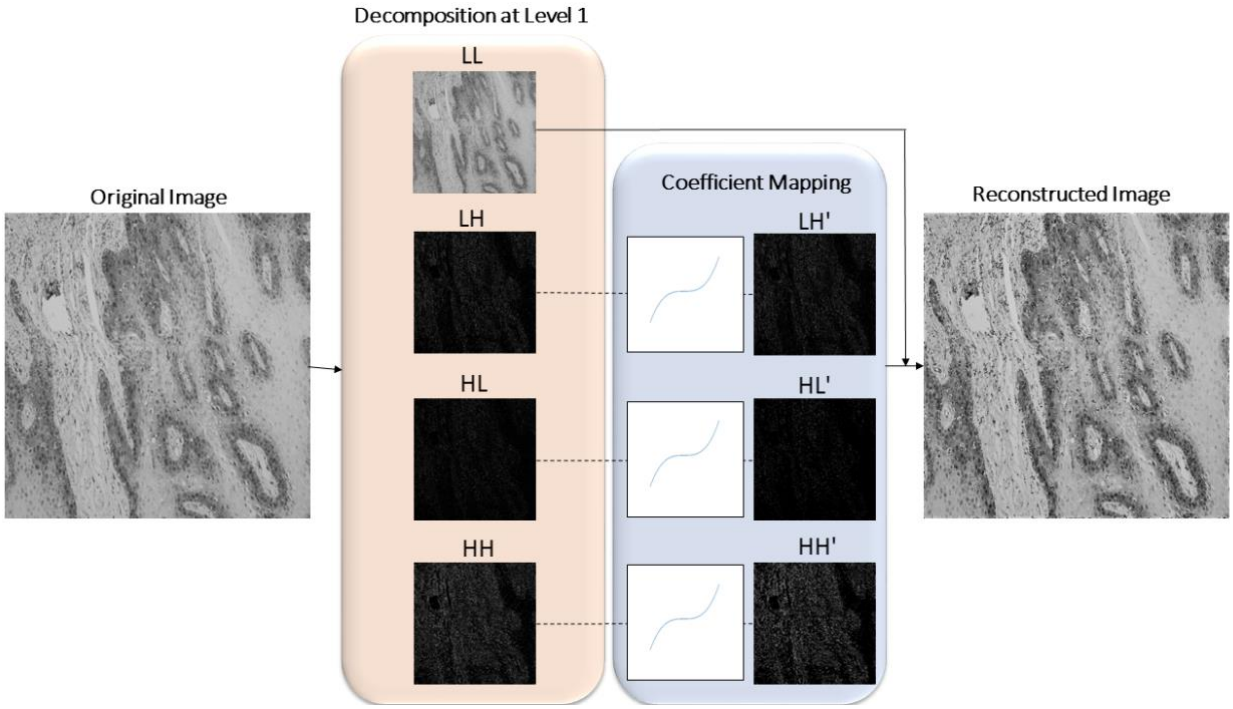

Figure 10.10. Level 1 SWT decomposition employing the Haar wavelet, coefficient mapping, and SWT reconstruction.

The main goal of Bayesian optimization was to determine the ideal wavelet mapping function constant values in order to maximize the performance measure. In this research, the AUC_{micro} performance metric was monitored throughout the optimization process. 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 shown in Table 10.1.

Table 10.1. Estimated constants for the coefficient mapping function obtained through Bayesian optimization along with corresponding 5-fold cross-validation performance.

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

10.2. Grad-CAM visualization

The following phase of the research used Gradient-weighted Class Activation Mapping to identify the areas of the image that showed the strongest impact on the model's predictions. In order to support diagnostic reasoning and boost confidence in automated systems, these visual explanations enhance interpretability by highlighting the regions of histopathology slides that are most suggestive of classifications. For the proposed model, the Grad-CAM visualizations are shown in Figures 10.11., 10.12., and 10.13. Figures show examples of histopathological images along with the corresponding Grad-CAM visualizations. In the left column, the original images are shown while the Grad-CAM heatmaps placed on the original tissue images are displayed in the left column. The heatmaps show the tissue regions that had strongest impact on the model's classification decision. The significance of various locations is indicated by the color spectrum, which ranges from blue to red. Blue shows places with the lowest activation, while red suggests areas with the highest activation.

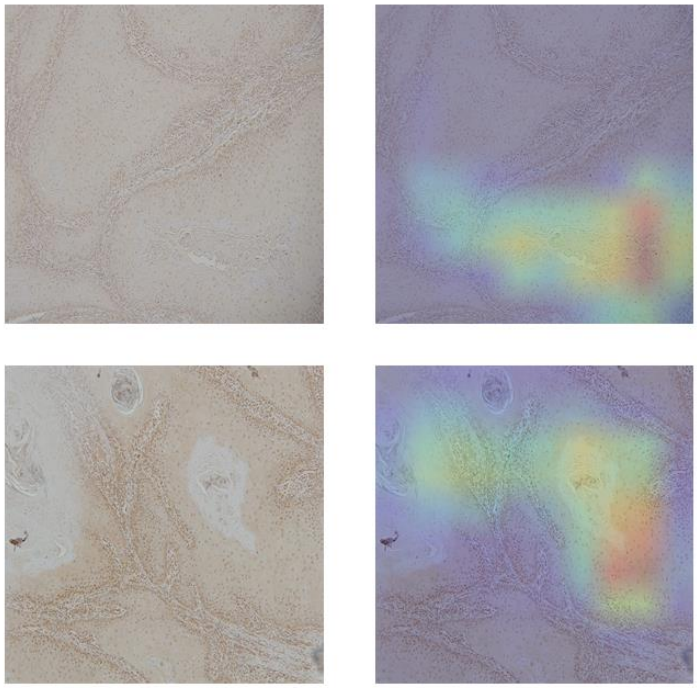

Figure 10.11. Grad-CAM application on histopathology images in order to highlight the Grade I discriminative regions

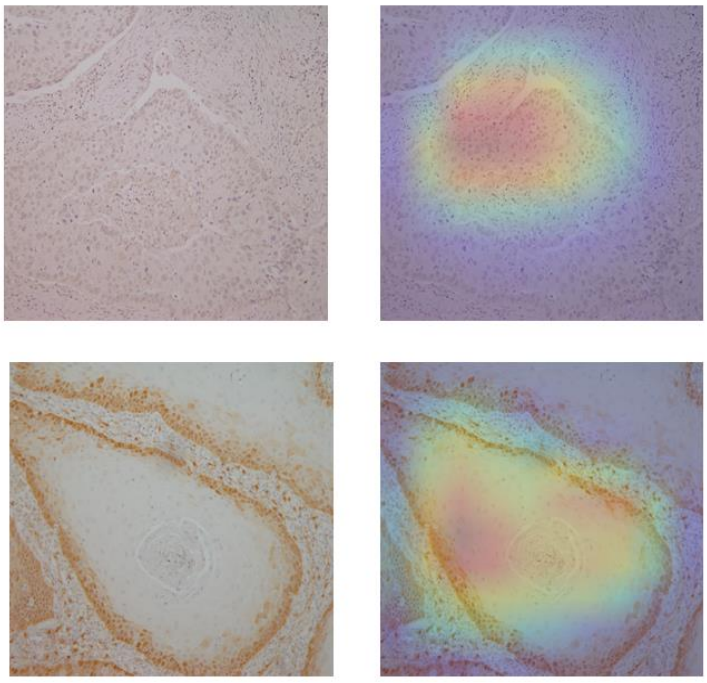

Figure 10.12. Grad-CAM application on histopathology images in order to highlight the Grade II discriminative regions

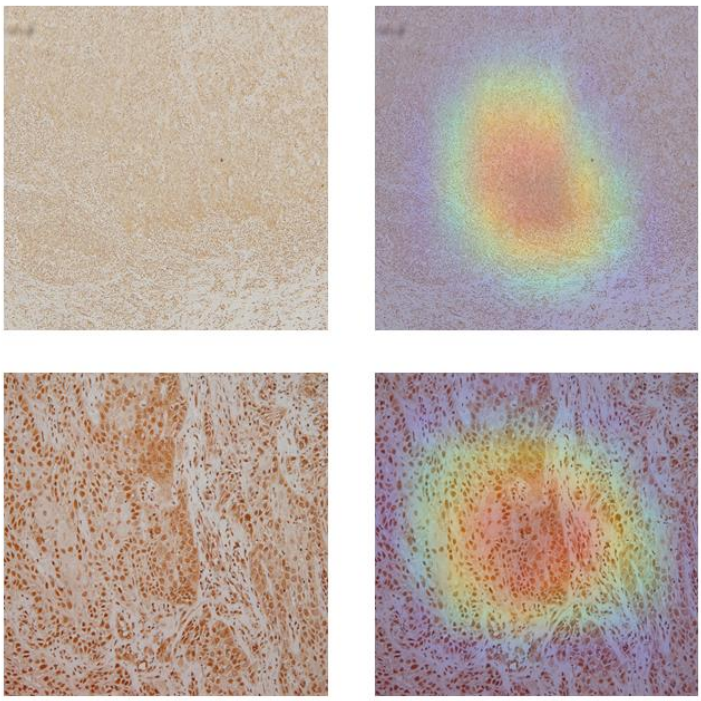

Figure 10.13. Grad-CAM application on histopathology images in order to highlight the Grade III discriminative regions

Grad-CAM was used to create heatmaps that show the most discriminative areas of histopathological images in the context of multiclass classification (Figures 10.11, 10.12, and 10.13). It captures gradients related to specific output classes, such as Grade I, Grade II, and Grade III, that flow into the final convolutional layers. To create a localization map, these gradients are pooled channel-wise, emphasizing important regions for class prediction. To create a heatmap, an input image is forward propagated through the network, computing gradients in relation to feature maps, spatially pooling these gradients, and combining weights with activation maps. By visualizing the model's decision-making process, this procedure verifies that the network focuses on pathologically significant regions rather than unimportant ones or artifacts.

However, some of the drawbacks limit the use of this method for comprehending deep learning models. It provides information about important aspects of the image but primarily focuses on high-level characteristics from later model layers, excluding details on how mid- or early-level features influence choices.

Spatial localization is limited by the output feature maps' resolution, which produces coarse heatmaps that may not accurately detect micro or subtle image features. Such features are crucial in clinical contexts.

The advantage of this method, however, is that the Grad-CAM can be applied to other areas outside the diagnosis of oral cancer from histopathology images. It can be used to create heatmaps that highlight significant regions that influence model predictions using a variety of medical imaging modalities, such as ultrasounds, CT scans, MRIs, and X-rays. Additionally, this interpretability can enhance trust in AI models for tasks such as detecting cardiovascular irregularities, brain tumors, lung tumors, or breast tumors.

10.3. Semantic segmentation

After multiclass grading of oral squamous cell carcinoma from histopathological images, the next step is semantic segmentation of tumor on the epithelial vs. stromal tissue.

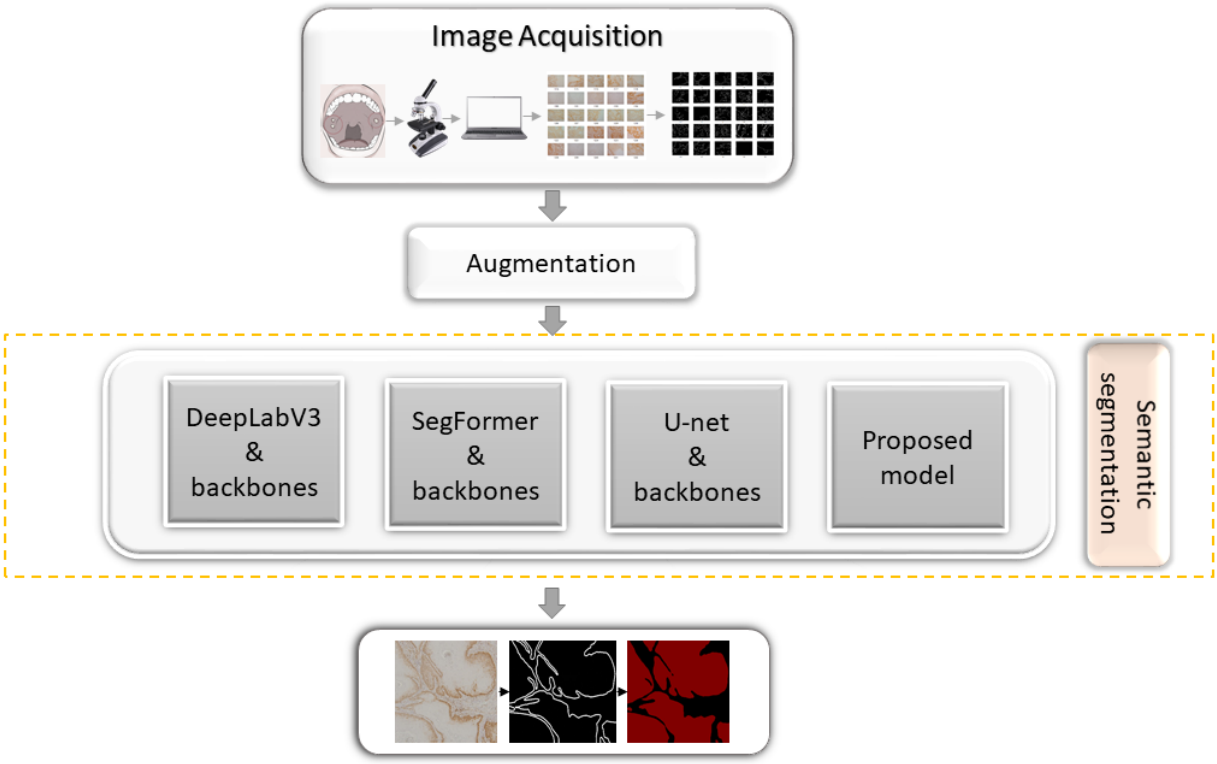

Figure 10.14. Framework for semantic segmentation approach.

The procedure for semantic segmentation designed for oral histopathology image analysis is shown in Figure 10.14. In the same manner as described for multiclass classification, image acquisition is the first step in the process, which follows data augmentation. Augmentation techniques such as geometric transformations are used to artificially increase the quantity of training samples. The images are then forwarded into several segmentation models, such as DeepLabV3, SegFormer, U-Net (each with its individual backbones) and proposed model. The proposed model consists of preprocessing method and transformer-based model. Preprocessing method based on Luminance Wavelet Enhancement is developed in order to improve the structural representation of immunohistochemistry images before they are transmitted to the segmentation model. The model results are shown by comparing the original histopathological image with its ground truth annotation and the predicted segmentation mask, as shown at the bottom of the figure. This shows how well the models detect relevant tissue features.

U-Net, DeepLabV3, and SegFormer were selected for the purpose of this research's assessment of semantic segmentation performance since they highlight substantial differences between generations and design approaches within segmentation architectures.

Unet:

- because of its encoder–decoder structure and skip connections, which maintain the fine spatial features necessary for tissue border recognition, U-Net has long been regarded as a benchmark model in biomedical image processing.

DeepLabV3+:

- is very good at capturing complicated structural variations that are frequently seen in histopathology images as it introduces enhanced atrous (dilated) convolutions and multi-scale context aggregation through ASPP modules.

SegFormer:

- offers strong global feature extraction and effective computation without depending on bulky decoders.

The performance evaluation of the DeepLabv3+ model with Xception_65 backbone utilizing a variety of segmentation metrics is shown in Figure 10.15. The line plot (green) shows the standard deviation (std) throughout experimental runs, while the bar chart (blue) shows the mean values of significant metrics.

The model demonstrates strong segmentation capabilities with balanced true positive and true negative detection, with 0.9466 ± 0.0049 accuracy, 0.9587 ± 0.0036 Dice coefficient, 0.9572 ± 0.0071 sensitivity, and 0.9602 ± 0.0048 precision. The lowest of the metrics is mIOU of 0.8898 ± 0.011. Specificity of 0.9275 ± 0.0039 is marginally lower, indicating a higher rate of false positives. Except for mIOU, which shows higher variability, the standard deviation across metrics is comparatively low, suggesting stable performance.

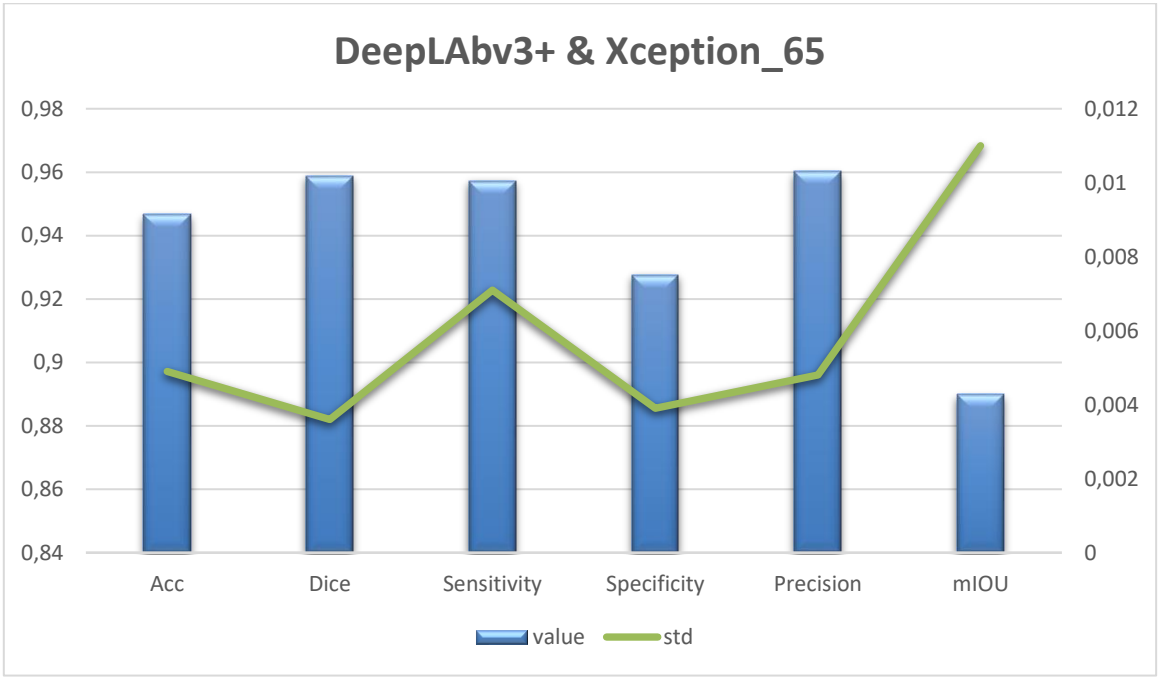

Figure 10.15. Visual representation of DeepLabv3+ and Xception_65 as backbone performance evaluation. Relevant segmentation metrics are shown in bar charts with corresponding standard deviation shown in line plot.

The segmentation performance of the DeepLabv3+ model using the ResNet101 backbone is shown in Figure 10.16.

The model offers balanced segmentation performance and reliable true positive identification with consistent high scores of 0.9545 ± 0.0063 accuracy, 0.9574 ± 0.0037 Dice coefficient, 0.9526 ± 0.0088 sensitivity, and 0.9622 ± 0.0032 precision. The specificity of 0.9314 ± 0,0039 indicates a slight decrease, suggesting limited false positive predictions. The mIOU of 0.8868 ± 0,0147 is the lowest, indicating difficulties with correct pixel-wise overlap.

The standard deviation is low and largely consistent when compared to other metrics. However, it significantly increases for mIOU, indicating a higher degree of variance in border alignment between samples. Even though fine-grained segmentation borders are still difficult to achieve, this performance trend shows that the ResNet101 backbone allows for robust feature extraction.

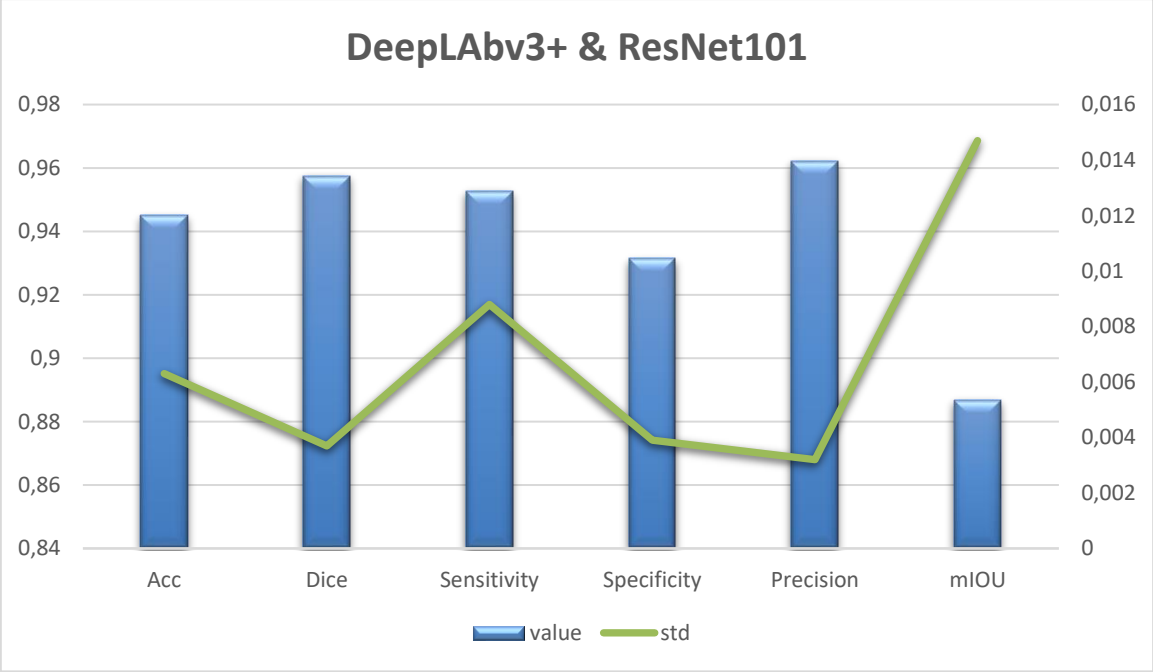

Figure 10.16. Visual representation of DeepLabv3+ and ResNet101 as backbone performance evaluation. Relevant segmentation metrics are shown in bar charts with corresponding standard deviation shown in line plot.

The DeepLabv3+ model's performance evaluation using MobileNetV2 as the backbone is shown in Figure 10.17. The results show a strong performance on most metrics, with Dice, sensitivity, and precision reaching values of 0.9499 ± 0.003, 0.951 ± 0.0096, and 0.949 ± 0.0047, respectively.

Overall robustness is established by an accuracy of 0.9351 ± 0,0049 and a specificity of 0.906 ± 0,0073. Despite having high Dice and sensitivity scores, the mIOU measure is lower (0.8674 ± 0.0119), indicating some limitations in pixel-wise overlap. All measures show relatively low standard deviations, with sensitivity and mIOU showing the most variability.

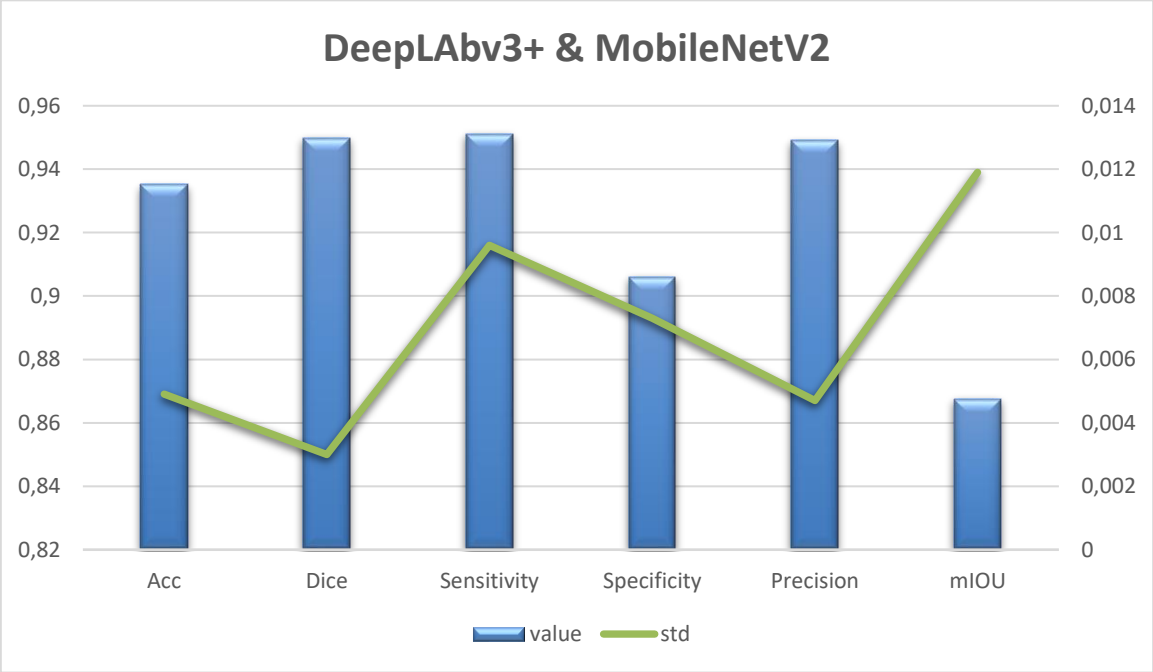

Figure 10.17. Visual representation of DeepLabv3+ and MobileNetV2 as backbone. Relevant segmentation metrics are shown in bar charts with corresponding standard deviation shown in line plot.

The SegformerB0 model's segmentation performance across several evaluation metrics is shown in Figure 10.18.

With a Dice of 0.9548 ± 0.0032, sensitivity of 0.9546 ± 0.0073, and precision of 0.9552 ± 0.0046, SegformerB0's overall performance is robust, indicating that the model offers accurate segmentation with a high degree of overlap between predictions and ground truth.

While the specificity of 0.9172 ± 0.0112 shows a relatively reduced ability to accurately identify negative cases compared to positive ones, the accuracy of 0.9415 ± 0.0055 confirms consistent overall performance. Despite strong global performance, the mIOU score of 0.8796 ± 0.0129 is lower than Dice, indicating difficulties in achieving pixel-level overlap. Although sensitivity and mIOU are slightly higher, the standard deviation is often low in terms of variability, indicating variability driven by dataset characteristics.

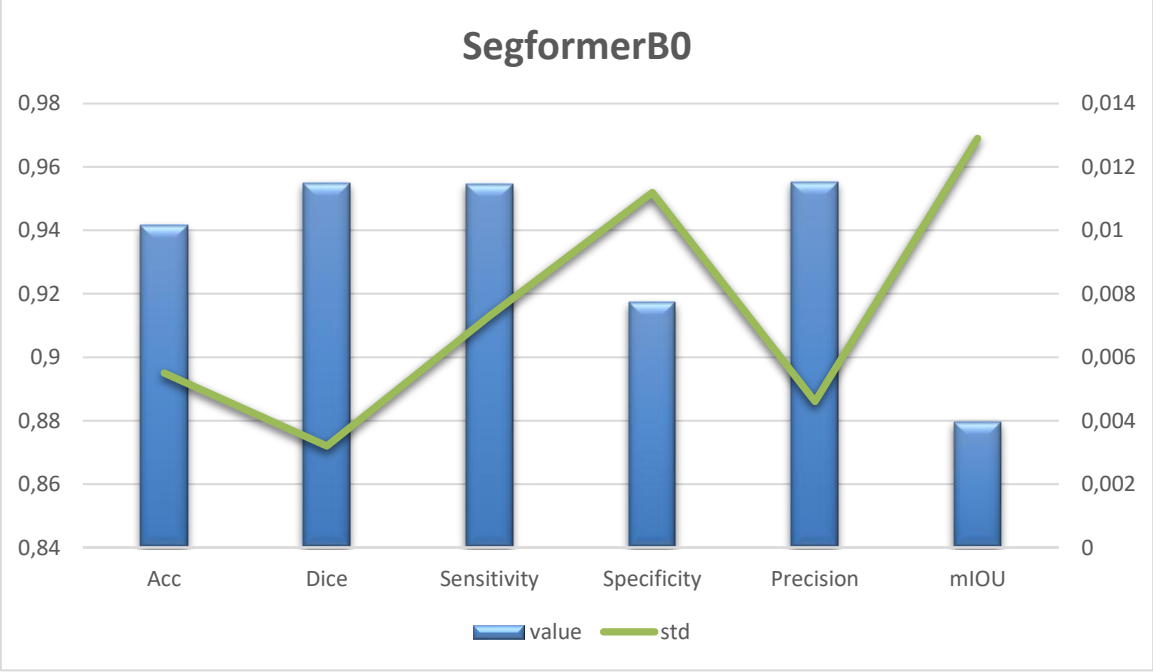

Figure 10.17. Visual representation of SegformerB0 performance evaluation. Relevant segmentation metrics are shown in bar charts with corresponding standard deviation shown in line plot.

Figure 10.18 presents the segmentation performance of the SegformerB3 model. With a Dice score of 0.9622 ± 0.0042, a sensitivity of 0.9631 ± 0.0052, and a precision of 0.9612 ± 0.0062, the results show continuously high performance, demonstrating high true positive detection and reliable prediction accuracy. While specificity is comparatively lower, showing a slightly reduced capacity to accurately categorize negative regions compared to positives, the accuracy of 0.9509 ± 0.0061 and the specificity of 0.9279 ± 0.0152 show strong performance. The mIOU of 0.8979 ± 0.0134 indicates some limitations in precise pixel-wise segmentation overlap.

The overall low variability is significantly higher for specificity and mIOU, indicating dataset-dependent variations in boundary delineation and the negative class.

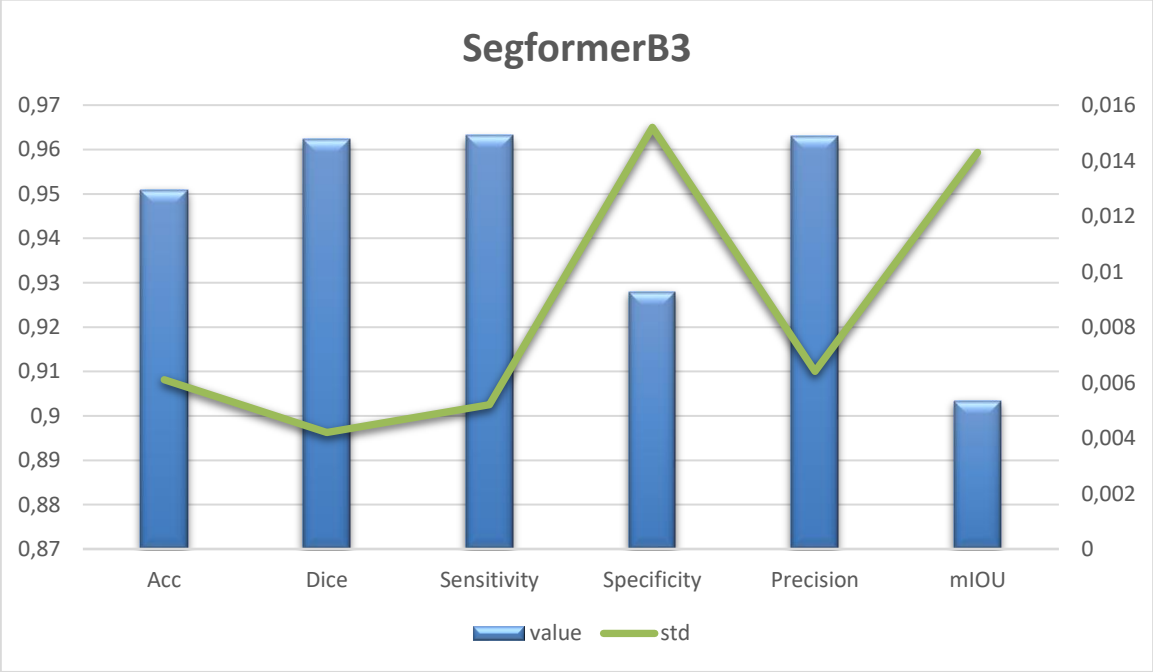

Figure 10.18. Visual representation of SegformerB3 performance evaluation. Relevant segmentation metrics are shown in bar charts with corresponding standard deviation shown in line plot.

Figure 10.19 shows the SegformerB5 model's performance metrics for a segmentation task.

When it comes to both positive and negative class detection, the model consistently performs well across key parameters, with 0.9533 ± 0.0066 accuracy, 0.9641 ± 0.0046 Dice, 0.9682 ± 0.005 sensitivity, 0.9253 ± 0.0172 specificity, and 0.9602 ± 0.0007 precision.

However, the much lower mIOU (0.9024 ± 0.0143) reflects a more severe evaluation of segmentation quality. Standard deviations for the majority of metrics are relatively low, suggesting steady performance.

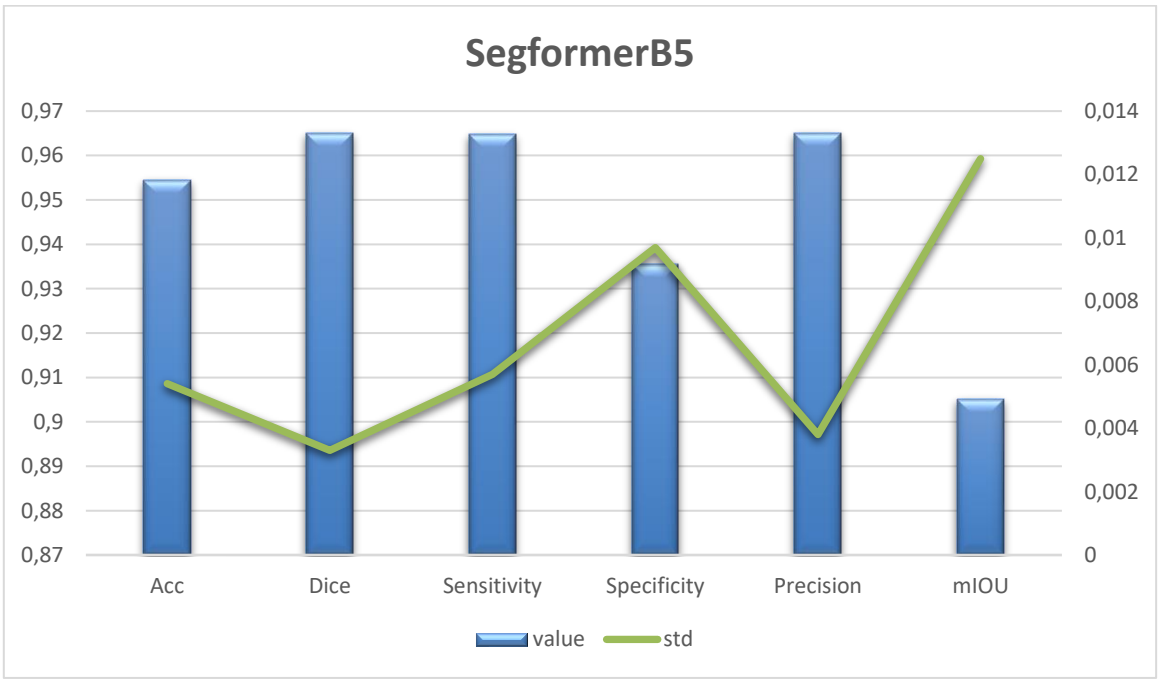

Figure 10.19. Visual representation of SegformerB5 performance evaluation. Relevant segmentation metrics are shown in bar charts with corresponding standard deviation shown in line plot.

The performance evaluation of a U-Net combined with ResNet50 architecture as backbone across a multiple of segmentation metrics is shown in Figure 10.20.

While the mIOU is somewhat lower (0.851 ± 0.0229), the blue bar graph shows high values for accuracy (0.9262 ± 0.0109), Dice coefficient (0.943 ± 0.0071), sensitivity (0.9412 ± 0.0151), specificity (0.8986 ± 0.0151), and precision (0.945 ± 0.0064), indicating great overall performance.

Except for mIOU, which shows the most variability (0.0229), the standard deviation of each metric is represented by the green line plot layered on top. The std is generally low and consistent, ranging from 0.01 to 0.015.

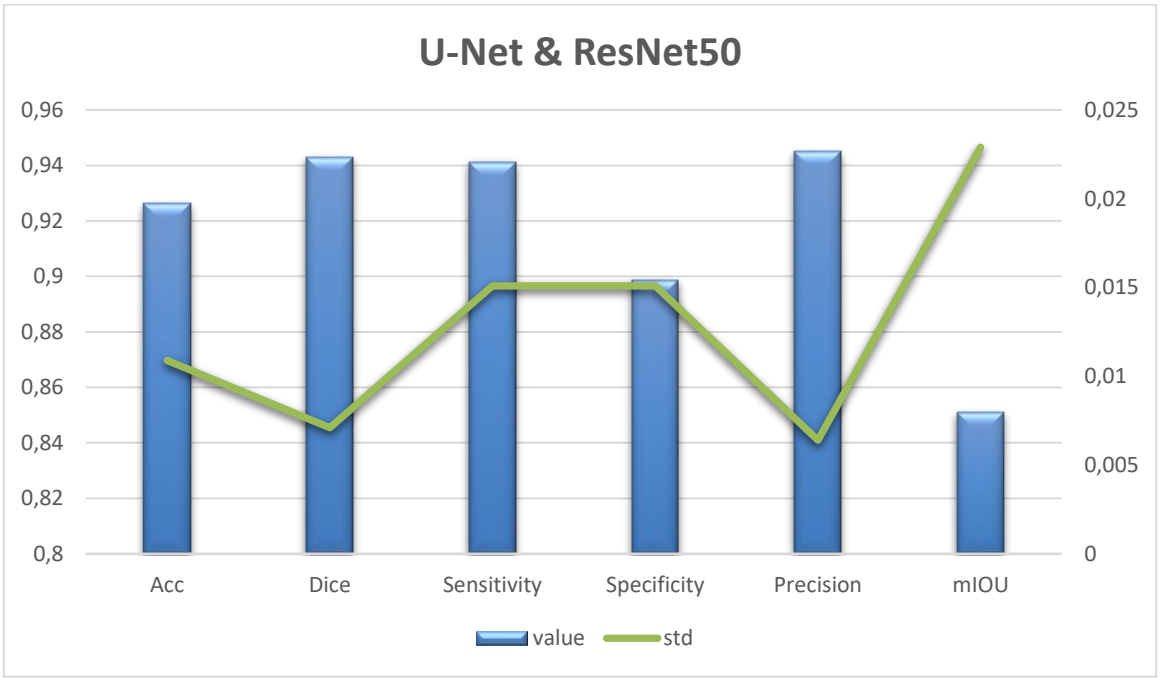

Figure 10.20. Visual representation of U-Net and ResNet50 as backbone performance evaluation. Relevant segmentation metrics are shown in bar charts with corresponding standard deviation shown in line plot.

The performance metrics of the image segmentation U-Net and InceptionV3 as backbone are shown in Figure 10.21.

The blue bar plot shows a high score in terms of accuracy (0.9214 ± 0.0114), precision (0.9372 ± 0.0132), Dice (0.9397 ± 0.0067), and sensitivity (0.9424 ± 0.0096). In contrast, specificity (0.8816 ± 0.0336) and mIOU (0.8413 ± 0.025) are relatively low.

The green overlay line figure indicates that specificity (0.0336) and mIOU (0.025) have significantly larger variability than accuracy, dice, sensitivity, and precision, which have very low variability (~0.01–0.015).

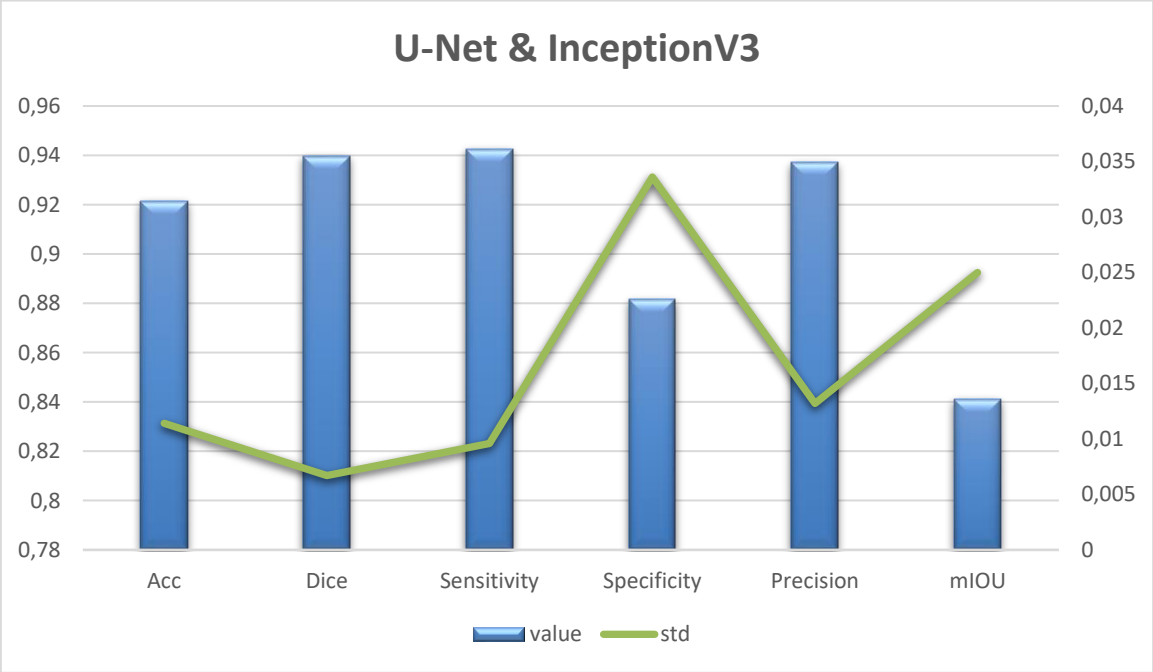

Figure 10.21. Visual representation of U-Net and InceptionV3 as backbone performance evaluation. Relevant segmentation metrics are shown in bar charts with corresponding standard deviation shown in line plot.

A quantitative assessment of the U-Net architecture with InceptionResNetV2 as the backbone is shown in Figure 10.22.

The outcomes demonstrate high performance on most metrics, with substantial segmentation capacity, as demonstrated by Dice of 0.9371 ± 0.8367, sensitivity of 0.9369 ± 0.0183, and precision of 0.9376 ± 0,0082. With a relatively lower specificity of 0.8856 ± 0,0122 and a comparatively high accuracy of 0.9184 ± 0,0109, the identification of true negatives appears to be fairly balanced.

However, the mIOU of (0.8367 ± 0,0225) is low, indicating that it is more difficult to achieve perfect spatial overlap. Other than mIOU, where variability is significantly higher, standard deviations are low overall (<0.0225), which shows that the model is stable. This indicates that region-level segmentation is inconsistent with pixel-wise measurements.

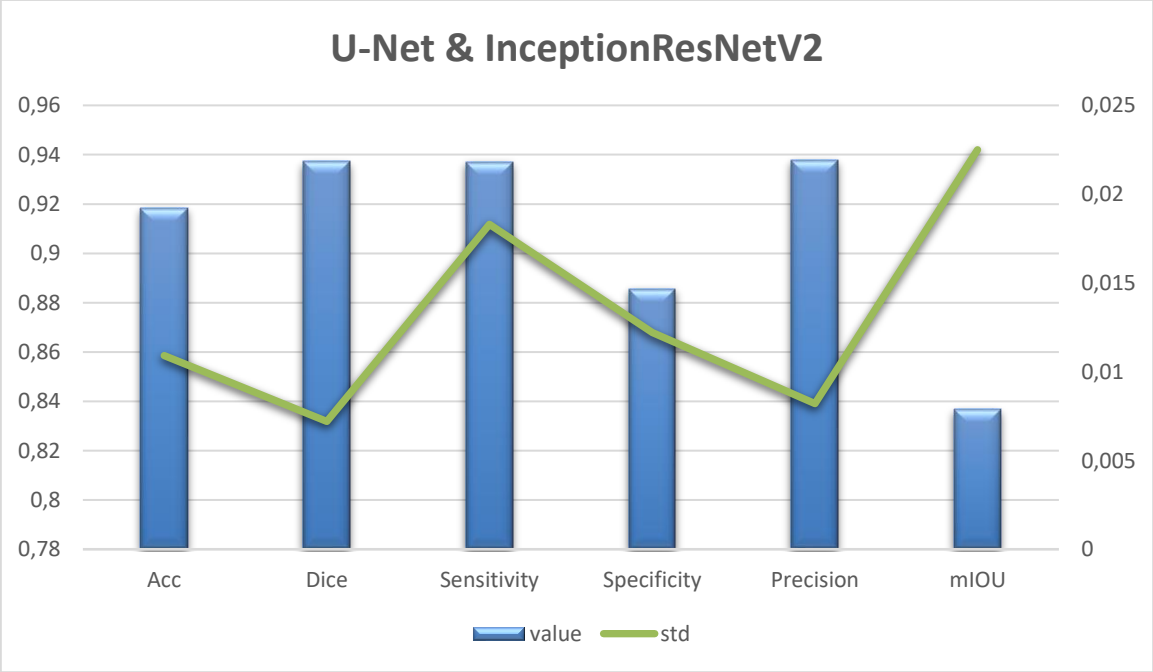

Figure 10.22. Visual representation of U-Net and InceptionResNetV2 as backbone performance evaluation. Relevant segmentation metrics are shown in bar charts with corresponding standard deviation shown in line plot.

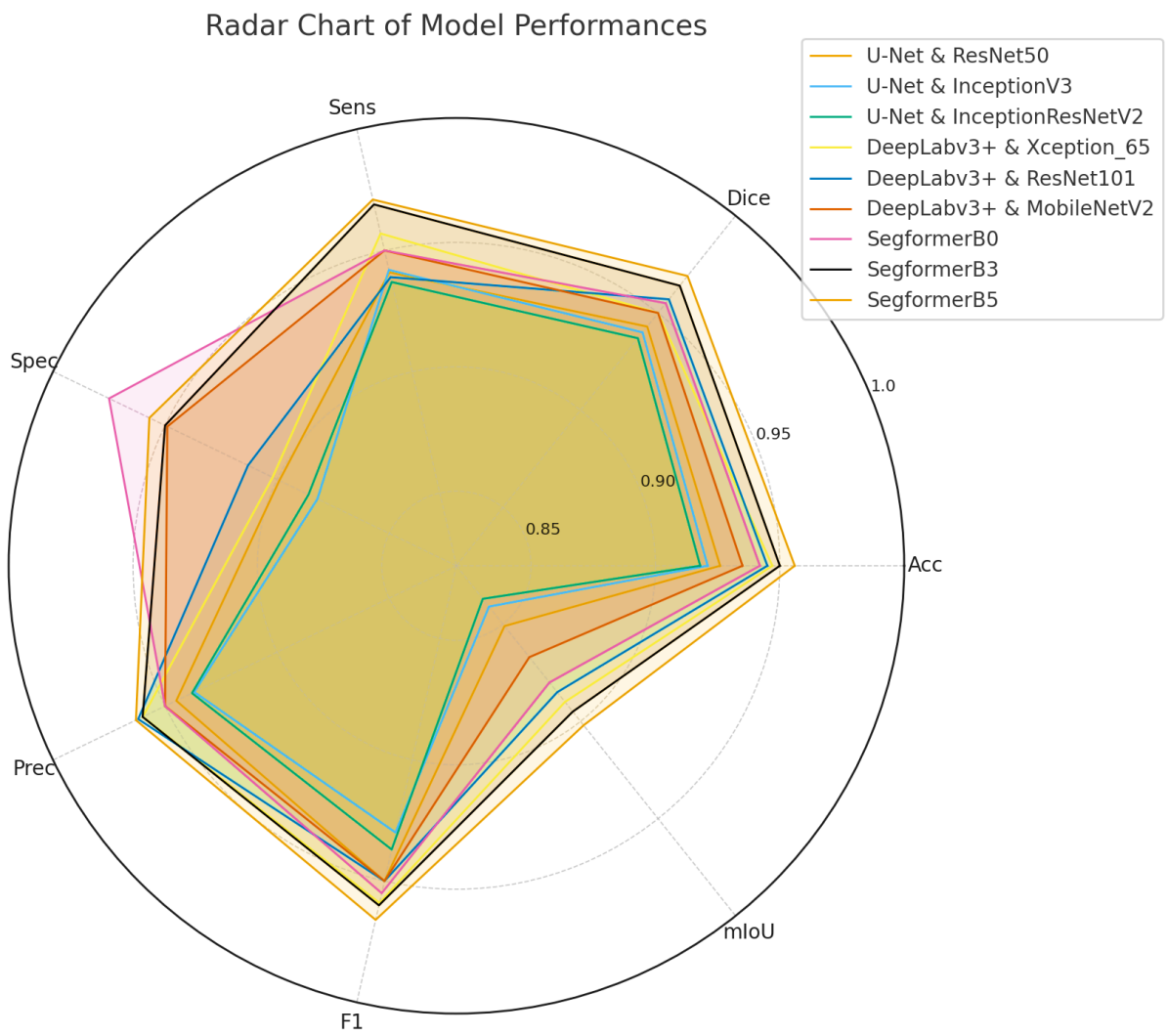

Figure 10.23. Radar chart of models' performances for semantic segmentation on tumor and stromal region

Significant variations in performance across several segmentation metrics can be observed by comparing the U-Net, DeepLabv3+, and Segformer models, as seen in Figure 10.23.

With accuracy values ranging from 0.918 to 0.926 and mIOU between 0.837 and 0.851, the performance of the traditional U-Net implementations (ResNet50, InceptionV3, and InceptionResNetV2) was the lowest across all metrics. These models showed limitations with overall consistency, although they performed well in terms of Dice and sensitivity values.

When compared to the U-Net, the DeepLabv3+ performed better. Stronger generalization was demonstrated by models using Xception_65 and ResNet101 backbones, which achieved satisfactory Dice scores (0.950–0.957) and mIOU values of nearly 0.89.

Furthermore, the Segformer outperformed DeepLabv3+ and U-Net. When it came to accuracy (0.953), Dice (0.964), sensitivity (0.968), and mIOU (0.902), SegformerB5 produced the best and most balanced results. With minimal difference, SegformerB3 achieved lower but still cutting-edge performance.

These results demonstrate how transformer-based segmentation models outperform conventional CNN-based designs. Since the SegFormerB5 model performed the best out of all the models evaluated it was of particular interest in this research. Building on its baseline performance, the proposed SegFormer-LWE model was created as an upgraded version of the SegFormer-B5 model in order to further enhance segmentation quality.

The baseline SegFormer-B5 achieves mIOU of 0.902 ± 0.014, which has been utilized as a benchmark to evaluate the effect of luminance wavelet improvement. The robustnes of the LWE technique in strengthening structural detail and improving overall segmentation performance is demonstrated by the fact that using the proposed preprocessing pipeline typically results in improvements across the majority for evaluation metrics.

A quantitative comparison of the proposed SegFormer-LWE model and the baseline SegFormer-B5 model is shown in Table 10.2. The parameters presented include mIOU, Dice score, accuracy, precision, sensitivity, and specificity, along with their standard deviation.

Table 10.2. A quantitative analysis between the baseline SegFormer-B5 model and the proposed SegFormer-LWE model developed utilizing various wavelet types and scale-factor configurations**.**

SegformerB5		mIOU ± σ	F1 ± σ	Accuracy ± σ	Precision ± σ	Sensitivity ± σ	Specificity ± σ
Original		0.902 ± 0.014	0.964 ± 0.004	0.953 ± 0.006	0.960 ± 0.007	0.967 ± 0.005	0.925 ± 0.017
Segformer-LWE							
Wavelet	Scale factor H, V, D						
Sym3	1.8, 1.8, 1.8	0.906 ± 0.013	0.966 ± 0.004	0.955 ± 0.006	0.960 ± 0.004	0.971 ± 0.008	0.924 ± 0.017
Sym5	2.2, 2.2, 2.2	0.905 ± 0.012	0.965 ± 0.003	0.954 ± 0.006	0.963 ± 0.007	0.967 ± 0.006	0.930 ± 0.015
Db3	2.3, 2.4, 2.3	0.906 ± 0.011	0.966 ± 0.003	0.955 ± 0.005	0.962 ± 0.005	0.969 ± 0.006	0.929 ± 0.015
Db5	2.2, 2.2, 2.2	**0.907 ± 0.011**	**0.967 ± 0.003**	**0.956 ± 0.005**	**0.965 ± 0.005**	**0.967 ± 0.008**	**0.934 ± 0.008**
Db6	2.4, 2.4, 1.6	0.904 ± 0.012	0.965 ± 0.003	0.955 ± 0.006	0.962 ± 0.004	0.967 ± 0.006	0.929 ± 0.015
Coif2	2.0, 2.0, 2.0	0.904 ± 0.013	0.965 ± 0.004	0.954 ± 0.006	0.960 ± 0.002	0.970 ± 0.007	0.925 ± 0.014
Bior4.4	2.5, 2.5, 1.1.	0.905 ± 0.014	0.966 ± 0.004	0.955 ± 0.007	0.961 ± 0.007	0.970 ± 0.004	0.927 ± 0.017

Table 10.2. shows the seven best-performing configurations of the proposed SegFormer-LWE model, determined by an exhaustive grid-search of multiple wavelet families and scale factors. Each configuration shown reflects one of the top parameter combinations that achieved the maximum performance across the evaluated metrics.

Among the analyzed wavelets, the Db5 with scale factors of H = 2.2, V = 2.2 and D = 2.2 achieves the highest overall performance. It produces a mIOU of 0.907 ± 0.011, reflecting the best improvement over the baseline, combined with the top Dice score (0.967 ± 0.003) and accuracy (0.956 ± 0.005). Furthermore, Db5 also gives the highest specificity (0.934 ± 0.008), showing a higher capacity to correctly identify background regions without increasing false positives.

Other versions of wavelets, such as sym3, sym5, db3, db6, coif2, and bior4.4, also exhibit modest increases over the baseline model but are lower than the performance achieved by Db5. These configurations generally produce slightly lower mIOU and accuracy values,

indicating that the degree of enhancement is strongly influenced by the wavelet family and scale factor selection.

The contribution of wavelet-enhanced preprocessing to better segmentation performance is shown graphically in Figure 10.24. The comparison displays both the original histopathology image and its LWE preprocessed version, which highlights tissue boundaries and subtle structures. This improvement enables the model to learn more discriminative features. Higher performance metrics are consistent with the preprocessed image's improved structural representation, which enables the network to more accurately identify tissue sections.

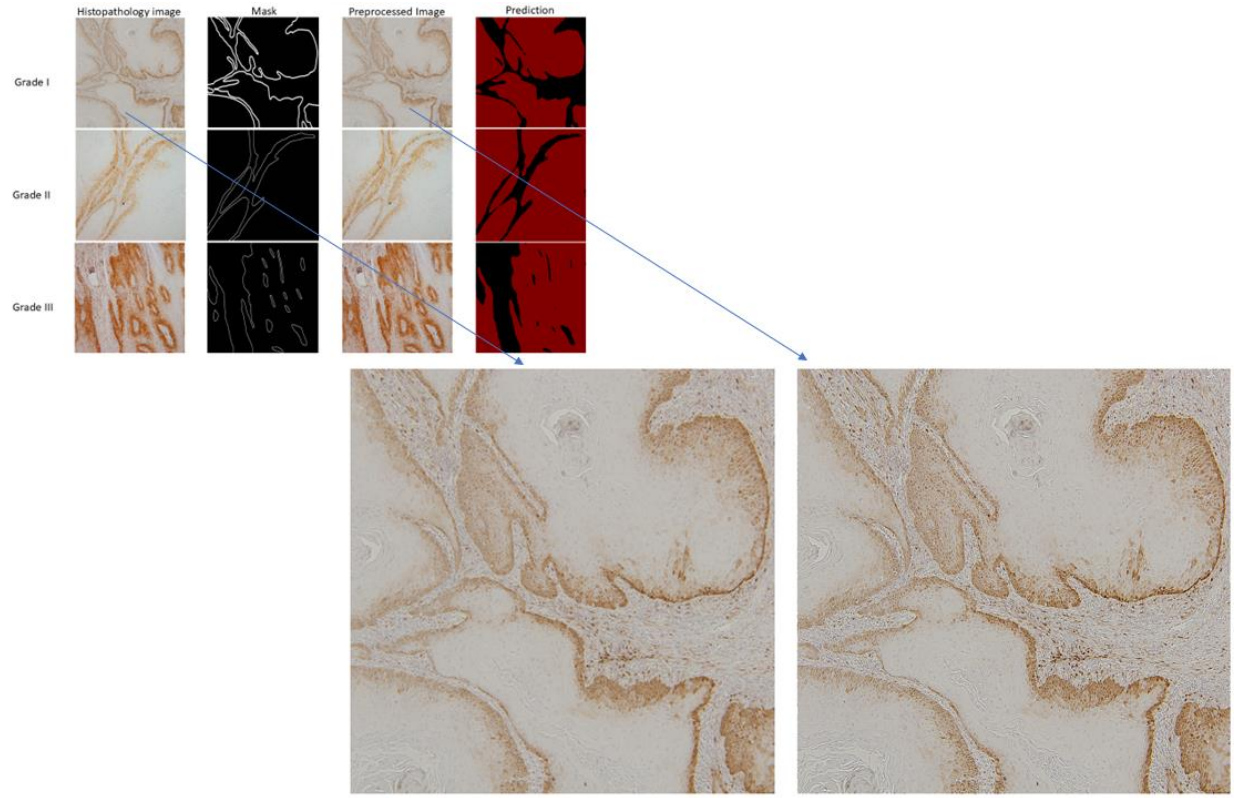

Figure 10.24. Visual representation of histopathology images, ground truth masks, preprocessed images, and semantic segmentation results. The original image and its LWE preprocessed equivalent are shown in the magnified photos on the right, giving a clear comparison of how preprocessing improves tissue appearance for additional analysis.

The results clearly demonstrate that the luminance-wavelet enhancement approach, as applied in SegFormer-LWE, boosts segmentation quality compared to the baseline SegFormer-B5 model.

The mIOU metric averages performance across all classes and all pixels, giving it a highly sensitive and robust measure. As a result, a 0.5% improvement shows that the model consistently produces more accurate pixel-level predictions throughout the whole dataset. Overall, even slight improvements in performance metrics, such as a 0.5% improvement in mIOU, are often considered significant.

Performance improvements have become even more challenging to achieve for cutting-edge transformer-based models, such SegFormer-B5. These models are substantially optimized through extensive pretraining and already function close to the upper bounds of representational capability. It frequently takes significant algorithmic or architectural innovation rather than simply hyperparameter tuning to get an additional 0.5% mIOU increase. Improvements at this level show that the proposed enhancement, the LWE preprocessing, is contributing important new information beyond what the original transformer can extract on its own. Furthermore, literature benchmarks commonly highlight increases of 0.3–0.7% as state-of-the-art advances, demonstrating that advancement in this research is competitive with leading research advancement.

10.4. Automatic quantification of TSR

Regions with the largest proportion of tumor-associated stroma were selected at 10× magnification for automatic TSR assessment. Analysis was limited to fields that included tumor cells on each of the four microscopic view borders. Cases in which the tumor-associated stroma occupied more than 50% of the selected field were classified as stroma-high, whereas those with 50% or less were defined as stroma-low. In earlier research, this 50% cutoff was frequently used as a reliable predictive subgroup discriminator. Areas with preexisting lymphoid clusters, necrosis, or other normal tissue components were not included in the research. If these factors were not completely preventable, they were not included in the tumor-associated stroma computational estimation. Furthermore, tumor and stromal regions were then automatically defined using semantic segmentation algorithm based on morphological, color, and texture characteristics.

The proportional area of each compartment was quantified computationally, TSR was computed as the ratio of tumor area to the overall area of tumor plus stroma. The automated workflow used to determine the tumor-stroma ratio is shown in Figure 10.25.

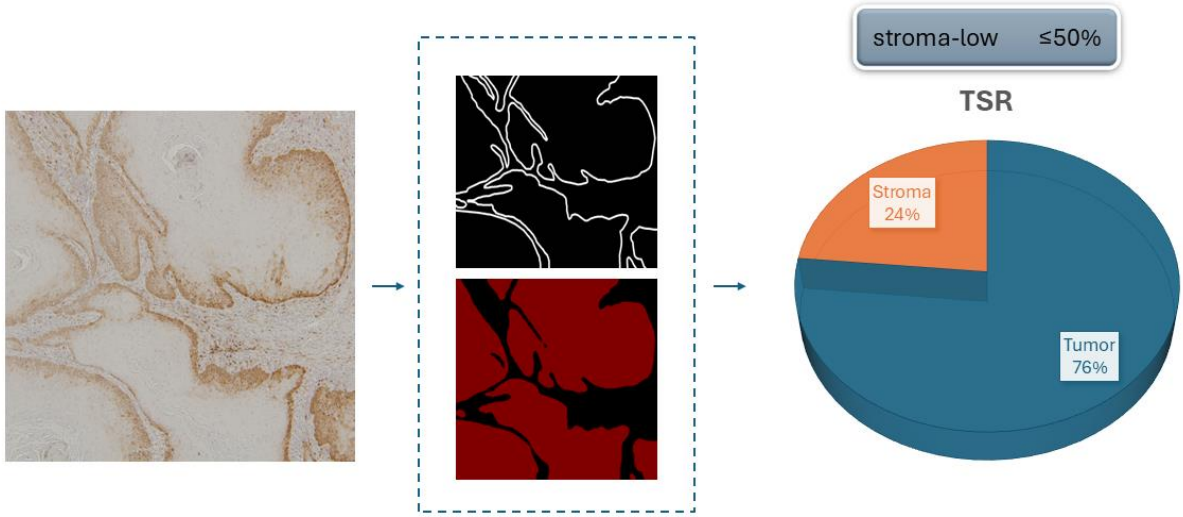

Figure 10.25. The automated process to assess the tumor-stroma ratio (TSR). A representative histologic image (left) that displays the surrounding tumor-associated stroma and tumor epithelial areas was prepared for digital segmentation (in the middle). While the lower panel displays classified regions with tumor (black area) and stroma (red area), the upper panel displays the tissue border detection map. A TSR of 76% tumor and 24% stroma was obtained by automatically calculating the proportionate areas of the two sections. This case was classified as stroma-low (≤50% stroma) based on the predetermined 50% limit.

In 40-patient cohort, the relationship between the tumor–stroma ratio and several clinicopathologic characteristics was assessed. As a histopathologic marker that represents the percentage of tumor-associated stroma in the tumor microenvironment, TSR was analyzed in order to determine if it correlated with known prognostic factors like patient age, lymph node status, tumor grade, and alcohol and smoke consumption.

Prior research on head and neck cancers has shown that a high stroma percentage (stroma-high) is frequently related to less favorable clinical outcomes and more aggressive tumor behavior. The correlation between TSR and clinicopathologic parameters was examined in the cohort to determine whether similar trends exist in this patient population. Table 10.3. summarizes the findings of this analysis.

Table 10.3. Correlation between the tumor-stroma ratio and the clinicopathologic characteristics of oral squamous cell carcinoma.

Variable	Total	Tumor-stroma ratio		*P*
		Stroma-low	Stroma-high	
	N = 40	Number (%) 28 (70%)	Number (%) 12 (30%)	
Gender				
Male	28	19 (68%)	9 (32%)	0.94
Female	12	9 (75%)	3 (25%)	
Alcohol intake				
Yes	15	10 (75%)	5 (25%)	1.00
No	25	18 (72)	7 (28%)	
Smoking				
Yes	22	14 (64%)	8 (36%)	0.53
No	18	14 (78%)	4 (22%)	
Age				
To 49	2	1 (50%)	1 (50%)	
50-59	5	5 (100%)	0 (0%)	0.28
60-69	22	16 (73%)	6 (27%)	
+70	11	6 (55%)	5 (45%)	
Grade				
I	18	12 (67%)	6 (33%)	0.74
II	16	11 (69%)	5 (31%)	
III	6	2 (33%)	4 (67%)	
Lymh Node Metastases				
Yes	21	14 (67%)	7 (33%)	0.89
No	19	14 (74%)	5 (26%)	

Several patterns observed in this research are biologically consistent with previously established findings confirming the predictive value of the tumor–stroma ratio (TSR) in OSCC, even if statistical significance was not attained for any of the clinicopathological markers in our cohort ($p > 0.05$). In particular, clinically unfavorable categories had a larger percentage of stroma-high tumors. A richer stromal environment may facilitate metastatic

spread, as demonstrated by the finding that that up to one-third of patients with lymph node metastases had a high stromal amount, compared to just over a quarter of individuals without nodal involvement. Furthermore, the finding that that the oldest age group (≥70 years), which usually correlates with poorer cancer survival, showed the highest relative proportion of stroma high tumors (45%) supports the theory that stroma high patterns may be linked to systemic and microenvironmental conditions that increase tumor aggressiveness in older people.

Although not statistically significant, the highest percentage of stroma-high tumors in G3 group is biologically significant since it indicates a more aggressive tumor microenvironment. A dense, active stroma rich in cancer-associated fibroblasts frequently supports high-grade OSCC, increasing invasion, metastasis, and resistance to therapy. Since stromal activity may actively promote tumor development in advanced disease, the higher percentage of stroma-high tumors in G3 patients emphasizes the potential value of TSR as a clinically relevant biomarker.

On the other hand, clinically less aggressive patient categories were associated with a higher probability of having stroma-low tumors. These comprised patients between the ages of 50 and 59, who showed only stroma low tumors (100%), as well as non-drinkers and non-smokers. These results indirectly corroborate with studies in the literature that stroma low cancers typically exhibit lower invasiveness, reduced metastatic potential, and slower development dynamics since these populations usually correspond to better outcomes.

Kaplan-Meier survival curves demonstrated that TSR had a strong predictive value for overall survival. As seen in Figure 10.26, patients with stroma-high tumors had significantly worse survival rates.

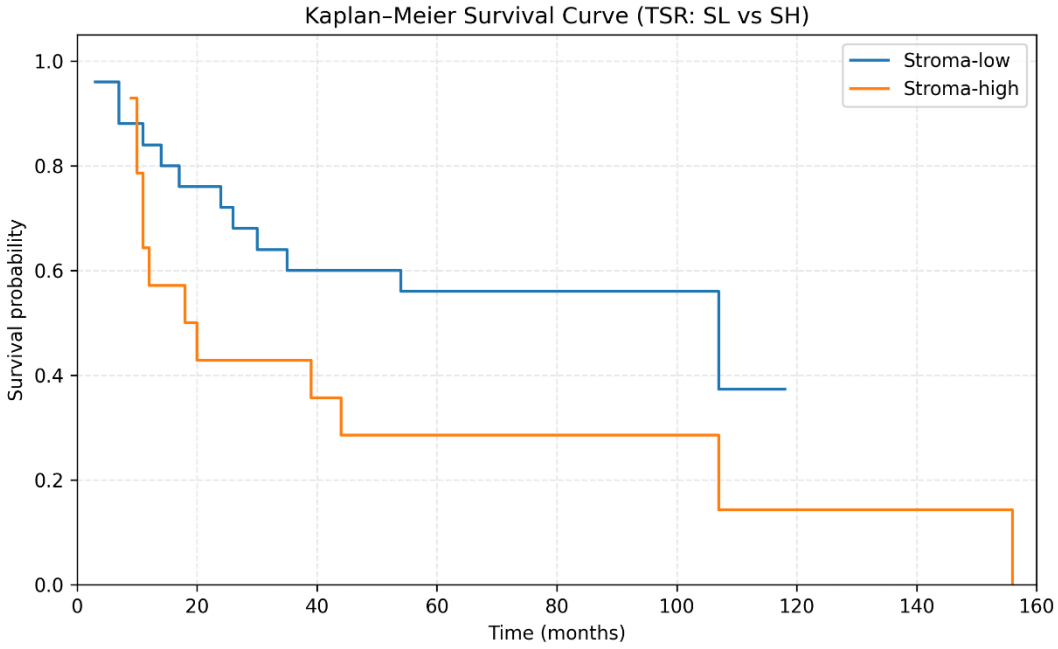

Figure 10.26. Kaplan-Meier analysis of overall survival in patients with stroma-low versus stroma-high OSCC tumor

The Kaplan–Meier survival curve shows that stroma-low (SL) and stroma-high (SH) tumors consistently have different overall survival rates. The survival probability declined more quickly in patients with SH tumors, especially in the early follow-up period, indicating earlier mortality and faster disease progression. The SL group, on the other hand, had higher survival probability for most of the observation period, with a more progressive decline and several long-term survivors who outlived the 100-month follow-up period.

The visual comparison of the two curves confirms prior studies indicating that TSR actively contributes to the aggressiveness of oral squamous cell carcinoma. A biologically active stromal environment, marked by cancer-associated fibroblasts and elevated pro-tumor signaling pathways that aid in invasion, rejection of the immune system, and metastatic growth, may be the cause of stroma-high cancers' worse survival rates. Thus, this research survival analysis's trend confirms TSR's possible predictive significance, particularly when used as a supporting biomarker in conjunction with well-established clinicopathologic variables.

10.5. Experimental Proof of Concept

In order to validate and confirm the results of the research, this chapter presents an experimental proof-of-concept (PoC) framework which demonstrates the consistency and robustness of the proposed methods. While acknowledging that bigger datasets are necessary for statistical generalization, the PoC evaluation was conducted on a small cohort of eight patients, which is adequate for verifying the observed research patterns and system behavior. Multiclass classification and semantic segmentation are two hybrid approaches that are integrated into the proposed framework, as shown in Figure 10.27.

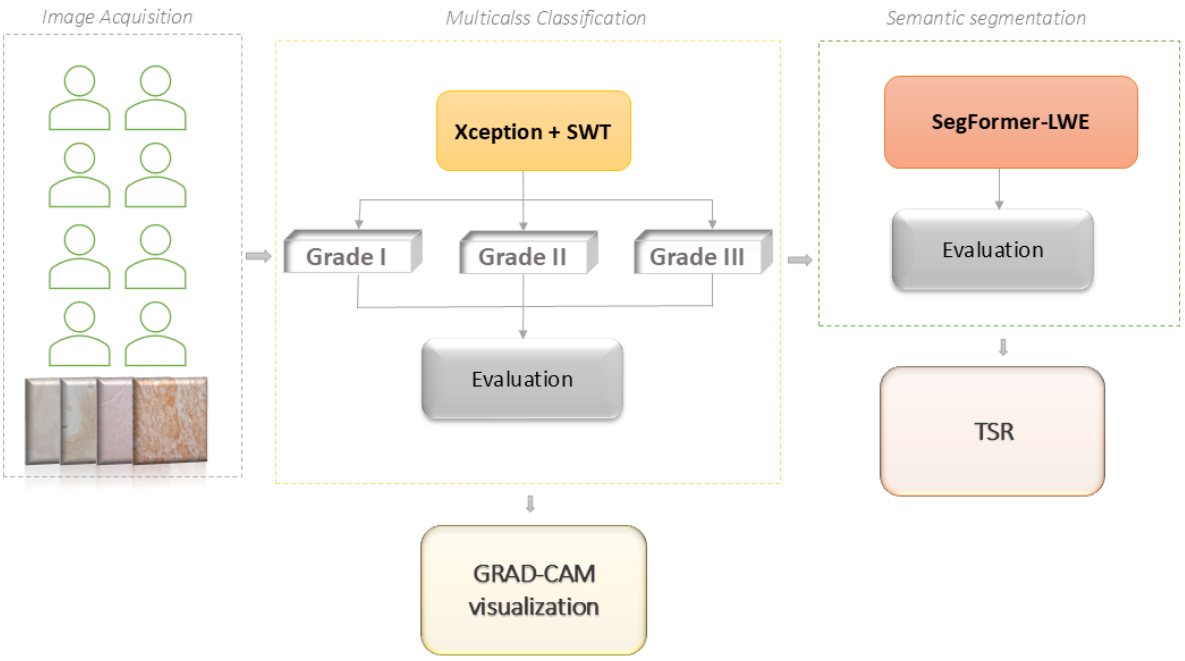

Figure 10.27. An outline of the proposed experimental framework for proof-of-concept. The first step in the process is obtaining medical images from a small group of patients. Then, the images go through two analytical branches. The first branch is a multiclass classification module that uses Grad-CAM visuals to support model interpretability. It is built on a hybrid SWT–Xception model. A semantic segmentation module using the SegFormer-B5 architecture and LWE preprocessing makes up the second branch. The Tumor–Stroma Ratio, a quantitative biomarker with clinical relevance, is calculated once the segmentation outcomes are evaluated.

The initial step consists of collecting medical images from the chosen patient group. These images serve as the raw input for pipeline used for multiclass classification and semantic segmentation. Despite the small sample size, the dataset represents realistic inter-patient heterogeneity, which is crucial for evaluating the resilience of the proposed system within a proof-of-concept context.

The aim of the multiclass classification task is to assign three classes (Grade I, Grade II, and Grade III) to the input images. This is achieved by using a hybrid deep learning model that combines Xception with Stationary Wavelet Transform for image preprocessing. Standard evaluation measures are used to evaluate model performance, while Grad-CAM visualizations are used to aid in qualitative interpretability.

The framework incorporates a semantic segmentation branch to precisely localize and delineate medically significant regions. This module uses a second hybrid model, which consists of SegFormer-B5, a transformer-based segmentation architecture, and LWE for image preprocessing. To determine the accuracy of region boundaries, segmentation outputs are assessed using relevant quantitative measures. The Tumor–Stroma Ratio is calculated as the ratio of tumor area to stromal area in the tissue under analysis based on the final segmentation masks.

The quantitative performance of the proposed hybrid models employed in the PoC for multiclass classification and semantic segmentation tasks can be seen in Table 10.3.

Table 10.4. Quantitative performance metrics of the proposed models in the proof-of-concept

	$\mathbf{AUC_{macro}}$	$\mathbf{AUC_{micro}}$				
Xception + SWT	0.992	0.973				
	mIOU	**F1**	**Accuracy**	**Precision**	**Sensitivity**	**Specificity**
SegFormer-LWE	0.897	0.964	0.956	0.965	0.964	0.939

With an AUC_{macro} of 0.992, the Xception + SWT model demonstrates balanced classification across all three classes. High overall classification performance is further confirmed by the AUC_{micro} value of 0.973. These findings show that, even with limited data, the proposed Xception + SWT model has considerable discriminative potential.

A Grad-CAM visualization of the proposed classification model is shown in Figure 10.28. The Grad-CAM visualization is displayed on the top, and the original histopathological image is displayed on the bottom.

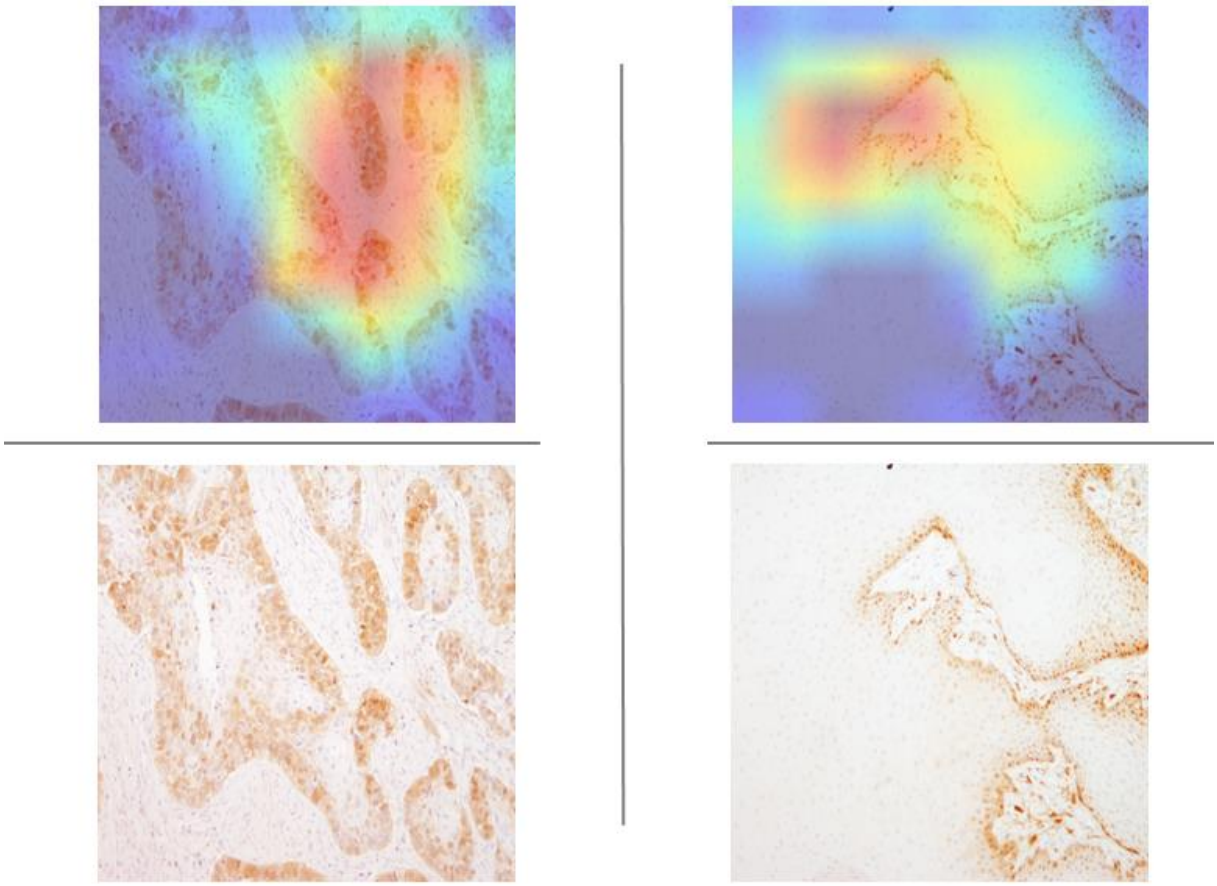

Figure 10.28. An illustration of the Grad-CAM heatmap that highlights discriminative tissue regions that the proposed Xception + SWT hybrid model uses to grade histological images.

In order to provide visual proof that the identified features correlate with pathological information, Grad-CAM visualization improves the interpretability of the proposed model and enhances the transparency of the classification results throughout the proof-of-concept framework.

Several metrics are used to evaluate the SegFormer–LWE model's performance for the semantic segmentation task. A significant level of spatial overlap between the predicted and ground-truth segmentation masks is reflected in the model's mIoU of 0.897. Strong and reliable pixel-level classification has been demonstrated by the Dice score of 0.964, high precision (0.965), and accuracy of 0.956. Furthermore, accurate non-target area discrimination is verified by sensitivity of 0.964 and specificity of 0.939.

The TSR estimation confirms the repeatability of the proposed system by reflecting the same patterns observed in results of the primary research. This agreement shows that the corresponding TSR values represent a stable and clinically significant biomarker within the proof-of-concept evaluation, and that the semantic segmentation approach accurately represents tumor and stromal regions.

11.Conclusions and Future Work

One method for classifying cancer cells based on tissue abnormalities is histology grading. It depends on the clinician's subjective component, which could have a negative impact on the patient's results and the most effective course of therapy. This research demonstrates the significant potential of using AI algorithms in conjunction with image processing approaches to improve OSCC prognosis and improve survival rates.

In the first stage of the research, the author demonstrates how to incorporate a wavelet coefficient mapping function and the SWT with deep convolutional neural networks for OSCC multiclass grading. According to experimental results, the Xception architecture and SWT combination produced the best classification performance, with AUC_{macro} and AUC_{micro} of 0.963± 0.042 and 0.966 ± 0.027, respectively.

The following stage was the implementation of Grad-CAM visualization. Grad-CAM is used to create heatmaps for multiclass classification, which emphasize important regions in histopathology images. These heatmaps help healthcare professionals distinguish pathologically significant features from irrelevant or potential artifacts by visually evaluating sensitivity of the model to critical regions. This method provides a more comprehensive analysis with less unpredictability and human error than conventional single-model approaches.

The third step involved semantic segmentation. With the Db5 wavelet and scale factors of H = 2.2, V = 2.2, and D = 2.2, the proposed SegFormer-LWE model produces the best overall results for semantic segmentation. It achieves the best improvement over the baseline model with a mIOU of 0.907 ± 0.011, along with the highest accuracy (0.956 ± 0.005) and Dice score (0.967 ± 0.003). Additionally, SegFormer-LWE has the highest specificity (0.934 ± 0.008), indicating a greater ability to accurately identify background regions without raising

false positives. 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.

In the last stage of the research, the tumor-stroma ratio was automatically quantified. Automated methods improve diagnostic consistency and reduce interobserver variability by precisely segmenting the tumor and stromal areas. According to the results of this research, OSCC patients with a low TSR (stroma-high tumors) have a unfavorable prognosis for survival.

Based on the results of the experimental proof of concept, an AI-based system has been proven successful in terms of multiclass grading, Grad-CAM visualization, semantic segmentation as well as automatic quantification of TSR and has a great potential in the prediction of tumor invasion and outcomes of patient with OSCC.

Further research should employ a dataset with more histopathological images to create a more dependable model, as the data availability of this research was limited. Additionally, a wider variety of oral cancer subtypes should be included in the dataset to increase the system's generalizability in various clinical applications. This would enable the system to record a wider variety of morphological traits.

In order to create a more comprehensive overview of tumor biology, future research should also consider incorporating multimodal data sources, such as molecular markers and genomic profiles. Precision oncology may benefit from the integration of various technologies since it will enable more precise prognostic evaluations and direct individualized treatment plans.

To increase practical relevance and generalizability of AI models, extensive prospective validation in real healthcare settings is required. In addition to proving the dependability of model in real-world scenarios, this kind of validation would highlight any potential drawbacks that would not be apparent in controlled experimental or retrospective research. A realistic approach to this process would be to include the AI-based system in actual diagnostic procedures, initially serving as an advisor or support system rather than a decision-maker on

its own. The ability to directly compare the output of AI with the skilled interpretations of pathologists would enable a methodical assessment of the AI-based system accuracy and potential utility.

Bibliography

[1] Addison, P. S. (2017). The illustrated wavelet transform handbook: Introductory theory and applications in science, engineering, medicine and finance. CRC Press.

[2] Adeli, H. (1988). Expert systems in construction and structural engineering. CRC Press.

[3] Afify, H. M., Mohammed, K. K., & Hassanien, A. E. (2023). Novel prediction model on OSCC histopathological images via deep transfer learning combined with Grad-CAM interpretation. Biomedical Signal Processing and Control, 83, 104704.

[4] Albasri, A. M., Ali, A. H., & Nathiha, A. A. (2015). Segmentation of immunohistochemical staining of β-catenin expression of oral cancer using EM algorithm. Journal of Taibah University Medical Sciences, 10(2), 169–174.

[5] Anghel, A., Stanisavljevic, M., Andani, S., Papandreou, N., Rüschoff, J. H., Wild, P., ... & Pozidis, H. (2019). A high-performance system for robust stain normalization of whole-slide images in histopathology. Frontiers in Medicine, 6, 193.

[6] Ayodele, T. O. (2010). Machine learning overview. In O. E. Olugbara (Ed.), New advances in machine learning (pp. 9–18). InTech.

[7] Bagan, J., Sarrion, G., & Jimenez, Y. (2010). Oral cancer: Clinical features. Oral Oncology, 46(6), 414–417

[8] Baik, J., Ye, Q., Zhang, L., Poh, C., Rosin, M., MacAulay, C., & Guillaud, M. (2014). Automated classification of oral premalignant lesions using image cytometry and random forests-based algorithms. Cellular Oncology, 37, 193–202.

[9] Banerjee, S., Karri, S. P. K., Chatterjee, S., Pal, M., Paul, R. R., & Chatterjee, J. (2016, January). Multimodal diagnostic segregation of oral leukoplakia and cancer. In 2016 International Conference on Systems in Medicine and Biology (ICSMB) (pp. 67–70). IEEE.

[10] Chen, H., & Sung, J. J. (2021). Potentials of AI in medical image analysis in gastroenterology and hepatology. Journal of Gastroenterology and Hepatology. Advance online publication.

[11] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4), 834–848.

[12] Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics, 21, 6.

[13] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1251–1258).

[14] da Silva, A. V. B., Saldivia-Siracusa, C., de Souza, E. S. C., Araújo, A. L. D., Lopes, M. A., Vargas, P. A., ... & Quiles, M. G. (2024, July). Enhancing Explainability in Oral Cancer Detection with Grad-CAM Visualizations. In International Conference on Computational Science and Its Applications (pp. 151-164). Cham: Springer Nature Switzerland.

[15] Das, D. K., Koley, S., Chakraborty, C., & Maiti, A. K. (2014, December). Automated segmentation of mitotic cells for in vitro histological evaluation of oral squamous cell carcinoma. In 2014 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT) (pp. 354–357). IEEE.

[16] Das, D. K., Chakraborty, C., Sawaimoon, S., Maiti, A. K., & Chatterjee, S. (2015). Automated identification of keratinization and keratin pearl area from in situ oral histological images. Tissue and Cell, 47(4), 349–358.

[17] Das, D. K., Mitra, P., Chakraborty, C., Chatterjee, S., Maiti, A. K., & Bose, S. (2017). Computational approach for mitotic cell detection and its application in oral squamous cell carcinoma. Multidimensional Systems and Signal Processing, 28, 1031–1050.

[18] Das, D. K., Bose, S., Maiti, A. K., Mitra, B., Mukherjee, G., & Dutta, P. K. (2018). Automatic identification of clinically relevant regions from oral tissue histological images for oral squamous cell carcinoma diagnosis. Tissue and Cell, 53, 111–119.

[19] Das, D. K., Koley, S., Bose, S., Maiti, A. K., Mitra, B., Mukherjee, G., & Dutta, P. K. (2019). Computer aided tool for automatic detection and delineation of nucleus from oral histopathology images for OSCC screening. Applied Soft Computing, 83, 105642.

[20] Das, N., Hussain, E., & Mahanta, L. B. (2020). Automated classification of cells into multiple classes in epithelial tissue of oral squamous cell carcinoma using transfer learning and convolutional neural network. Neural Networks, 128, 47-60.

[21] Das, M., Dash, R., Mishra, S. K., & Dalai, A. K. (2024). An Ensemble deep learning model for oral squamous cell carcinoma detection using histopathological image analysis. IEEE Access.

[22] Dharani, R., & Danesh, K. (2024). Oral cancer segmentation and identification system based on histopathological images using MaskMeanShiftCNN and SV-OnionNet. Intelligence-Based Medicine, 100185.

[23] Deepa, C., & Tharageswari, K. (2021). Deep learning for healthcare. In Deep Learning and IoT in Healthcare Systems (pp. 1–25). Apple Academic Press.

[24] Deo, B. S., Pal, M., Panigrahi, P. K., & Pradhan, A. (2024). An ensemble deep learning model with empirical wavelet transform feature for oral cancer histopathological image classification. International Journal of Data Science and Analytics, 1–18.

[25] Douglas, L. (2011). Making oral cancer screening a routine part of your patient care, Part 1. Vital, 9(1), 44–47.

[26] El-Naggar, A. K., Chan, J. K., Grandis, J. R., Takata, T., & Slootweg, P. J. (Eds.). (2017). WHO classification of head and neck tumours. International Agency for Research on Cancer (IARC).

[27] Ettinger, K. S., Ganry, L., & Fernandes, R. P. (2019). Oral cavity cancer. Oral and Maxillofacial Surgery Clinics, 31(1), 13–29.

[28] Feller, L., & Lemmer, J. (2012). Oral squamous cell carcinoma: Epidemiology, clinical presentation and treatment. Journal of Cancer Therapy, 3(4), 263–268.

[29] Feurer, M., & Hutter, F. (2019). Hyperparameter optimization. In Automated Machine Learning (pp. 3–33). Springer, Cham.

[30] Folmsbee, J., Liu, X., Brandwein-Weber, M., & Doyle, S. (2018, April). Active deep learning: Improved training efficiency of convolutional neural networks for tissue classification in oral cavity cancer. In 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018) (pp. 770-773). IEEE.

[31] Ganesh, D., Sreenivasan, P., Öhman, J., Wallström, M., Braz-Silva, P. H., Giglio, D., ... & Hasseus, B. (2018). Potentially malignant oral disorders and cancer transformation. Anticancer Research, 38(6), 3223–3229.

[32] Gerdes, A. (2024). The role of explainability in AI-supported medical decision-making. Discover Artificial Intelligence, 4(1), 29.

[33] Gunawardana, A., & Shani, G. (2009). A survey of accuracy evaluation metrics of recommendation tasks. Journal of Machine Learning Research, 10(12), 1–39.

[34] Gupta, R. K., Kaur, M., & Manhas, J. (2019). Tissue level based deep learning framework for early detection of dysplasia in oral squamous epithelium. Journal of Multimedia Information System, 6(2), 81-86.

[35] Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., & Yener, B. (2009). Histopathological image analysis: A review. IEEE Reviews in Biomedical Engineering, 2, 147–171.

[36] Hagenaars, S. C., Vangangelt, K. M., Van Pelt, G. W., Karancsi, Z., Tollenaar, R. A., Green, A. R., ... & Mesker, W. E. (2022). Standardization of the tumor-stroma ratio scoring method for breast cancer research. Breast Cancer Research and Treatment, 193(3), 545–553.

[37] Halicek, M., Shahedi, M., Little, J. V., Chen, A. Y., Myers, L. L., Sumer, B. D., & Fei, B. (2019, March). Detection of squamous cell carcinoma in digitized histological images from the head and neck using convolutional neural networks. In Medical Imaging 2019: Digital Pathology (Vol. 10956, pp. 112–120). SPIE.

[38] Hameed, K. S., Abubacker, K. S., Banumathi, A., & Ulaganathan, G. (2021). Immunohistochemical analysis of oral cancer tissue images using support vector machine. Measurement, 173, 108476.

[39] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770–778).

[40] Hong, Y., Heo, Y. J., Kim, B., Lee, D., Ahn, S., Ha, S. Y., ... & Kim, K. M. (2021). Deep learning-based virtual cytokeratin staining of gastric carcinomas to measure tumor–stroma ratio. Scientific Reports, 11(1), 19255.

[41] Houben, S., Abrecht, S., Akila, M., Bär, A., Brockherde, F., Feifel, P., ... et al. (2021). Inspect, understand, overcome: A survey of practical methods for AI safety. arXiv preprint arXiv:2104.14235.

[42] Hoque, M. Z., Keskinarkaus, A., Nyberg, P., & Seppänen, T. (2024). Stain normalization methods for histopathology image analysis: A comprehensive review and experimental comparison. Information Fusion, 102, 101997.

[43] Janani, S., Marisuganya, R., & Nivedha, R. (2013). MRI image segmentation using stationary wavelet transform and FCM algorithm. International Journal of Computer Applications, 0975–8887, 1–6.

[44] Kehl, K. L., Elmarakeby, H., Nishino, M., Van Allen, E. M., Lepisto, E. M., Hassett, M. J., Johnson, B. E., & Schrag, D. (2019). Assessment of deep natural language processing in ascertaining oncologic outcomes from radiology reports. JAMA Oncology, 5(10), 1421–1429.

[45] Khanagar, S. B., Alkadi, L., Alghilan, M. A., Kalagi, S., Awawdeh, M., Bijai, L. K., Vishwanathaiah, S., Aldhebaib, A., & Singh, O. G. (2023). Application and performance of artificial intelligence (AI) in oral cancer diagnosis and prediction using histopathological images: A systematic review. Biomedicines, 11(6), 1612.

[46] Kumar, R., Srivastava, R., & Srivastava, S. (2015). Detection and classification of cancer from microscopic biopsy images using clinically significant and biologically interpretable features. Journal of Medical Engineering, 2015(1), 457906.

[47] Leonard, L. C. (2017). Web-based behavioral modeling for continuous user authentication (CUA). In Advances in Computers (Vol. 105, pp. 1–44). Elsevier.

[48] Ling, X., Alexander, G. S., Molitoris, J., Choi, J., Schumaker, L., Mehra, R., ... & Ren, L. (2023). Identification of CT-based non-invasive radiomic biomarkers for overall survival prediction in oral cavity squamous cell carcinoma. Scientific Reports, 13(1), 21774.

[49] Lu, C., Lewis Jr, J. S., Dupont, W. D., Plummer Jr, W. D., Janowczyk, A., & Madabhushi, A. (2017). An oral cavity squamous cell carcinoma quantitative histomorphometric-based image classifier of nuclear morphology can risk stratify patients for disease-specific survival. Modern Pathology, 30(12), 1655–1665.

[50] Macenko, M., Niethammer, M., Marron, J. S., Borland, D., Woosley, J. T., Guan, X., ... & Thomas, N. E. (2009, June). A method for normalizing histology slides for quantitative analysis. In 2009 IEEE international symposium on biomedical imaging: from nano to macro (pp. 1107-1110). IEEE.

[51] Maia, B. M. S., de Assis, M. C. F. R., de Lima, L. M., Rocha, M. B., Calente, H. G., Correa, M. L. A., ... & Krohling, R. A. (2024). Transformers, convolutional neural networks, and few-shot learning for classification of histopathological images of oral cancer. Expert Systems with Applications, 241, 122418.

[52] Markopoulos, A. K. (2012). Current aspects on oral squamous cell carcinoma. The Open Dentistry Journal, 6, 126–130.

[53] Marur, S., & Forastiere, A. A. (2008, April). Head and neck cancer: Changing epidemiology, diagnosis, and treatment. In Mayo Clinic Proceedings (Vol. 83, No. 4, pp. 489–501). Elsevier.

[54] Matias, A. V., Cerentini, A., Macarini, L. A. B., Amorim, J. G. A., Daltoé, F. P., & von Wangenheim, A. (2021). Segmentation, detection, and classification of cell nuclei on oral cytology samples stained with Papanicolaou. SN Computer Science, 2(4), 285.

[55] Mehlum, C. S., Larsen, S. R., Kiss, K., Groentved, A. M., Kjaergaard, T., Möller, S., & Godballe, C. (2018). Laryngeal precursor lesions: Interrater and intrarater reliability of histopathological assessment. The Laryngoscope, 128(10), 2375–2379.

[56] Mella, M. H., Chabrillac, E., Dupret-Bories, A., Mirallie, M., & Vergez, S. (2023). Transoral robotic surgery for head and neck cancer: Advances and residual knowledge gaps. Journal of Clinical Medicine, 12(6), 2303.

[57] Meyyappan, M., Verma, A., Kaushik, A., Sathyapriya, L., & Shanmugam, S. K. (2024, July). Oral Cancer detection using Histopathology Images. In 2024 Second International Conference on Advances in Information Technology (ICAIT) (Vol. 1, pp. 1-5). IEEE.

[58] Milas, Z., & Schellenberger, T. D. (2018). The Head and Neck Cancer Patient: Neoplasm Management, An Issue of Oral and Maxillofacial Surgery Clinics of North America, E-Book (Vol. 31, No. 1). Elsevier Health Sciences.

[59] Millar, E. K., Browne, L. H., Beretov, J., Lee, K., Lynch, J., Swarbrick, A., & Graham, P. H. (2020). Tumour stroma ratio assessment using digital image analysis predicts survival in triple negative and luminal breast cancer. Cancers, 12(12), 3749.

[60] Mohan, R., Rama, A., Raja, R. K., Shaik, M. R., Khan, M., Shaik, B., & Rajinikanth, V. (2023). OralNet: Fused optimal deep features framework for oral squamous cell carcinoma detection. Biomolecules, 13(7), 1090.

[61] Momtaz, W., Ali, S. S. A., Yasin, M. A. M., & Malik, A. S. (2018). A machine learning framework involving EEG-based functional connectivity to diagnose major depressive disorder (MDD). Medical & Biological Engineering & Computing, 56(2), 233–246.

[62] Musulin, J., Štifanić, D., Zulijani, A., Ćabov, T., Dekanić, A., & Car, Z. (2021). An enhanced histopathology analysis: An AI-based system for multiclass grading of oral squamous cell carcinoma and segmenting of epithelial and stromal tissue. Cancers, 13(8), 1784.

[63] Nawandhar, A., Kumar, N., & Yamujala, L. (2019, August). Performance analysis of neighborhood component feature selection for oral histopathology images. In 2019 PhD Colloquium on Ethically Driven Innovation and Technology for Society (PhD EDITS) (pp. 1–2). IEEE.

[64] Naughton-Rockwell, M. (2022). Using deep learning with satellite imagery to estimate deforestation rates. Environmental Modelling & Software, 153, 105388.

[65] Neves, M., & Ševa, J. (2021). An extensive review of tools for manual annotation of documents. Briefings in Bioinformatics, 22(1), 146–163.

[66] Oya, K., Kokomoto, K., Nozaki, K., & Toyosawa, S. (2023). Oral squamous cell carcinoma diagnosis in digitized histological images using convolutional neural network. Journal of Dental Sciences, 18(1), 322–329.

[67] Omar, E. A. (2013). The outline of prognosis and new advances in diagnosis of oral squamous cell carcinoma (OSCC): Review of the literature. Journal of Oral Oncology, 2013(1), 519312.

[68] Panigrahi, S., & Swarnkar, T. (2019, November). Automated classification of oral cancer histopathology images using convolutional neural network. In 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 1232–1234). IEEE.

[69] Panigrahi, S., Das, J., & Swarnkar, T. (2022). Capsule network based analysis of histopathological images of oral squamous cell carcinoma. Journal of King Saud University-Computer and Information Sciences, 34(7), 4546–4553.

[70] Peng, C., Liu, Y., Yuan, X., & Chen, Q. (2022). Research of image recognition method based on enhanced inception-ResNet-V2. Multimedia Tools and Applications, 81(24), 34345–34365.

[71] Qayyum, H., Majid, M., Anwar, S. M., & Khan, B. (2017). Facial expression recognition using stationary wavelet transform features. Mathematical Problems in Engineering, 2017, 1–12.

[72] Ragab, M., & Asar, T. O. (2024). Deep transfer learning with improved crayfish optimization algorithm for oral squamous cell carcinoma cancer recognition using histopathological images. Scientific Reports, 14(1), 25348.

[73] Rahman, T. Y., Mahanta, L. B., Chakraborty, C., Das, A. K., & Sarma, J. D. (2018). Textural pattern classification for oral squamous cell carcinoma. Journal of Microscopy, 269(1), 85–93.

[74] Rahman, T. Y., Mahanta, L. B., Das, A. K., & Sarma, J. D. (2020). Automated oral squamous cell carcinoma identification using shape, texture and color features of whole image strips. Tissue and Cell, 63, 101322.

[75] Rahman, T. Y., Mahanta, L. B., Choudhury, H., Das, A. K., & Sarma, J. D. (2020). Study of morphological and textural features for classification of oral squamous cell carcinoma by traditional machine learning techniques. Cancer Reports, 3(6), e1293.

[76] Rahman, A. U., Alqahtani, A., Aldhafferi, N., Nasir, M. U., Khan, M. F., Khan, M. A., & Mosavi, A. (2022). Histopathologic oral cancer prediction using oral squamous cell carcinoma biopsy empowered with transfer learning. Sensors, 22(10), 3833.

[77] Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 658–666).

[78] Rivera, C., & Venegas, B. (2014). Histological and molecular aspects of oral squamous cell carcinoma. Oncology Letters, 8(1), 7–11.

[79] Rich, J. T., Neely, J. G., Paniello, R. C., Voelker, C. C., Nussenbaum, B., & Wang, E. W. (2010). A practical guide to understanding Kaplan-Meier curves. Otolaryngology—Head and Neck Surgery, 143(3), 331–336.

[80] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18 (pp. 234–241). Springer International Publishing.

[81] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4510–4520).

[82] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual explanations from deep networks via gradient-based localization. International Journal of Computer Vision, 128, 336–359.

[83] Shetty, S., & Patil, A. P. (2023). Oral cancer detection model in distributed cloud environment via optimized ensemble technique. Biomedical Signal Processing and Control, 81, 104311.

[84] Sheu, R. K., & Pardeshi, M. S. (2022). A survey on medical explainable AI (XAI): recent progress, explainability approach, human interaction and scoring system. Sensors, 22(20), 8068.

[85] Shukla, R., Ajwani, B., Sharma, S., & Das, D. (2024, April). Identifying oral carcinoma from histopathological image using unsupervised nuclear segmentation. In 2024 IEEE 9th International Conference for Convergence in Technology (I2CT) (pp. 1–6). IEEE.

[86] Singh, T., & Vishwakarma, D. K. (2021). A deeply coupled ConvNet for human activity recognition using dynamic and RGB images. Neural Computing and Applications, 33(1), 469–485.

[87] Smit, M. A., Ciompi, F., Bokhorst, J. M., van Pelt, G. W., Geessink, O. G., Putter, H., ... & van der Laak, J. A. (2023). Deep learning based tumor–stroma ratio scoring in colon cancer correlates with microscopic assessment. Journal of pathology informatics, 14, 100191.

[88] Souza da Silva, R. M., Queiroga, E. M., Paz, A. R., Neves, F. F., Cunha, K. S., & Dias, E. P. (2021). Standardized assessment of the tumor-stroma ratio in colorectal cancer: Interobserver validation and reproducibility of a potential prognostic factor. Clinical Pathology, 14, 2632010X21989686.

[89] Suara, S., Jha, A., Sinha, P., & Sekh, A. A. (2023, November). Is Grad-CAM explainable in medical images? In International Conference on Computer Vision and Image Processing (pp. 124–135). Cham: Springer Nature Switzerland.

[90] Sujatha, K. B., Nageswari, D., Geethamahalakshmi, G., & Jayasankar, T. (2021). Classification of standard oral cancer using textural analysis and hybrid Hopfield neural networks. Turkish Journal of Physiotherapy and Rehabilitation, 32(2), 2811–2819.

[91] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians.

[92] Swersky, K., Snoek, J., & Adams, R. P. (2013). Multi-task Bayesian optimization. Advances in Neural Information Processing Systems, 26, 2004–2012.

[93] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

[94] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the AAAI conference on artificial intelligence (Vol. 31, No. 1).

[95] Štifanić, D., Musulin, J., Miočević, A., Baressi Šegota, S., Šubić, R., & Car, Z. (2020). Impact of covid-19 on forecasting stock prices: an integration of stationary wavelet transform and bidirectional long short-term memory. Complexity, 2020.

[96] Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In International Conference on Machine Learning (pp. 6105–6114). PMLR.

[97] Tan, Y., Wang, Z., Xu, M., Li, B., Huang, Z., Qin, S., ... & Huang, C. (2023). Oral squamous cell carcinomas: State of the field and emerging directions. International Journal of Oral Science, 15(1), 44.

[98] Tharwat, A. (2020). Classification assessment methods. Applied Computing and Informatics.

[99] Tlsty, T. D., & Coussens, L. M. (2006). Tumor stroma and regulation of cancer development. Annu. Rev. Pathol. Mech. Dis., 1(1), 119-150.

[100] Umaldi, N., Asari, V. K., & Rahman, Z. U. (2008, April). Fast and robust wavelet-based dynamic range compression with local contrast enhancement. In Visual Information Processing XVII (Vol. 6978, pp. 42-53). SPIE.

[101] Umapathy, V. R., Natarajan, P. M., Swamikannu, B., Jaganathan, S., Rajinikanth, S., & Periyasamy, V. (2024). Role of artificial intelligence in oral cancer. Advances in Public Health, 2024(1), 3664408.

[102] van Pelt, G. W., Kjær-Frifeldt, S., van Krieken, J. H. J., Al Dieri, R., Morreau, H., Tollenaar, R. A., ... & Mesker, W. E. (2018). Scoring the tumor-stroma ratio in colon cancer: Procedure and recommendations. Virchows Archiv, 473(4), 405–412.

[103] Wang, R., Naidu, A., & Wang, Y. (2021). Oral cancer discrimination and novel oral epithelial dysplasia stratification using FTIR imaging and machine learning. Diagnostics, 11(11), 2133.

[104] Warin, K., Limprasert, W., Suebnukarn, S., Jinaporntham, S., & Jantana, P. (2021). Automatic classification and detection of oral cancer in photographic images using deep learning algorithms. Journal of Oral Pathology & Medicine, 50(9), 911–918.

[105] Warnakulasuriya, S., Reibel, J., Bouquot, J., & Dabelsteen, E. (2008). Oral epithelial dysplasia classification systems: predictive 716 value, utility, weaknesses and scope for improvement. Journal of Oral Pathology & Medicine, 37(3), 127-133. DOI: 717 10.1111/j.1600-0714.2007.00584.x 718

[106] Werb, Z., & Lu, P. (2015). The role of stroma in tumor development. The Cancer Journal, 21(4), 250–253.

[107] Wetzer, E., Gay, J., Harlin, H., Lindblad, J., & Sladoje, N. (2020, April). When texture matters: Texture-focused CNNs outperform general data augmentation and pretraining in oral cancer detection. In 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI) (pp. 517–521). IEEE.

[108] Wu, Y., Cheng, M., Huang, S., Pei, Z., Zuo, Y., Liu, J., ... & Shao, W. (2022). Recent advances of deep learning for computational histopathology: Principles and applications. Cancers, 14(5), 1199.

[109] Wu, Y., Koyuncu, C. F., Toro, P., Corredor, G., Feng, Q., Buzzy, C., ... & Madabhushi, A. (2022). A machine learning model for separating epithelial and stromal regions in oral cavity squamous cell carcinomas using H&E-stained histology images: A multi-center, retrospective study. Oral Oncology, 131, 105942.

[110] Wu, I. C., Chen, Y. C., Karmakar, R., Mukundan, A., Gabriel, G., Wang, C. C., & Wang, H. C. (2024). Advancements in hyperspectral imaging and computer-aided diagnostic methods for the enhanced detection and diagnosis of head and neck cancer. Biomedicines, 12(10), 2315.

[111] Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., & Luo, P. (2021). SegFormer: Simple and efficient design for semantic segmentation with transformers. Advances in Neural Information Processing Systems, 34, 12077–12090.

[112] Xu, Y., & Goodacre, R. (2018). On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. Journal of Analysis and Testing, 2(3), 249–262.

[113] Zhang, R., Song, W., Wang, K., & Zou, S. (2017). Tumor-stroma ratio (TSR) as a potential novel predictor of prognosis in digestive system cancers: A meta-analysis. Clinica Chimica Acta, 472, 64–68.

[114] Zhang, D. (2019). Wavelet transform. In Fundamentals of Image Data Mining (pp. 35–44). Springer, Cham.

[115] Zhao, K., Li, Z., Yao, S., Wang, Y., Wu, X., Xu, Z., ... & Liu, Z. (2020). Artificial intelligence quantified tumour-stroma ratio is an independent predictor for overall survival in resectable colorectal cancer. EBioMedicine, 61, 103033.

[116] Zheng, Q., Jiang, Z., Ni, X., Yang, S., Jiao, P., Wu, J., ... & Liu, X. (2023). Machine learning quantified tumor-stroma ratio is an independent prognosticator in muscle-invasive bladder cancer. International Journal of Molecular Sciences, 24(3), 2746.

[117] Zoph, B., Vasudevan, V., Shlens, J., & Le, Q. V. (2018). Learning transferable architectures for scalable image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8697-8710).

[118] Zuraw, A., & Aeffner, F. (2022). Whole-slide imaging, tissue image analysis, and artificial intelligence in veterinary pathology: An updated introduction and review. Veterinary Pathology, 59(1), 6–25.

List of Figures

List of Tables

List of Abbreviations

ACC	Accuracy
AHA	Artificial Hummingbird Algorithm
AI	Artificial Intelligence
AL	Active Learning
AUC	Area Under the Curve
CAD	Computer-Aided Diagnosis Systems
CNN	Convolutional Neural Network
COE	Conventional Oral Examination
CV	Computer Vision
DFS	Disease-Free Survival
DL	Deep Learning
DWT	Discrete Wavelet Transform
ECM	Extracellular Matrix
FFPE	Formalin-Fixed Paraffin-Embedded
FN	False Negatives
FP	False Positives
FPR	False Positive Rate
Grade I	Well Differentiated Tumor
Grade II	Moderately Differentiated Tumor
Grade III	Poorly Differentiated Tumor
Grad-CAM	Gradient Weighted Class Activation Mapping
H&E	Hematoxylin And Eosin
IHC	Immunohistochemical
IOU	Intersection-Over-Union
KBC Rijeka	Clinical Hospital Center Rijeka

KNN	K-Nearest Neighbourhood
LBEC	Liquid-Based Exfoliative Cytology
LIME	Local Interpretable Model-Agnostic Explanations
LRP	Layer-Wise Relevance Propagation
ML	Machine Learning
MLP	Multilayer Perceptron
NLP	Natural Language Processing
NSCLC	Non-Small Cell Lung Cancer
OC	Oral Cancer
OCT	Optical Coherence Tomography
OD	Optical Density
OS	Overall Survival
OSCC	Oral Squamous Cell Carcinoma
SHAP	Shapley Additive Explanations
SCD	Stain Color Descriptor
SPCN	Structure Preserving Color Normalization
SVM	Support Vector Machines
SVD	Singular Value Decomposition
SWT	Stationary Wavelet Transform
TP	True Positives
TPR	True Positive Rate
TN	True Negatives
TNM	Tumor-Node-Metastasis
TSR	Tumor-Stroma Ratio
WT	Wavelet Transform
WSI	Whole Slide Imaging
XSAI	Explainable Artificial Intelligence
XDL	Explainable Deep Learning

Acknowledgment

This research was (partly) supported by;
Erasmus+ AISE, under grant 2023-1-494 EL01-KA220-SCH-000157157;
and by the CZI 'BrainClock' project under grant NPOO.C3.2.R3-495 I104.0089.

Curriculum Vitae

JELENA ŠTIFANIĆ

ABOUT

Jelena Štifanić, Master of Electrical Engineering, is an assistant at the Faculty of Engineering at the Catholic University of Croatia. She graduated in 2019 with a thesis focused on applying artificial intelligence to improve bladder cancer diagnostics. She gained international experience at Johannes Kepler University in Linz, which enhanced her interdisciplinary perspective and communication skills. Jelena is proficient in various programming languages and tools aa well as specialized analytical software. Her research focuses on digital image processing and the application of artificial intelligence in medicine, robotics, and engineering systems. Her work contributes to the development of advanced technological solutions and the improvement of automated medical diagnostics.

List of Selected Publications

Scientific papers published in journals

1. Musulin, J., Štifanić, D., Zulijani, A., Ćabov, T., Dekanić, A., & Car, Z. (2021). An enhanced histopathology analysis: An ai-based system for multiclass grading of oral squamous cell carcinoma and segmenting of epithelial and stromal tissue. Cancers, 13(8), 1784.
2. Štifanić, J., Štifanić, D., Anđelić, N., & Car, Z. (2025). Explainable AI for Oral Cancer Diagnosis: Multiclass Classification of Histopathology Images and Grad-CAM Visualization. Biology, 14(8), 909.

Scientific papers published as bookchapter

1. Štifanić, J., Štifanić, D., Zulijani, A., & Car, Z. (2022, May). Application of AI in histopathological image analysis. In Serbian International Conference on Applied Artificial Intelligence (pp. 121-131). Cham: Springer International Publishing.

Scientific papers published in conferences

1. Musulin, J., Štifanić, D., Zulijani, A., Šegota, S. B., Lorencin, I., Anđelić, N., & Car, Z. (2021, October). Automated grading of oral squamous cell carcinoma into multiple classes using deep learning methods. In 2021 IEEE 21st international conference on bioinformatics and bioengineering (BIBE) (pp. 1-6). IEEE.
2. Musulin, J., Štifanić, D., Zulijani, A., & Car, Z. (2021). Semantic segmentation of oral squamous cell carcinoma on epithellial and stromal tissue. In Book of proceedings 1st International Conference on Chemo and BioInformatics (ICCBIKG 2021) (pp. 194-197). Kragujevac: Institute for Information Technologies, University of Kragujevac.
3. Musulin, J., Štifanić, D., Zulijani, A., Šegota, S. B., Anđelić, N., Lorencin, I., ... & Car, Z. (2022). Histopathological H&E-Stained Image Analysis Based on AI. In First Serbian Conference on Applied Artificial Intelligence.

Appenices

Table A. Comparison of mean AUC_{macro} and -$_{micro}$ values of three different optimizers (SGD, ADAM, and RMSprop) on pre-trained model - MobileNetv2

MobileNetv2

	AUCmacro	AUCmicro
SGD	0,877	0,901
Adam	0,762	0,613
RMSprop	0,745	0,592

Table B. Comparison of mean AUC_{macro} and -$_{micro}$ values of three different optimizers (SGD, ADAM, and RMSprop) on pre-trained model – ResNet50

ResNet50

	AUCmacro	AUCmicro
SGD	0,822	0,788
Adam	0,871	0,864
RMSprop	0,833	0,832

Table C. Comparison of mean AUC_{macro} and $-_{micro}$ values of three different optimizers (SGD, ADAM, and RMSprop) on pre-trained model – ResNet101

ResNet101

	AUCmacro	AUCmicro
SGD	0,86	0,834
Adam	0,882	0,89
RMSprop	0,829	0,836

Table D. Comparison of mean AUC_{macro} and $-_{micro}$ values of three different optimizers (SGD, ADAM, and RMSprop) on pre-trained model - NasNet

NasNet

	AUCmacro	AUCmicro
SGD	0,845	0,869
Adam	0,89	0,909
RMSprop	0,849	0,854

Table E. Comparison of mean AUC_{macro} and $-_{micro}$ values of three different optimizers (SGD, ADAM, and RMSprop) on pre-trained model – InceptionResNet2

InceptionResNetv2

	AUCmacro	AUCmicro
SGD	0,807	0,823
Adam	0,92	0,931
RMSprop	0,914	0,917

Table F. Comparison of mean AUC_{macro} and $-_{micro}$ values of three different optimizers (SGD, ADAM, and RMSprop) on pre-trained model – IneptionV3

InceptionV3

	AUCmacro	AUCmicro
SGD	0,824	0,854
Adam	0,932	0,934
RMSprop	0,923	0,933

Table G. Comparison of mean AUC_{macro} and $-_{micro}$ values of three different optimizers (SGD, ADAM, and RMSprop) on pre-trained model – EfficientNetB3

EfficientNetB3

	AUCmacro	AUCmicro
SGD	0,751	0,796
Adam	0,911	0,915
RMSprop	0,902	0,898

Table H. Comparison of mean AUC_{macro} and $-_{micro}$ values of three different optimizers (SGD, ADAM, and RMSprop) on pre-trained model – Xception

Xception

	AUCmacro	AUCmicro
SGD	0,818	0,85
Adam	0,924	0,933
RMSprop	0,929	0,942

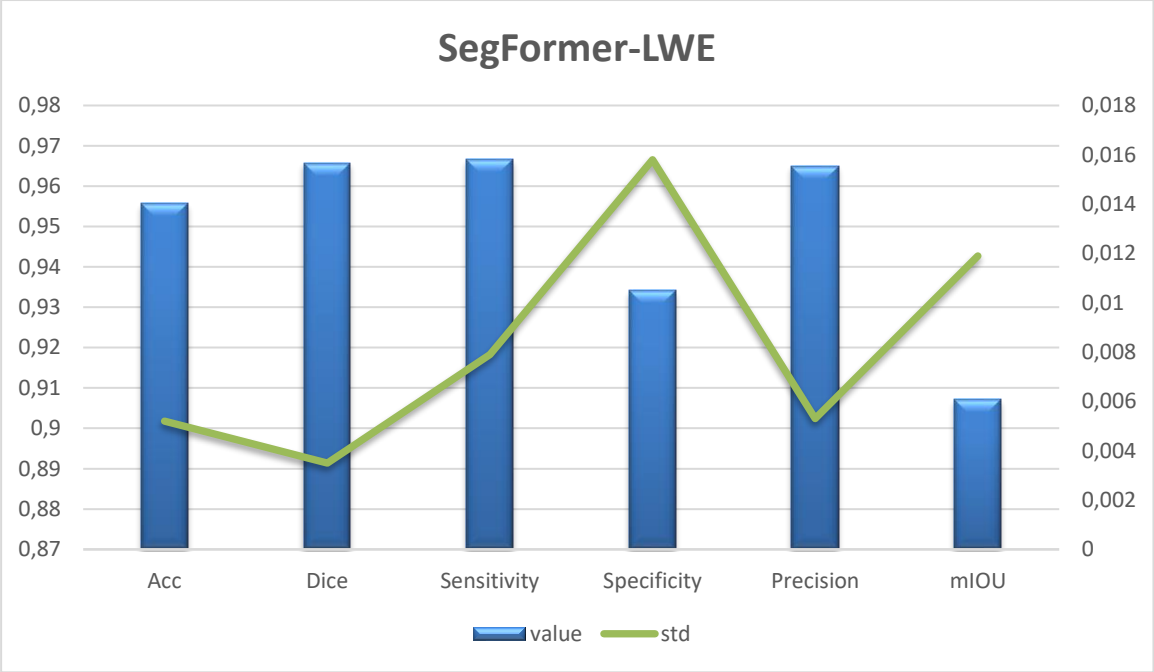

Figure A. Visual representation of SegFormer-LWE performance evaluation. Relevant segmentation metrics (Accuracy, Dice coefficient, Sensitivity, Specificity, Precision, and mIOU) are shown in bar charts with corresponding standard deviation shown in line plot.

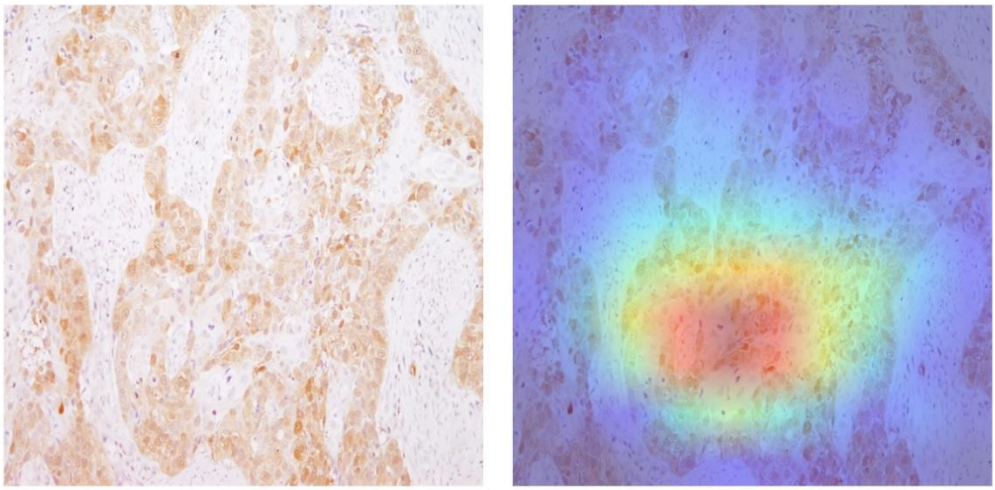

Figure B. An illustration of the Grad-CAM heatmap that highlights discriminative tissue regions that the proposed Xception–SWT model uses to grade histological images – patient 1 - PoC

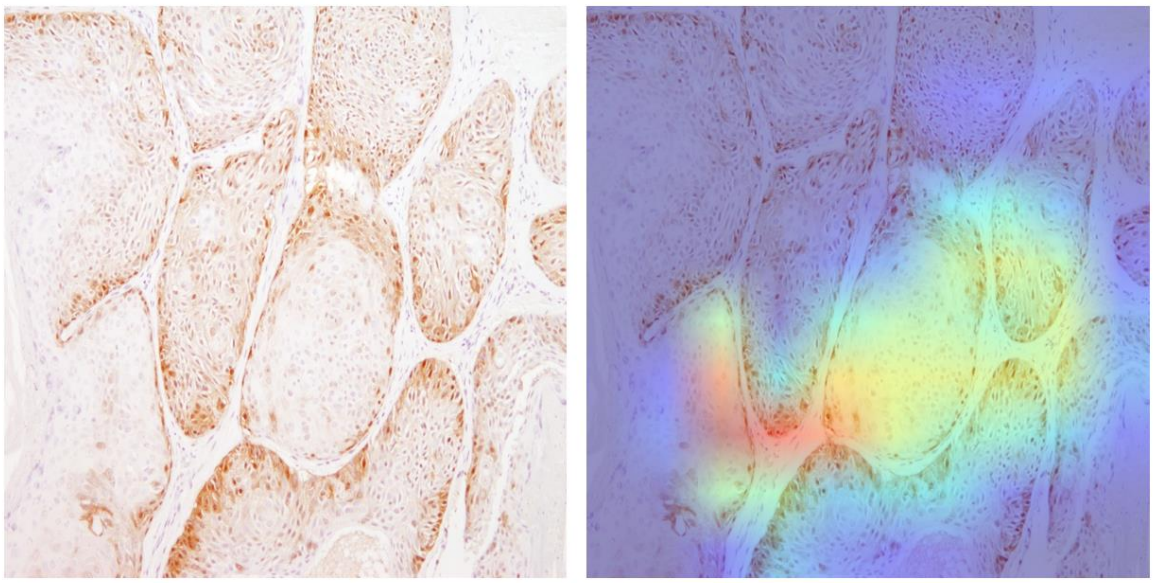

Figure C. An illustration of the Grad-CAM heatmap that highlights discriminative tissue regions that the proposed Xception–SWT model uses to grade histological images – patient 2 - PoC

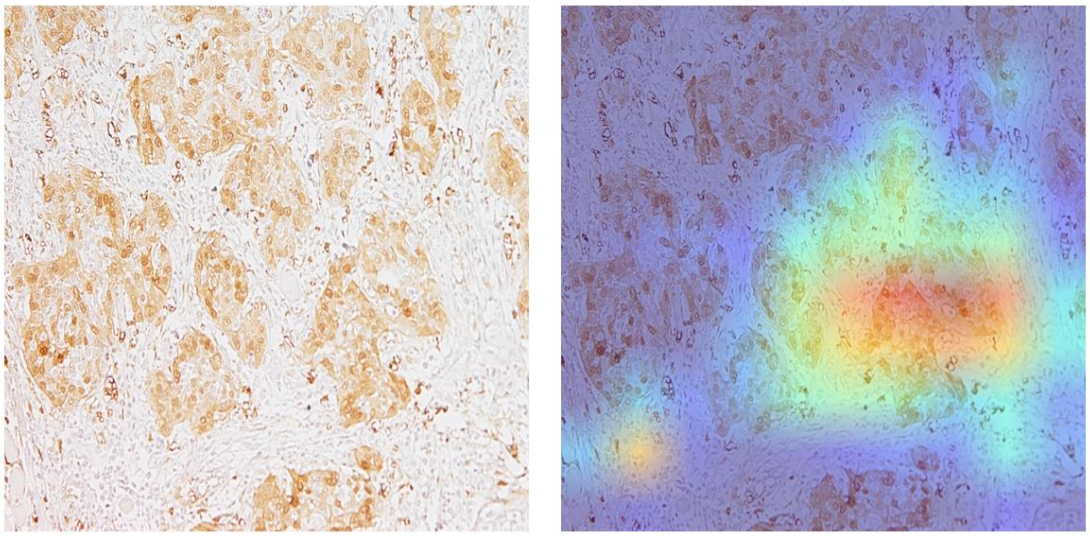

Figure D. An illustration of the Grad-CAM heatmap that highlights discriminative tissue regions that the proposed Xception–SWT model uses to grade histological images – patient 3 - PoC

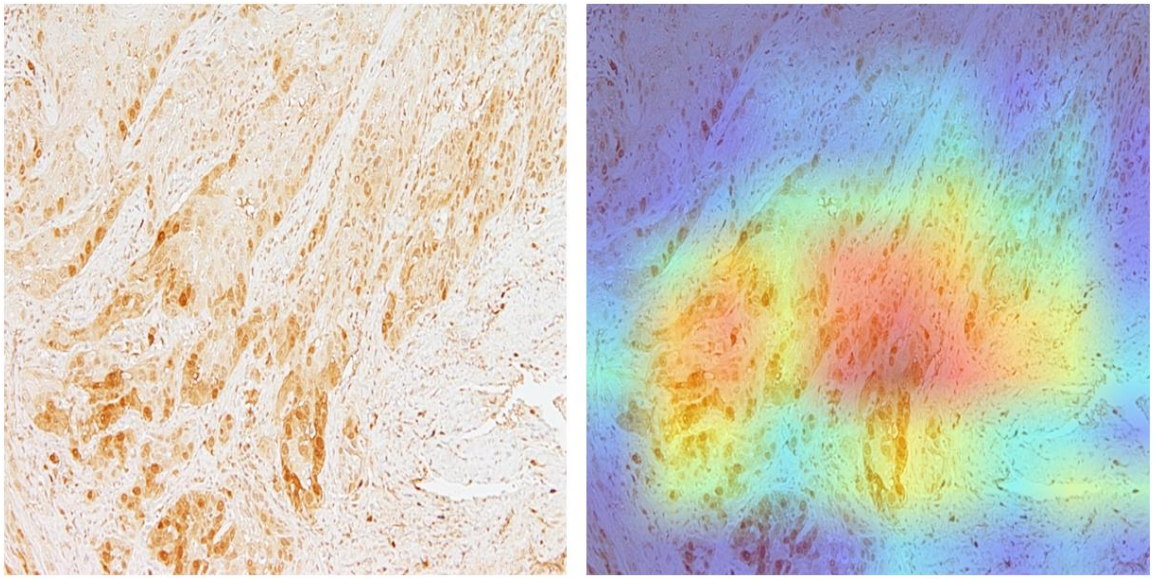

Figure E. An illustration of the Grad-CAM heatmap that highlights discriminative tissue regions that the proposed Xception–SWT model uses to grade histological images – patient 4 – PoC

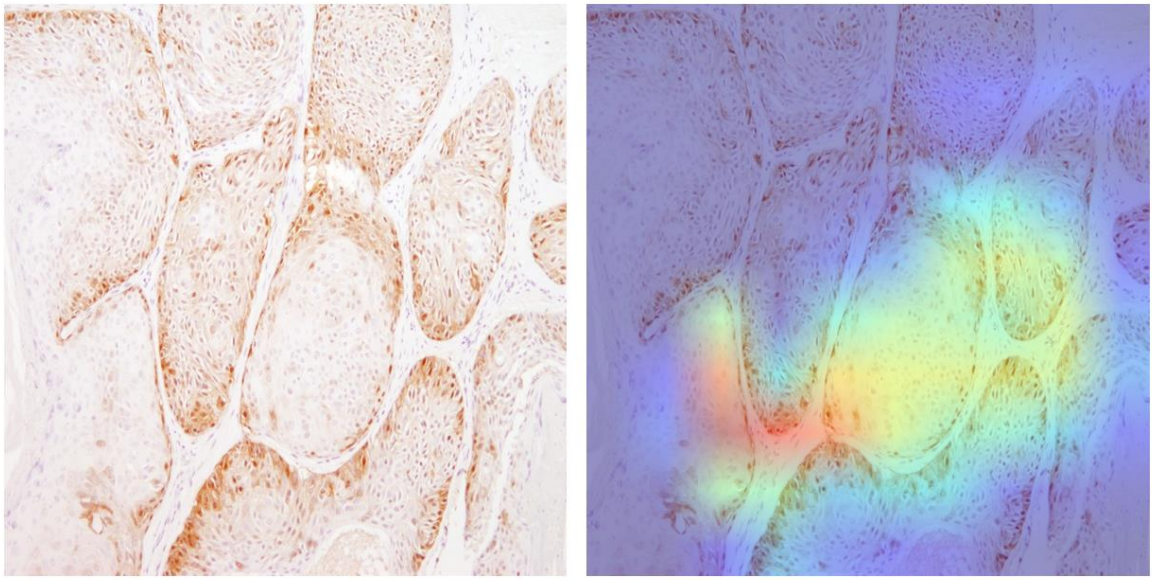

Figure F. An illustration of the Grad-CAM heatmap that highlights discriminative tissue regions that the proposed Xception–SWT model uses to grade histological images – patient 5 – PoC

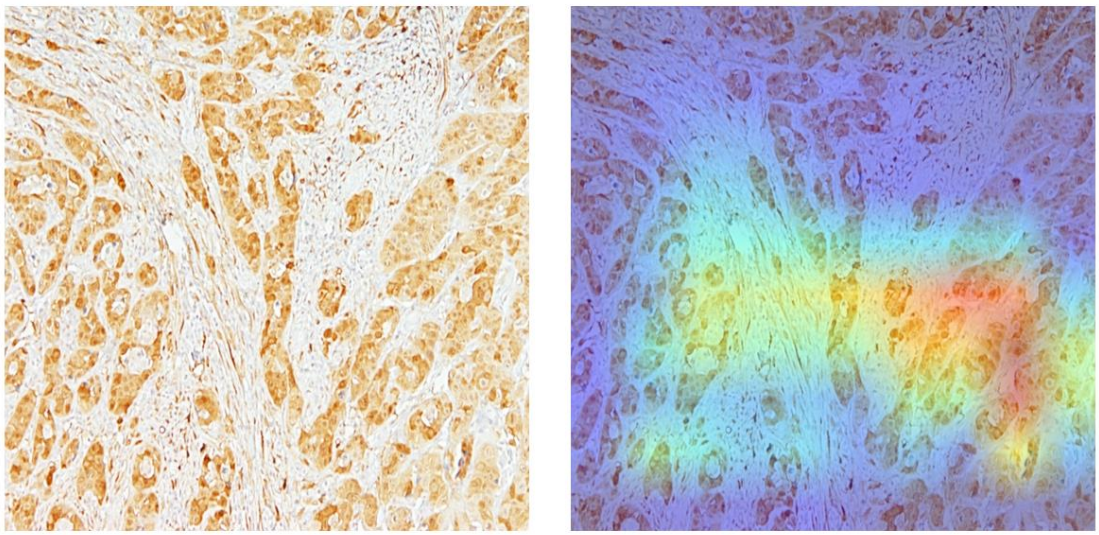

Figure G. An illustration of the Grad-CAM heatmap that highlights discriminative tissue regions that the proposed Xception–SWT model uses to grade histological images – patient 6 – PoC

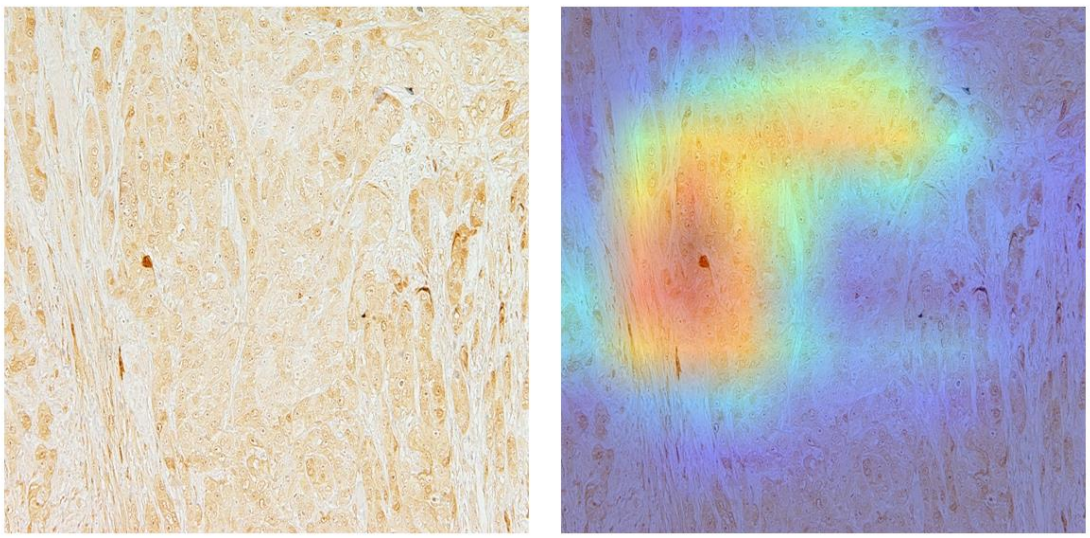

Figure H. An illustration of the Grad-CAM heatmap that highlights discriminative tissue regions that the proposed Xception–SWT model uses to grade histological images – patient 7 – PoC

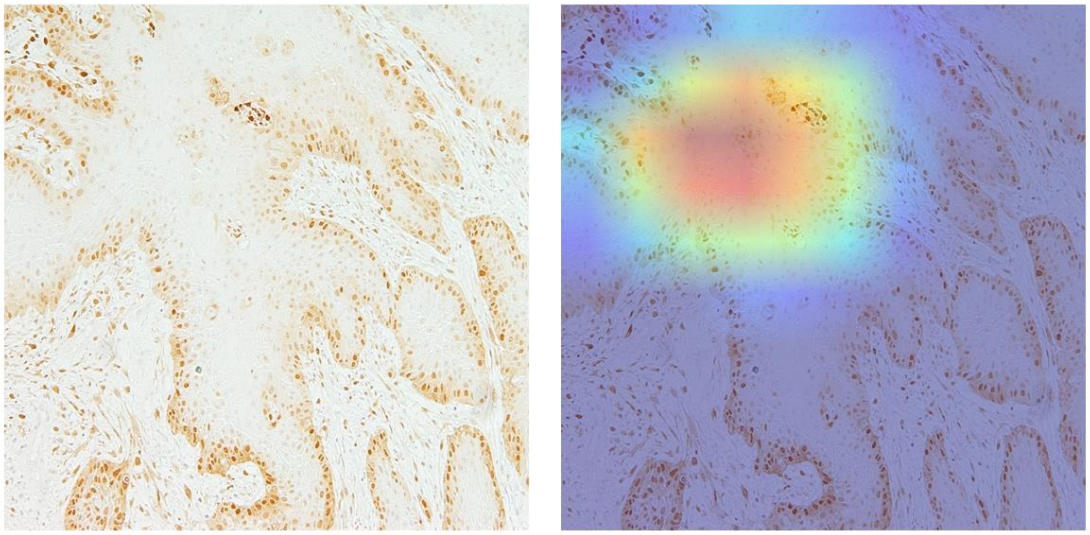

Figure I. An illustration of the Grad-CAM heatmap that highlights discriminative tissue regions that the proposed Xception–SWT model uses to grade histological images – patient 8 – PoC