ENHANCEMENT OF LUNG CANCER HISTOPATHOLOGICAL IMAGE CLASSIFICATION USING K-FOLD CROSS VALIDATION AND VAHANADE-BASED DIGITAL IMAGE PROCESSING PIPELINE

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

OF

ATILIM UNIVERSITY

MEHMET ÇAĞLAR VESEK

A MASTER OF SCIENCE THESIS IN THE DEPARTMENT OF SOFTWARE ENGINEERING

JANUARY, 2025

ENHANCEMENT OF LUNG CANCER HISTOPATHOLOGICAL IMAGE CLASSIFICATION USING K-FOLD CROSS VALIDATION AND VAHANADE-BASED DIGITAL IMAGE PROCESSING PIPELINE

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
ATILIM UNIVERSITY

BY

MEHMET ÇAĞLAR VESEK

IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR
THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF SOFTWARE ENGINEERING

JANUARY, 2025

Approval of the Graduate School of Natural and Applied Sciences, Atilim University.

_____

Prof. Dr. Ender Keskinkılıç
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of **Master of Science in Software Engineering, Atılım University**.

_____

Prof. Dr. Ali Yazıcı
Head of Department

This is to certify that we have read the thesis "Enhancement of Lung Cancer Histopathological Image Classification Using K-Fold Cross Validation and Vahanade-Based Digital Image Processing Pipeline" submitted by Mehmet Çağlar Vesek and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

_____

Asst. Prof. Dr. Erhan Gökçay
Supervisor

Examining Committee Members:

Asst. Prof. Dr. Erhan Gökçay
Software Engineering Department, Atılım University _____

Assoc. Prof. Dr. Çiğdem Turhan
Software Engineering Department, Atılım University _____

Assoc. Prof. Dr. Emre Sümer
Computer Engineering Department,
Başkent University _____

Date: *24.01.2025*

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name : Mehmet Çağlar Vesek

Signature :

ABSTRACT

ENHANCEMENT OF LUNG CANCER HISTOPATHOLOGICAL IMAGE CLASSIFICATION USING K-FOLD CROSS VALIDATION AND VAHANADE-BASED DIGITAL IMAGE PROCESSING PIPELINE

Vesek, Mehmet Çağlar
M.Sc., Department of Software Engineering
Supervisor : Asst. Prof. Dr. Erhan Gökçay

January 2025, 108 pages

Lung cancer, the leading cause of global cancer deaths, requires precise and efficient diagnostic solutions. This work introduces a novel preprocessing step that removes duplicates from the LC25000 dataset, reducing it to 14195 balanced images (4727/4744/4724) in three classes, which is a previously undocumented improvement. Initial classification using InceptionResNetV2 and ConViT-Small under 5-fold cross-validation (5 epochs) achieved almost perfect accuracy (≤6 errors), demonstrating unprecedented cost-effectiveness without hyperparameter tuning. ConViT-Small, selected due to its lower computational requirement, was further optimized via TIFF conversion and color normalization. Among the tested methods, Vahanade's transform outperformed Reinhard and hybrid techniques (e.g. DCT-DWT, CLAHE), achieving excellent accuracy (0 errors), Kappa, and MCC scores. By integrating dataset curation, lightweight deep learning, and robust image processing, this work advances high-accuracy lung cancer classification and provides scalable solutions for medical imaging.

Keywords: medical image classification, deep learning, transfer learning, digital image processing, lung cancer diagnosis, color normalization, LC25000 dataset, Vahanade method.

ÖZ

AKCİĞER KANSERİ HİSTOPATOLOJİK GÖRÜNTÜ SINIFLANDIRMASININ K-KAT ÇAPRAZ DOĞRULAMA VE VAHANADE TABANLI DİJİTAL GÖRÜNTÜ İŞLEME HATTI KULLANILARAK GELİŞTİRİLMESİ

Vesek, Mehmet Çağlar
Yüksek Lisans, Yazılım Mühendisliği Bölümü
Tez Yöneticisi : Dr. Öğr. Üyesi Erhan Gökçay

Ocak 2025, 108 sayfa

Küresel kanser ölümlerinin önde gelen nedeni olan akciğer kanseri, kesin ve etkili tanı çözümleri gerektirir. Bu çalışma, LC25000 veri setinden yinelenenleri kaldırarak, üç sınıfta 14195 dengeli görüntüye (4727/4744/4724) düşüren yeni bir ön işleme adımı sunar; bu daha önce belgelenmemiş bir iyileştirmedir. 5 katlı çapraz doğrulama (5 epoch) altında InceptionResNetV2 ve ConViT-Small kullanılarak yapılan ilk sınıflandırma, hiperparametre ayarlaması olmadan benzeri görülmemiş bir maliyet etkinliği göstererek neredeyse mükemmel doğruluk (≤6 hata) elde etti. Daha düşük hesaplama gereksinimi nedeniyle seçilen ConViT-Small, TIFF dönüştürme ve renk normalizasyonu yoluyla daha da optimize edildi. Test edilen yöntemler arasında, Vahanade'nin dönüşümü Reinhard ve hibrit tekniklerden (örn. DCT-DWT, CLAHE) daha iyi performans göstererek, mükkemel doğruluk (0 hata), Kappa ve MCC puanları elde etti. Veri seti iyileştirme, hafif derin öğrenme ve sağlam görüntü işlemeyi entegre ederek, bu çalışma yüksek doğruluklu akciğer kanseri sınıflandırmasını ilerletiyor ve tıbbi görüntüleme için ölçeklenebilir çözümler sunuyor.

Anahtar Kelimeler: tıbbi görüntü sınıflandırma, derin öğrenme, transfer öğrenme, sayısal görüntü işleme, akciğer kanseri teşhisi, renk normalizasyonu, LC25000 veri kümesi, Vahanade yöntemi.

To my grandmother, my fianceé and Mehmet Fuat Sunguroğlu

ACKNOWLEDGMENTS

First of all, I would like to thank Prof. Dr. Ali Yazıcı for the support he gave me throughout my Master's degree. If he had not helped me especially when I was having difficulty, I would not have been able to finish this master's degree and I would not have been able to complete this beautiful comprehensive thesis.

I would like to express my gratitude to my valuable supervisor Asst. Prof. Dr. Erhan Gökçay for all the support he gave me throughout this thesis.

I would also like to thank our jury members Assoc. Prof. Dr. Çiğdem Turhan and Assoc. Prof. Dr. Emre Sümer for their contributions and valuable opinions.

I am eternally grateful and thankful to my fianecé who inspired me and never withheld his support for 9 years, and to my dear Mehmet Fuat Sunguroğlu who I see as my second father without withholding his support during all the difficulties in Ankara.

I would also like to thank my sweet grandmother and my family who did not withhold their financial and moral support.

Publicly available datasets were analyzed in this study. These data can be found here: LC25000 (https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images) and Cancer Imaging Archive (https://www.cancerimagingarchive.net/).

Finally, I would also like to thank Asst. Prof. Dr. Güler Kalem for her support with whom we published a conference paper.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF SYMBOLS/ABBREVIATIONS

NSCLC	Non-Small Cell Lung Cancer
SCLC	Small Cell Lung Cancer
WSI	Whole-Slide Images
CT	Computed Tomography
MRI	Magnetic Resonance Imaging
LUAD	Lung Adenocarcinoma
LUSC/LSCC	Lung Squamous Cell Carcinoma
NC	Normal Cell
H&E	Hematoxylin and Eosin
ViTs	Vision Transformers
ConViT	The Convolution-Free Transformer
CNNs	Convolutional Neural Networks
RNNs	Recurrent Neural Networks
ROC AUC	Receiver Operating Characteristic Area Under the Curve
TIFF	Tagged Image File Format
JPEG	Joint Photographic Experts Group
CLAHE	Contrast Limited Adaptive Histogram Equalization
GHE	Global Histogram Equalization
SNMF	Sparse Non-Negative Matrix Factorization
SPCN	Structure-Preserving Color Normalization
OD	Optical Density
NMF	Non-Negative Matrix Factorization
SISR	Single Image Super-Resolution
ESRGAN	Enhanced Super-Resolution Generative Adversarial Networks
PSNR	Peak Signal to Noise Ratio
RRDB	Dual in Residual Dense Blocks
DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform
TIMM	PyTorch Image Models

AdamW	Adaptive Moment Estimation with Weight Decay
MCC	Matthew Correlation Coefficient

CHAPTER 1

INTRODUCTION

1.1 Introduction to Lung cancer

Lung cancer is among the primary causes of cancer-related mortality globally and presents significant health challenges due to subtype variations, such as Non-Small Cell Lung Cancer (NSCLC) and Small Cell Lung Cancer (SCLC) [1]. NSCLC constitutes approximately 85% of all lung cancer cases [2], and SCLC, which has poor prognosis, accounts for approximately 15% of all lung cancers [6]. According to the 2020 cancer data, lung cancer remains the leading cause of cancer mortality, with an estimated 1.8 million deaths (18%) [7]. Owing to the high mortality rate of lung cancer, early diagnosis is crucial to improve survival rates and treatment efficacy [3]. Treatment modalities may vary according to lung cancer subtype. For instance, in SCLC, chemotherapy is typically administered without surgical intervention because of the advanced stage of the cancer, whereas in NSCLC, surgical intervention may be applicable based on cancer stage, and chemotherapy is administered based on its subtypes [8].

1.2 Small Cell Lung Carcinoma (SCLC)

SCLC is a neuroendocrine tumor typically associated with smoking and presents with aggressive behavior. It often exhibits rapid doubling times and early dissemination to distant sites, contributing to its poor prognosis [18]. Due to its high mitotic rate, it is often detected in advanced stages, making surgical resection rarely an option for curative treatment [18]. Histologically, SCLC cells are small, round, and densely packed, with minimal cytoplasm and absent nucleoli. Their morphology reflects neuroendocrine differentiation, as evidenced by markers such as chromogranin and synaptophysin [18]. Despite the grave prognosis, due to its neuroendocrine origin, SCLC responds well to chemotherapy and radiation therapy initially, but recurrences are common [18].

1.3 Non Small Cell Lung Carcinoma (NSCLC)

NSCLC constitutes the remaining 85% of lung cancers, with adenocarcinoma being the most prevalent subtype [16]. NSCLC progresses more gradually than SCLC and is subdivided into several histological categories, including adenocarcinoma, squamous cell carcinoma, and large cell carcinoma [18]. Adenocarcinoma has recently surpassed squamous cell carcinoma as the predominant subtype, particularly among non-smokers and women. It frequently originates in the peripheral regions of the lungs and is more likely to be detected at an early stage, thereby increasing the potential for surgical resection and targeted therapies [16]. The clinical and molecular classification of NSCLC is crucial for determining the appropriate course of treatment, including the utilization of tyrosine kinase inhibitors (TKIs) for tumors harboring mutations such as EGFR or ALK [18].

Pulmonary adenocarcinoma, a subtype of NSCLC, has undergone significant revisions in its classification to better align with advances in molecular biology and radiology. The 2015 World Health Organization (WHO) classification now categorizes invasive adenocarcinomas by their predominant histologic growth patterns, which include lepidic, acinar, papillary, micropapillary, and solid structures. These classifications not only aid in diagnosis but also carry prognostic implications [17].

The lepidic growth pattern refers to tumor cells proliferating along pre-existing alveolar structures without invasion into the underlying stroma. This pattern was previously termed "bronchioloalveolar carcinoma" but is now reclassified under the broader term of adenocarcinoma. Lepidic-predominant adenocarcinoma is generally associated with a better prognosis when compared to other invasive subtypes [18]. It is typically nonmucinous, and the tumor cells are bland, with minimal cytological atypia. Radiologically, these tumors often appear as ground-glass opacities, correlating with the non-invasive lepidic growth seen in histology [17].

Acinar adenocarcinoma predominantly consists of well-organized glandular structures, characterized by round, small gland-like formations, known as acini, often with clear or vacuolated cytoplasm that may contain mucin.

This subtype may include a range of configurations, from simple, well-formed glands to more intricate structures, including cribriform patterns, which are typically more invasive than lepidic growths still tend to respond well to early intervention but associated with a less favorable prognosis. Acinar adenocarcinomas are known for their association with genetic alterations such as EGFR mutations, which render them potential candidates for targeted therapies [16].

Papillary adenocarcinoma is typified by tumor cells growing along fibrovascular cores. This subtype often exhibits a complex architectural pattern with well-defined fibrovascular structures surrounded by layers of neoplastic cells. It is generally more invasive compared to the lepidic pattern but still shows intermediate aggressiveness [18]. Papillary adenocarcinomas may present with significant structural complexity and are often associated with a worse prognosis when compared to less invasive forms. However, papillary adenocarcinoma remains an important subtype in understanding lung cancer, especially in the context of emerging targeted therapies [17].

The solid pattern of adenocarcinoma is predominantly characterized by sheets or nests of tumor cells with little or no glandular differentiation. Solid adenocarcinomas are more aggressive and often associated with worse outcomes due to their rapid growth and higher likelihood of metastasis [18]. Histologically, the cells within this pattern often exhibit nuclear pleomorphism and prominent nucleoli, contributing to its classification as a high-grade malignancy. Solid adenocarcinomas are also associated with intratumoral mucin, and immunohistochemical staining is crucial for differentiating them from other solid tumors [17].

Micropapillary adenocarcinoma is a distinctive and aggressive subtype, characterized by small clusters of tumor cells that float within alveolar spaces without fibrovascular cores. This pattern is strongly associated with poor prognosis, due to its high propensity for lymphatic invasion and distant metastasis. The micropapillary structure's ability to invade the surrounding stroma makes it one of the most challenging subtypes to treat effectively [18].

Invasive mucinous adenocarcinoma, previously known as mucinous bronchioloalveolar carcinoma, is distinguished by the presence of tumor cells that produce abundant mucin. This subtype tends to have a diffuse growth pattern and is often multifocal, which can complicate surgical resection. These tumors frequently exhibit mutations in genes such as KRAS, which are less responsive to targeted therapies that are effective against other types of adenocarcinoma, making treatment more challenging [16].

1.4 Lung Cancer Detection Methods

Various methods can be used to detect lung cancers. Advanced imaging techniques play a particularly significant role in definitive diagnosis and categorization. X-rays, whole-slide images (WSI), computed tomography (CT) scans, and magnetic resonance imaging (MRI) are frequently used in the classification and diagnosis of lung cancer [4]. Additionally, certain deep learning models such as ResNet-50, EfficientNet-B3, and ResNet-101 offer innovative approaches to ameliorate disease and reduce mortality [5]. The application of these analytical methods and advanced technologies in the field of image analysis for lung cancer aims not only to enhance accuracy, but also to optimize treatment planning and patient outcomes, underscoring its critical importance in modern healthcare.

Progress in lung cancer image evaluation has leveraged a variety of imaging methodologies, including whole-slide imaging (WSI), Computed Tomography (CT), and Magnetic Resonance Imaging (MRI), which are instrumental for diagnosis and research applications. WSI, also known as virtual microscopy, involves the production and digitization of digital slides and is used in procedures such as research, diagnosis, and education [9]. Although CT is an effective method for determining tumor spread and location, enlarged lymph nodes, and metastatic disease, CT-integrated PET can combine information from CT with metabolic information [10]. MRI is a versatile imaging technique employed in lung cancer to distinguish between solid and vascular hilar masses and from diaphragmatic abnormalities to detecting mediastinal lymphoma [11].

Advancements in artificial intelligence (AI)-based approaches, such as machine learning and deep learning, facilitate the detection and classification of lung cancer. Several methodologies have been applied in machine learning. For instance, Machine Learning methods such as Random Forest (RF), XGBoost, Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), Naïve Bayes, and Stochastic Diffusion Search (SDS) that aid in feature selection to enhance the performance of ML can be utilized in lung cancer classification [12]. Furthermore, in Deep Learning methods, AlexNet, VGG16, and VGG19 networks with an optimization technique based on a Genetic Algorithm (GA) can be employed for lung nodule classification and early detection of these nodules in low-dose computed tomography (LDCT) [13].

1.5 Software Tools in Lung Cancer Imaging Analysis

Advancements in lung cancer imaging analysis have progressed through the utilization of various software tools and frameworks, with platforms such as Python and MATLAB becoming essential in this domain. Python and MATLAB are two highly regarded programming platforms in the field of medical imaging, particularly in lung cancer diagnosis, owing to their robust support for machine-learning and image-analysis tools. Python, with its versatility and extensive library ecosystem, including TensorFlow, Keras, and PyTorch, is gaining prominence for implementing complex deep-learning models [14]. According to a survey [14], these platforms are instrumental in leveraging convolutional neural networks (CNNs) and other deep learning models to achieve more accurate results in cancer classification tasks, particularly for modalities such as CT scans, MRIs, and histopathological images. Furthermore, Python's adaptability to integration with various AI-based approaches renders it a powerful tool for researchers and clinicians [15].

MATLAB is renowned for its specialized capabilities in image processing and quantitative analysis, making it a preferred choice for many medical imaging communities. Its built-in functions in signal processing and image segmentation are particularly valuable for analyzing CT scans and other imaging methods used in lung cancer diagnosis [15]. These platforms enable the execution of complex tasks such as image segmentation, feature extraction, and nodule classification. Researchers can apply CNN architectures, such as U-Net and ResNet, to automate and enhance

diagnostic processes. MATLAB's extensive toolboxes facilitate the implementation of advanced deep learning techniques, including convolutional neural networks (CNNs) and support vector machines (SVMs) [15].

These software environments have become integral to lung cancer research because they enable the utilization of complex deep learning algorithms, such as stacked sparse autoencoders and iterative neural networks, that surpass traditional machine learning approaches in terms of both accuracy and efficiency [14]. As machine learning continues to evolve, tools such as Python and MATLAB facilitate groundbreaking advancements in early tumor detection and classification, significantly improving the precision of lung cancer diagnosis. Furthermore, the integration of AI-driven methods, as outlined [15], underscores the importance of computational models that automate and enhance diagnostic workflows.

1.6 Study's Explanation and Contribution

In this study, firstly, the definition of lung cancer, its subtypes, tools, and artificial intelligence models used are given in the introduction section.

The focus of this study is to classify histopathological images of lung cancer in the most efficient way. Then, in the second section, important studies recently conducted in the field are mentioned. It focuses on the models and differences in the studies conducted with the LC25000 dataset. In the third section, repeated images that were not performed in previous studies are eliminated and digital image processing techniques that will make the classification more efficient are explained and then their application in the study is explained. In the fourth section, deep learning techniques such as transfer learning and key fold cross validation are explained and how the classification is done in the study is mentioned. In the fifth section, the findings of the study are shared and a literature comparison is made. In the last section, the summary of the study and future studies are mentioned.

CHAPTER 2

LITERATURE REVIEW

Luo et al. [19] introduced a model based on pathology image analysis to forecast the outcome of patients with lung adenocarcinoma (ADC), with validation across several independent cohorts. Morphological characteristics were extracted from H&E-stained sections of FFPE tumor samples. The model was constructed using pathology images from 91 stage I ADC patients at the Chinese Academy of Medical Sciences (CAMS) and independently verified using ADC patients from the National Lung Screening Trial (NLST) and University of Texas Special Program of Research Excellence (SPORE) cohorts.

The morphological features derived from CAMS cohort training were found to be crucial in predicting overall survival in both the NLST and SPORE cohorts. The predictive model classified patients into high- and low-risk categories, revealing significantly worse survival outcomes for the high-risk group in both test cohorts. This study confirmed the value of pathology images in cancer diagnosis and prognosis, suggesting that such analyses could be incorporated into clinical practice.

A study by Wang et al. [20] introduced a novel approach called multi-scale pathology image texture signature (MPIS) to forecast the outcome of patients with operable lung adenocarcinoma (LUAD). The research encompassed patients from four separate cohorts and employed an automated system to extract texture features from H&E-stained whole-slide images at various magnifications. MPIS was developed using the most distinctive texture features identified using the LASSO method. Its predictive capability for overall survival (OS) was assessed in an initial set and three external validation cohorts.

The findings revealed that a higher MPIS was correlated with significantly poorer OS across all cohorts. Combining MPIS with clinicopathological variables enhances the prognostic classification of patients with LUAD. Furthermore, this investigation examined the relationships between texture features and biological pathways, uncovering significant associations with processes such as cytokine activity and

extracellular matrix structure. This study suggests that MPIS serves as a robust and interpretable biomarker, potentially improving personalized post-surgery care for patients with LUAD.

A deep learning system for classifying histologic patterns in lung adenocarcinoma and predicting tumor mutation burden (TMB) status using Hematoxylin and Eosin (H&E)-stained whole-slide images was developed by Sadhwani et al. [21]. They trained a convolutional neural network (CNN) to identify histologic features across the whole slide images, achieving a patch-level area under the receiver operating characteristic curve (AUC) of 0.78-0.98 for nine histologic features. This model was combined with the clinical and demographic information to create an interpretable TMB classification model. The system's performance was evaluated in 172 cases from The Cancer Genome Atlas (TCGA), achieving an AUC of 0.71 for TMB classification.

The study also evaluated a weakly supervised deep learning model that relied solely on case-level TMB data, which demonstrated comparable performance with an AUC of 0.72. These results indicate that incorporating histological patterns in biomarker prediction offers substantial advantages, and interpretable approaches utilizing these patterns can perform similarly to less interpretable models.

Talib et al. [22] introduce two innovative approaches for detecting lung cancer using histopathological images: TransSegNet for segmentation and MinClassNet for classification. TransSegNet employs transformer-based semantic segmentation by dividing the input images into patches and processing them through four transformer blocks. The model was trained with specific hyperparameters (Adam optimizer, 10 epochs, batch size of 4) and achieved 99.62% detection accuracy on the ACDC-LungHP dataset, with 49.81% mean Intersection over Union (IoU) and 99.02% validation accuracy.

MinClassNet, designed for classification, comprises seven layers: convolution, ReLU, and Softmax components. Trained using an Adam optimizer over 10 epochs with a batch size of eight, MinClassNet differentiates between healthy and abnormal cells with 98.39% accuracy on a publicly accessible histopathological lung cancer dataset. Both models demonstrated substantial improvements compared to existing methods, highlighting the efficacy of transformer-based and deep learning techniques in lung cancer detection.

Coudray et al. [23] present a deep learning-based approach using convolutional neural networks (CNNs) to classify non-small cell lung cancer (NSCLC) subtypes and predict mutations from histopathology images. They utilized the Inception v3 architecture to differentiate between lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), and normal lung tissues using whole-slide images (WSIs) obtained from The Cancer Genome Atlas (TCGA). The model achieved an average area under the curve (AUC) of 0.97, which is comparable to that of expert pathologists. Validation was performed using independent datasets, including frozen sections, formalin-fixed paraffin-embedded (FFPE) tissues, and biopsies.

Additionally, the CNN was trained to predict the mutational status of ten commonly mutated genes in LUAD, such as EGFR, KRAS, and TP53, using only the WSIs. The model successfully identified six gene mutations with AUCs ranging from 0.733 to 0.856. This study demonstrated the potential of deep learning models to assist pathologists in diagnosing lung cancer subtypes and predicting genetic mutations, which could lead to more personalized treatment strategies.

Ji et al. [24] developed a novel automated system for classifying lung and colon cancer utilizing histopathological images. The system was constructed using the LC25000 dataset, which comprises images of both lung and colon tissues. The core model of the system is based on Swin Transformer V2, which integrates self-attention mechanisms with a hierarchical structure to enhance the performance on smaller datasets. The system was evaluated using five-fold cross-validation and compared with various convolutional neural networks (CNNs) and other vision transformers (ViTs). The Swin Transformer V2 model achieved optimal results in all performance metrics, including accuracy, precision, recall, and F1 score, surpassing the other models utilized in this study and previous investigations.

The authors posit that while the Swin Transformer V2 demonstrates the potential to assist pathologists in cancer classification, the LC25000 dataset may be insufficient for clinical application owing to its limitations, such as a lack of diverse cancer subtypes and real-world image complexities. They proposed the development of a new and more comprehensive dataset for further research and validation.

Afshar et al. [25] introduce a novel deep learning-based radiomics model, DRTOP, designed to predict clinical outcomes for lung cancer patients based on time-to-event

data. In contrast to traditional handcrafted radiomics, which relies on predefined features and accurate tumor segmentation, DRTOP utilizes raw imaging data directly as input. The model was trained using a dataset of 132 lung cancer patients by employing both the CT and PET components of PET/CT scans. Two parallel convolutional neural networks (CNNs) were utilized to analyze CT and PET data separately, which were subsequently combined with clinical parameters (such as age, sex, SUV, and radiation dose) using a Cox proportional hazards model (PHM). The objective of this study was to predict overall survival (OS), recurrence-free survival (RFS), local control (LC), and distant control (DC).

The results demonstrate that DRTOP exhibits superior performance to handcrafted radiomics in predicting OS, achieving a concordance index of 68% compared to 51% for handcrafted methods. While both methods performed comparably for DC prediction, DRTOP did not significantly enhance RFS prediction. The model demonstrates the potential to provide complementary information for personalized management of lung cancer patients, offering a more automated and less subjective approach than hand-crafted radiomics.

Zheng et al. [26] developed a hybrid deep learning model to predict the two-year overall survival (OS) of patients with stage I-IIIA non-small cell lung cancer (NSCLC) who received definitive radiotherapy. The model integrates clinical data, such as age and clinical stage, with imaging features extracted from pretreatment CT scans. This study utilized data from 270 NSCLC patients treated with radiotherapy at the University Medical Center Groningen (UMCG) and validated the model externally on 228 patients from the Maastro clinic.

The results indicated that combining clinical variables with imaging features led to improved prognostic performance. The hybrid model achieved a median AUC of 0.76 [95% CI: 0.65–0.86] on the UMCG test set and 0.64 [95% CI: 0.58–0.70] on the Maastro test set, both demonstrating significant differentiation between low- and high-risk mortality groups. This suggests that the hybrid approach could effectively identify patients with a higher mortality risk, potentially facilitating personalized treatment planning.

Xu et al. [27] investigated the application of deep learning models for predicting clinical outcomes in patients with locally advanced non-small cell lung cancer

(NSCLC) utilizing serial computed tomography (CT) scans obtained before and after treatment. The study comprised 179 patients with stage III NSCLC who underwent chemoradiation therapy, employing pre-treatment and follow-up CT scans at 1, 3, and 6 months post-treatment. Researchers have integrated convolutional neural networks (CNNs) with recurrent neural networks (RNNs) through transfer learning to analyze time-series data.

The findings demonstrated that the deep learning models significantly predicted overall survival and other cancer-specific outcomes, including progression, distant metastasis, and locoregional recurrence. The predictive accuracy of the models improved with the incorporation of each additional follow-up scan, achieving an area under the curve (AUC) of 0.74 for 2-year overall survival. Moreover, the models effectively stratified patients into high- and low-risk groups and exhibited a strong association with survival outcomes. The study concluded that deep learning methods could enhance clinical decision-making by providing noninvasive, cost-effective radiomics biomarkers for monitoring treatment response.

Wang et al. [28] developed a prognostic model to predict survival outcomes in non-small cell lung cancer (NSCLC) patients using computed tomography (CT) radiomics. This study retrospectively analyzed data from 173 patients with NSCLC, aiming to forecast the three-year survival limit. Lung tumors were segmented from the CT scans, and radiomics features were extracted and optimized using a feature weighting algorithm. These features were subsequently utilized to train several machine-learning classifiers to create a model capable of predicting a patient's survival time range.

The model demonstrated an accuracy of 88.7% during cross-validation, and achieved a high prediction accuracy of 79.6% when tested on an independent dataset. The study identified key prognostic radiomics features, such as inverse different moments, lobulation signs, and angular second moments, which were significantly associated with survival outcomes. These findings suggest that CT radiomics can effectively support clinicians in making more accurate survival predictions and optimizing postoperative treatment plans for NSCLC patients, ultimately improving patient outcomes.

Yang et al. [29] developed a deep learning-based method to classify different subtypes of lung cancer, including large cell neuroendocrine carcinoma (LCNEC), from biopsy

slides using convolutional neural networks (CNNs). The investigation evaluated several models, including four pre-trained CNNs—ResNet152, VGG19, Xception, and NASNetLarge—alongside a newly developed CNN constructed from scratch. The models were assessed on a dataset comprising whole slide images (WSIs) of lung cancer biopsy samples segmented into smaller image patches.

The results indicated that the CNN model constructed from scratch performed comparably to the best pre-trained models, achieving a macro average area under the curve (AUC) of 0.90 for distinguishing all subtypes, including LCNEC. When excluding LCNEC from the classification, the model's performance improved, achieving an AUC of 0.97, and for three main categories—non-small cell lung cancer (NSCLC), small cell lung cancer (SCLC), and non-tumor tissues—the AUC reached 0.96. These findings suggest that a custom-built CNN can achieve similar or superior accuracy compared with pre-trained models for specific tasks, such as classifying lung cancer subtypes.

Aonpong et al. [30] propose a genotype-guided radiomics (GGR) approach to enhance the prediction accuracy of cancer recurrence in non-small cell lung cancer (NSCLC) patients utilizing computed tomography (CT) images and gene expression data. This investigation employed a public dataset comprising both CT scans and gene expression profiles. The GGR methodology incorporates two models: the first estimates gene expression from radiomic features and deep learning-derived features extracted from CT images, whereas the second utilizes these estimated gene expressions to predict recurrence. This hybrid approach, which combines handcrafted and deep learning-based features, demonstrated a significant improvement in prediction accuracy, achieving 83.28% compared to 78.61% with existing radiomics methods and 79.09% with a deep learning model (ResNet50).

Researchers emphasize that by employing the GGR method, the necessity for invasive procedures to obtain gene expression data during the testing phase is eliminated, rendering it a more cost-effective and accessible approach. This methodology could potentially assist clinicians in predicting recurrence more accurately, facilitating timely intervention and improving the management of patients with NSCLC.

Huang et al. [31] developed an Integrated Deep Learning Evaluation (IDLE) score to predict the risk of lung cancer recurrence in stage IA non-small-cell lung cancer

(NSCLC) patients following complete surgical resection of the tumor. This study integrated features from preoperative low-dose CT scans with histological characteristics from hematoxylin and eosin (H&E)-stained tissue sections. A cohort of 182 patients with stage IA NSCLC from the National Lung Screening Trial was evaluated using the IDLE score, which incorporates both preoperative imaging and postoperative pathology data.

The findings indicated that the IDLE score achieved a five-year area under the curve (AUC) of 0.817, significantly outperforming traditional TNM staging and tumor grade, which exhibited AUCs of 0.561 and 0.573, respectively. The enhanced predictive accuracy of the IDLE score was attributed to the integration of tumor characteristics from both CT imaging and microscopic tissue features, suggesting that combining data from multiple platforms provides a more robust prediction of cancer recurrence than utilizing similar types of measurements in isolation.

Pham et al. [32] developed a two-step deep learning method to enhance the detection of lymph node metastases in lung cancer utilizing whole-slide histopathologic images. The investigation employed 349 digital slides, comprising 233 slides for training, 10 for validation, and 106 for testing. The initial step of the approach involved a deep learning model that targeted noncancerous regions frequently misidentified as tumors, such as lymphoid follicles, to eliminate false positives. Subsequently, a specialized deep learning classifier was employed to identify cancer cells.

The two-step method significantly reduced false-positive rates by an average of 36.4% and up to 89% in cases with reactive lymphoid follicles. Furthermore, it achieved 100% sensitivity for detecting macrometastases, micrometastases, and isolated tumor cells (ITCs). By incorporating thresholds of 0.6 mm and 0.7 mm for foci size, the sensitivity and specificity of the model were further refined to 79.6% and 96.5%, and 75.5% and 98.2%, respectively. This approach provides an effective strategy for identifying lymph node metastases with high accuracy and minimal false positives, rendering it a valuable tool for clinical pathology workflows.

Ding et al. [33] investigated the spatial and molecular characteristics of tumor-infiltrating lymphocytes (TILs) in non-small cell lung cancer (NSCLC), specifically comparing lung adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC). The study analyzed over 1,000 H&E-stained images of LUAD (421) and LUSC (438)

tumors, revealing that the density and spatial patterns of TILs differed significantly between these two subtypes. In LUAD, higher TIL density was associated with improved overall survival (OS), whereas in LUSC, the spatial arrangement of TILs was more predictive of patient outcomes.

Further analysis of external validation datasets demonstrated that the LUAD-specific TIL signature, comprising primarily CD4+ and CD8+ T cells, was prognostic of OS and predictive of response to therapy in NSCLC patients treated with various chemotherapy regimens. Conversely, LUSC-specific TIL patterns included CD4+ T, CD8+ T, and CD20+ B cells. These molecular differences underscore the necessity of subtype-specific models to better predict survival risk and treatment responses, emphasizing the role of unique immune patterns in different histologic subtypes of NSCLC.

Sha et al. [34] developed a deep learning model to predict the status of programmed death ligand 1 (PD-L1) in non-small cell lung cancer (NSCLC) using whole-slide Hematoxylin and Eosin (H&E) images. The investigation included 130 NSCLC patients whose samples were allocated to the training and test cohorts. The model employed a multi-field-of-view approach with a residual neural network (ResNet) backbone to analyze over 145,000 tiles extracted from H&E slides. It accurately predicted PD-L1 status in a balanced test cohort, achieving an area under the receiver operating characteristic curve (AUC) of 0.80.

The model demonstrated robustness across varying PD-L1 positivity thresholds (AUC = 0.67–0.81) and maintained performance even when simulating inter-observer variability (AUC = 0.63–0.77). These findings suggest that PD-L1 expression is associated with the morphological characteristics of the tumor microenvironment, indicating that deep learning models can complement traditional immunohistochemistry (IHC) testing in clinical settings where resources are limited.

Kanavati et al. [35] developed a deep learning model utilizing convolutional neural networks (CNN) and recurrent neural networks (RNN) to classify challenging cases of lung carcinoma from small transbronchial lung biopsy (TBLB) specimens. These cases frequently present diagnostic challenges for pathologists using only Hematoxylin and Eosin (H&E)-stained slides and typically require further immunohistochemistry (IHC) or surgical resection to confirm the diagnosis. The model was trained on 579 whole

slide images (WSIs) and evaluated on an independent test set of 83 indeterminate cases, achieving an area under the receiver operating characteristic curve (AUC) of 0.99.

To validate the model further, it was tested on additional datasets, including 502 TBLB specimens and 1,905 surgical specimens from different medical centers, with AUCs ranging from 0.94 0.99. These findings suggest that this deep learning approach could serve as a valuable adjunct diagnostic tool, aiding in the sub-classification of lung cancer types, such as adenocarcinoma (ADC), squamous cell carcinoma (SCC), small-cell lung cancer (SCLC), and non-neoplastic tissues, particularly in cases with poorly differentiated carcinoma cells.

Lu et al. [36] introduced a prognostic model designed to predict the five-year overall survival of patients with early-stage non-small cell lung cancer (NSCLC) based on cellular diversity features extracted from hematoxylin and eosin (H&E)-stained histology images. This study included 1,057 patients from four distinct centers, and the model specifically utilized features representing local cellular morphological diversity (termed CellDiv) within the tumor epithelium region. A Cox proportional hazards model was developed separately for lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC), employing data from two cohorts and validated on two independent cohorts.

The investigation revealed that patients classified as high-risk by the CellDiv model exhibited significantly worse five-year overall survival in both validation cohorts (hazard ratios ranged from 1.48 to 2.24, p-values ranging from 0.022 to 0.0058). The features of CellDiv were also associated with biological pathways related to apoptotic signaling and cell differentiation, suggesting that the model could identify patients who might benefit from additional adjuvant therapy. These findings underscore the potential of image-based risk models to provide a non-invasive method for enhancing prognostic assessments in early stage NSCLC.

Yu et al. [37] developed a fully automated system to predict the prognosis of non-small cell lung cancer (NSCLC) patients using quantitative features extracted from hematoxylin and eosin (H&E) stained histopathology images. The study analyzed 2,186 whole-slide images from The Cancer Genome Atlas (TCGA) and 294 images from the Stanford Tissue Microarray (TMA) database, focusing on lung

adenocarcinoma and squamous cell carcinoma. A total of 9,879 image features were extracted, and machine learning techniques were applied to identify the most relevant features that could differentiate between patients with different survival outcomes.

The predictive models, validated using independent datasets, successfully distinguished short-term survivors from long-term survivors of stage I adenocarcinoma and squamous cell carcinoma ($p < 0.003$ and $p = 0.023$, respectively). The results indicated that these automatically derived image features can provide valuable prognostic information for lung cancer patients, potentially facilitating more personalized treatment approaches. This methodology could also be extended to the analysis of histopathological images of other cancers, contributing to the advancement of precision oncology.

Wang et al. [38] developed and validated a computational pathology risk score (CoRiS) to predict overall survival (OS) and disease-free survival (DFS) in patients with early-stage non-small cell lung cancer (ES-NSCLC). This study encompassed 1,330 patients from three independent cohorts. CoRiS was constructed by applying an elastic-net Cox model to H&E-stained tissue images and identifying prognostic features to stratify patients into different risk groups. The prognostic ability of the model was validated across multiple datasets, demonstrating significant associations with both OS and DFS, independent of clinicopathologic factors.

Furthermore, CoRiS has also been demonstrated to be predictive of the benefit of adjuvant chemotherapy (ACT) in patients with ES-NSCLC. High-risk patients identified by CoRiS who received ACT exhibited improved survival outcomes compared to those who did not receive ACT, whereas low-risk patients did not show significant survival benefits from ACT. This suggests that CoRiS could function as a cost-effective, non-destructive tool to guide the management of patients with ES-NSCLC, potentially identifying those who would benefit most from additional chemotherapy following surgery.

Lin et al. [39] proposed a novel prognostic tool for non-small cell lung carcinoma (NSCLC) based on the digital quantification of tumor cellularity. Using digital pathology tools, they analyzed histopathological images from The Cancer Genome Atlas, which encompasses 213 cases of lung adenocarcinoma (ADC) and 90 cases of lung squamous cell carcinoma (SCC). The investigators calculated tumor cellularity

by determining the number of tumor cells per unit area and multiplying it by tumor size, thereby providing a comprehensive measure of cell density.

The investigation demonstrated that high-grade histological patterns, such as micropapillary and solid patterns in ADC and poorly differentiated types in SCC, were associated with higher tumor cell densities. Patients with lower tumor cellularity had significantly better overall and progression-free survival rates than those with higher cellularity. These findings suggest that tumor cellularity, quantified through digital methods, could serve as a robust prognostic marker, potentially enhancing patient stratification for the personalized treatment of NSCLC.

Shim et al.[40] proposed the DeepRePath model, which utilizes multi-scale pathological images and deep convolutional neural networks (CNNs) to predict the recurrence of early stage lung adenocarcinoma (LUAD). This study involved pre-training the model on whole-slide images from the Cancer Genome Atlas and further training it on a separate dataset of 393 resected LUAD specimens. The DeepRePath model demonstrated an average area under the curve (AUC) score of 0.77 and 0.76 on internal and external validation cohorts, respectively. Key pathological features associated with a high recurrence risk identified by the model included tumor necrosis, discohesive tumor cells, and atypical nuclei. Despite the limitations related to sample size, the study suggests that the DeepRePath model could improve treatment decisions for patients with early stage LUAD by accurately predicting recurrence risk.

Wang et al. [41] developed ConvPath, a software tool that utilizes convolutional neural networks (CNNs) to analyze digital pathology images of lung adenocarcinoma (ADC). The tool is designed to automate the identification and classification of cell types in histopathological images, including tumor cells, stromal cells, and lymphocytes. ConvPath employs a multistep pipeline that involves nuclei segmentation, cell type classification, and extraction of features related to the tumor microenvironment. The software was trained and validated using datasets from multiple independent cohorts, achieving classification accuracies of 92.9% for the training data and 90.1% for the external test datasets.

ConvPath transforms pathology images into "spatial maps" that illustrate the distribution of different cell types, facilitating a more comprehensive analysis of their organization and interactions. This approach enables the development of prognostic

models that consider the tumor microenvironment, which has been validated as an independent prognostic factor for lung ADC patient outcomes, even after adjusting for clinical variables, such as age, sex, smoking status, and cancer stage.

Wu et al. [42] developed a deep learning framework named DeepLRHE, which employs convolutional neural networks (CNNs) to predict the risk of recurrence and metastasis in lung cancer patients by analyzing histopathological images. The model incorporates multiple steps including tumor region identification, image normalization, biomarker identification, and classification. It was trained and validated on 110 lung cancer samples from The Cancer Genome Atlas (TCGA) and tested on an independent set of 101 samples. The model achieved an area under the receiver operating characteristic curve (AUC) of 0.79, indicating good predictive performance. DeepLRHE is a cost-effective and efficient alternative to traditional methods, such as circulating tumor cell detection and next-generation sequencing, which are often expensive and time-consuming. This framework has the potential to assist clinicians in determining which patients might benefit from additional adjuvant therapy, thereby enhancing personalized treatment strategies for lung cancer.

CHAPTER 3

DATASET AND PREPROCESSING

3.1 Dataset

The LC25000 dataset was obtained from James A. Haley Veterans Hospital. It initially comprised 750 lung tissue images and 500 colon tissue images, encompassing five distinct classes of 250 images each with dimensions of 1024 × 768 pixels. These images, categorized as benign lung tissue, lung adenocarcinoma, lung squamous cell carcinoma, benign colon tissue, and colon adenocarcinoma, were subsequently cropped to 768 × 768 pixel square dimensions using Python. Employing Augmentor, 25-degree left and right rotations with a probability of 1.0, and horizontal and vertical flips with a probability of 0.5 were applied. Consequently, the number of images in each class was increased to 5000, resulting in a total of 25000 images, which were compressed into a 1.85 GB zip file and saved in JPEG format[43].

Furthermore, the dataset's adherence to HIPAA compliance and the measures implemented for de-identification facilitate the ethical management of sensitive medical information, thereby fostering confidence and ensuring the observance of ethical standards in AI research[43]. The utilization of the LC25000 dataset is of significant importance for the advancement of diagnostic pathology and enhancement of cancer diagnostic practices through the application of machine learning and deep learning algorithms[43].

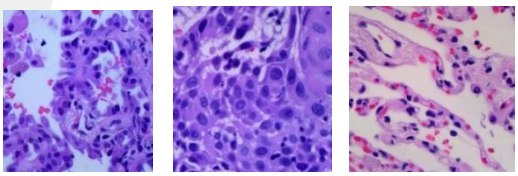


Figure 3.1 Sample Histopathological Lung Cancer Images of Original LC25000 Dataset are lungaca4991.jpeg, lungscc542.jpeg, lungn1287.jpeg respectively.

3.2 Preprocessing

Duplicate images were initially eliminated from the dataset, and the remaining images were converted from JPEG format to TIFF format. Subsequently, the retained images in the dataset were processed using a series of digital image processing techniques to enhance classification. Given that the classification models rapidly learned the normal cell class, these techniques were exclusively applied to the adenocarcinoma class (lung_aca) and squamous cell carcinoma class (lung_scc).

To prevent memory leaks during all digital image processing techniques, unused variables were cleared, and Python's garbage collector (`gc.collect()`) was invoked to free up the memory. After processing each image, the memory usage was monitored using the "psutil" library. For large datasets, the garbage collector was called after every 10 images to optimize the performance and memory usage. The directories containing images for each lung cancer type were specified, and output directories for saving the processed images were created. Any potential errors during the processing were captured to ensure that the process could continue without interruption.

3.2.1 Detection and Removal of Duplicated Histopathological Images in Original Image Folders

The presence of duplicate images can potentially introduce bias into the results of machine learning and deep learning applications, particularly in the context of medical image analysis. In the present work, duplicate images are systematically identified and moved to specified directories by removing them from the original folders. In the DuplicateImageFinder class, dictionaries named “hash_map” to store unique hashes of images and allow efficient comparison, and duplicates to save “duplicates” by folder have been created respectively. Additionally, a “transform” attribute is defined to convert images to tensors, ensuring compatibility with a variety of formats and facilitating image comparison.

The function of “create_duplicate_folders” checks whether each duplicate folder exists in the root directory containing the source and duplicate folders. If not, it creates the index and ensures that copies can be safely relocated without overwriting the source images. A hash-based method is used to detect duplicates. The function of “compute_hash” generates a unique hash for each image. These images are then converted to bytes and hashed using MD5 to produce a consistent hash value for identical images. By using the hash as a unique identifier, the method ensures an accurate and efficient comparison across large datasets.

The function of "process_images" iterates through each source folder, opening and processing images in supported formats such as PNG, jpg, jpeg and TIFF. It converts each image to a tensor and then to a NumPy array, which facilitates transformation and manipulation. If an image hash already exists in “hash_map”, the image is identified as a duplicate and recorded in “duplicates”; if the hash is new, it is added to “hash_map” along with the associated folder and filename.

The “handle_duplicates” function handles the saved duplicates by moving them to the specified duplicate folders and recording the details in a log file. It uses “shutil.move” to move duplicates from the source folder to the specified duplicate folder so that duplicate bias does not occur in the source data. In addition, each duplicate pair (original and duplicate file) is saved in a “duplicate_images.txt” file for reference.

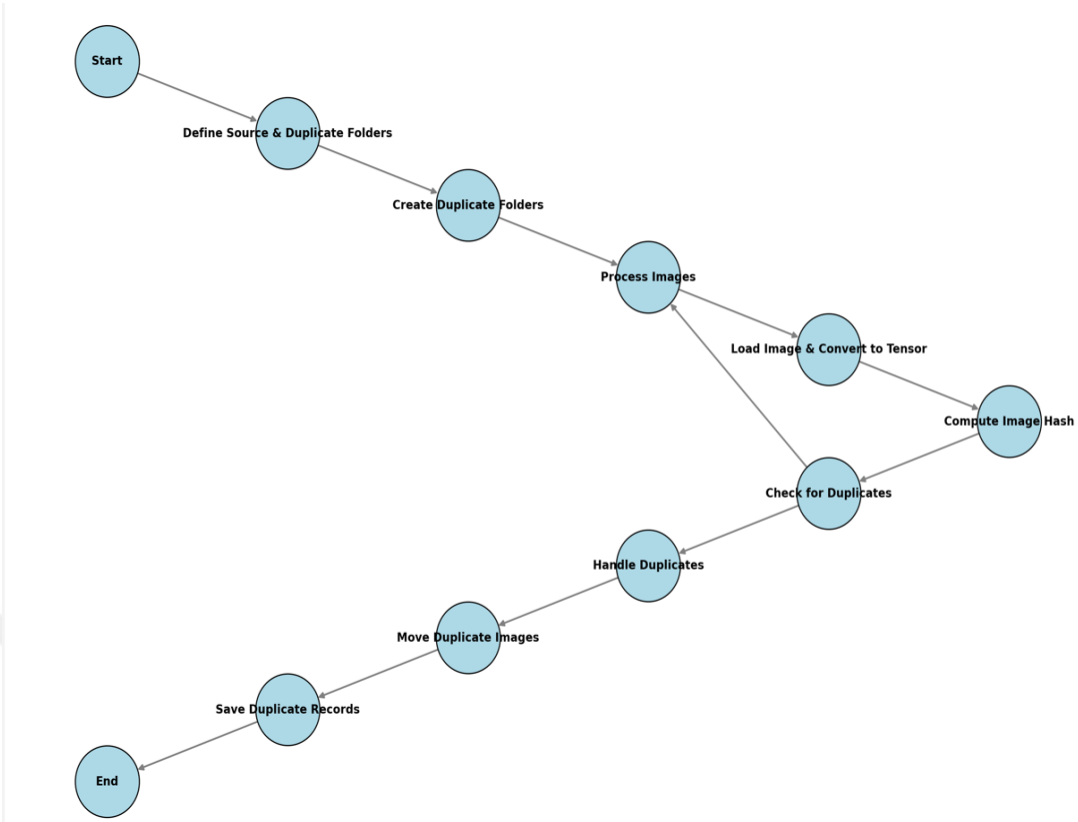


Figure 3.2 Dublicate Image Removal Pipeline

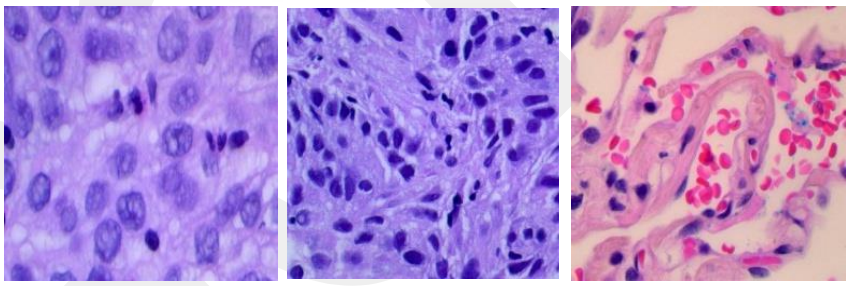


Figure 3.3 Sample Duplicate Images are lungscc1346.jpeg, lungaca270.jpeg, lungn29.jpeg respectively.

3.2.2 Conversion to tiff files and Metadata Preservation for Medical Image Datasets Using ExifTool

Images suffer data loss when saving JPEG format files. Since the images in the LC25000 dataset are in JPEG format, any digital image processing applied to these images will still save them lossy. In addition, the discrete wavelet transform applied just before the classification process is not effective on JPEG format files. Therefore, ExifTool automates the conversion of medical images to TIFF format with LZW or DEFLATE compression while preserving all metadata.

The purpose of this process is to standardize the image format and preserve the metadata required for subsequent medical image analysis tasks.

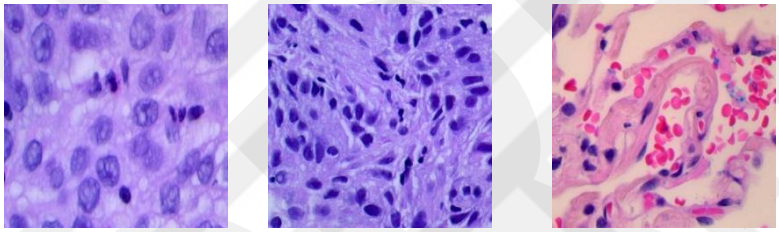


Figure 3.4 Sample of images converted to TIFF format are lungscc1346.tiff, lungaca270.tiff, lungn29.tiff respectively.

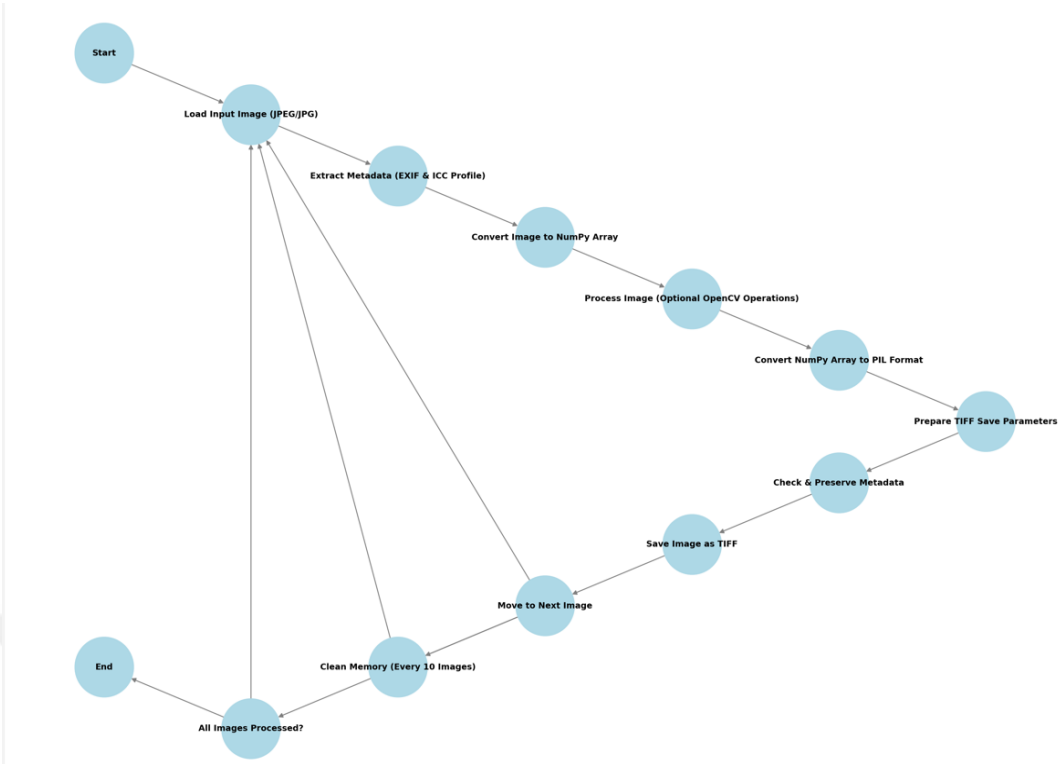


Figure 3.5 TIFF Conversion Pipeline

3.2.3 Reinhard Method for Image Enhancement

The Reinhard method is a linear contrast-enhancement technique that was initially introduced to transfer color statistics from a reference (target) image to an original image. In contrast to traditional histogram equalization methods such as Global Histogram Equalization (GHE), CLAHE, and other nonlinear methods, the Reinhard method preserves nearly all the information from the source image, ensuring minimal data loss during the enhancement process. This characteristic renders it particularly suitable for medical imaging, particularly for tasks such as tumor detection or cancer diagnosis, where preserving subtle features is crucial [44]. Furthermore, color normalization has demonstrated significant value in digital pathology by enabling comparison and analysis of images from diverse sources. This standardization is particularly relevant for maintaining interpretive consistency, as histopathological slides are susceptible to color shifts depending on the dye type, slide thickness, and scanning method employed [46]. Techniques such as the Reinhard method, which

operates in a perceptual color space such as Lab, facilitate the transfer of mean color statistics from a target image, while preserving the structural fidelity of the source. This unique combination of consistency and detail retention supports more reliable and reproducible analyses, especially within the framework of automated diagnostic tools [45]. The key characteristics of the Reinhard Method include linear transformation, contrast, color transfer, and an automatic approach. Unlike nonlinear methods, linear transformation preserves the original structure and data integrity of the image, which may result in significant data loss. The linearity of the Reinhard transformation ensures that the correlation between the processed image and source image is close to 1, thereby minimizing the loss of crucial image details. For medical images, contrast and color transfer methods have demonstrated efficacy in enhancing the contrast without introducing noise or other undesirable artifacts. These techniques highlight essential features such as tumor boundaries while preserving other details. Unlike methods such as CLAHE, which necessitate manual selection of parameters such as the contrast clipping limit, the Reinhard method is fully automated. The user is only required to select an appropriate target image with a sufficient contrast, which simplifies the application process. The Reinhard method is particularly effective in enhancing images for tumor-detection tasks. This method ensures that diagnostically significant details, such as differences in tissue textures or tumor boundaries, are maintained without being obscured by artifacts introduced during enhancement[47]. Histopathological reference images were used in this study. These reference images provided the target color distribution used to align the dataset[48].

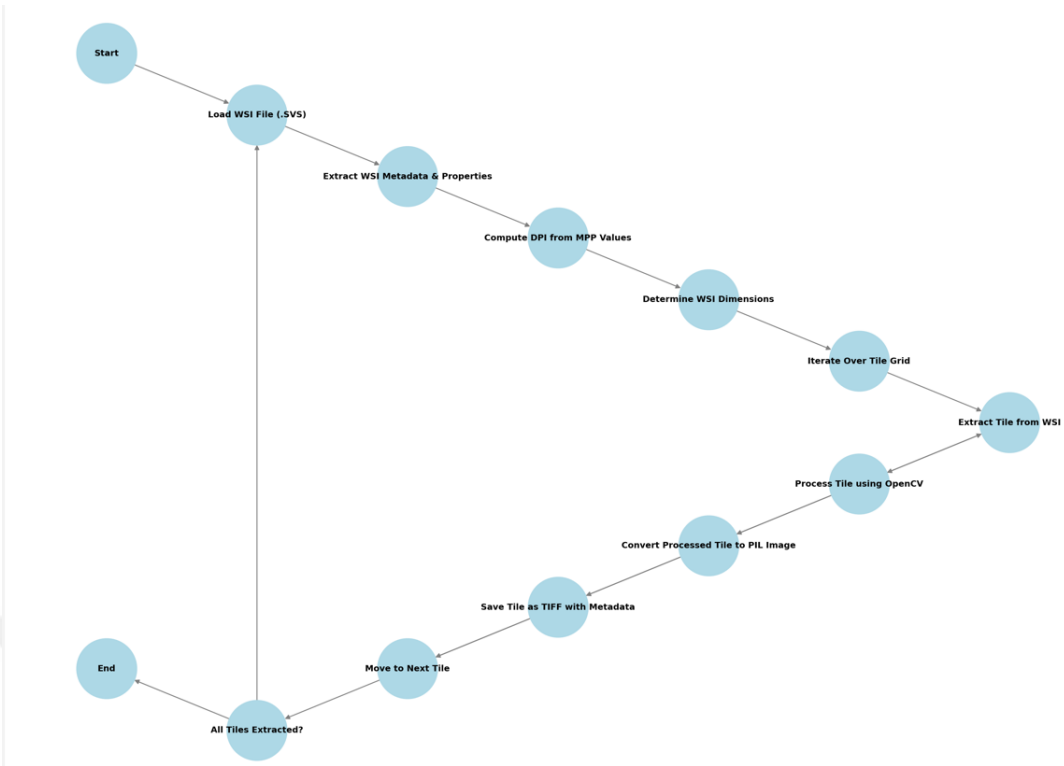
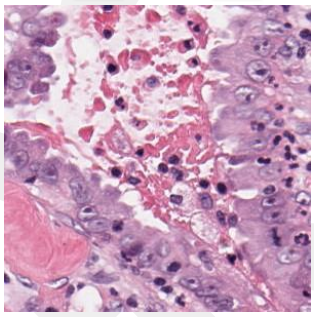
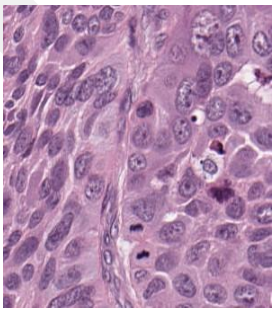


Figure 3.6 WSI Tile Extraction Pipeline



PKG - CPTAC-LUAD_v12



PKG - CPTAC-LSCC_v10

Figure 3.7 The histopathological reference images used in this study were derived from the Cancer Imaging Archive, specifically from the Whole Slide Images (WSIs) named PKG - CPTAC-LUAD_v12 and PKG - CPTAC-LSCC_v10, by dividing them into tiles[48].

In this study, the source and target images were read and initially converted to the RGB format to ensure consistency in color representation. Grayscale images were converted

into a color format to maintain uniformity across the dataset. All images were then converted to the 'float32' data type and pixel values were scaled to fall within the range [0, 1] for optimized processing and compatibility in subsequent analysis steps.

RGB images were transformed into Lab color space, which aligns more closely with human visual perception by representing color in terms of a lightness (L) channel along with two color channels, a and b.

The mean and standard deviation values for the L, a, and b channels were calculated for both source and target images, with a small epsilon value added to the standard deviation to prevent potential division by zero errors.

For each channel, the mean value of the source image was first subtracted, followed by scaling of each channel by the ratio of the standard deviations of the target and source images. Finally, the mean values of the target image are added back to each channel.

The processed L, a, and b channels were merged to form a new Lab image, which was then converted back into the RGB color space. Pixel values were clipped to the valid range [0, 1], rescaled to [0, 255], and finally converted to "uint8" format.

Processed images were saved while preserving the original image format and metadata, with appropriate compression and quality settings applied to the JPEG, PNG, and TIFF formats. Additionally, metadata such as EXIF and ICC profile data were retained to ensure consistency with the original files.

The directories containing images for each lung cancer type were specified, and output directories for saving the processed images were created.

The "process_dataset" function was utilized to apply Reinhard color transfer to all images in the dataset in batches.

Any potential errors during the processing were captured to ensure that the process could continue without interruption.

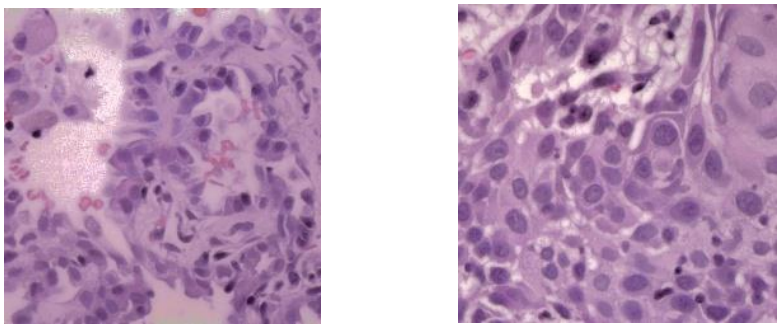


Figure 3.8 The Sample Images lungaca4991.tiff and lungscc542.tiff are the Reinhard Color Transformation-applied versions.

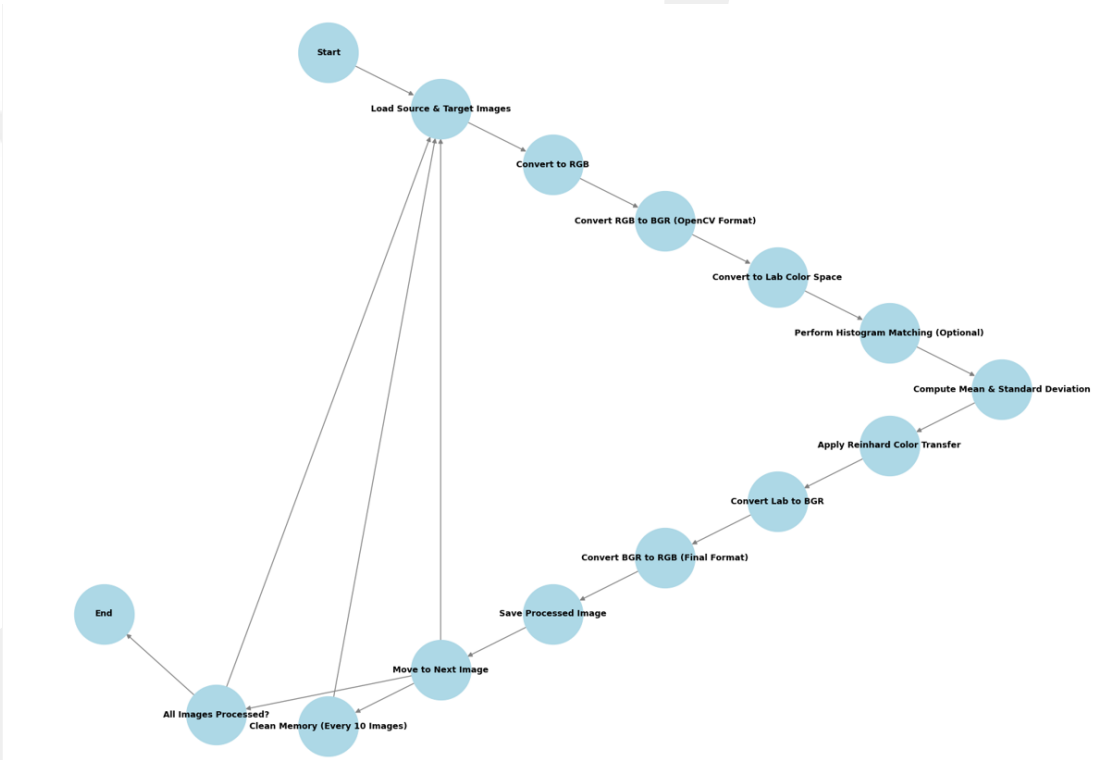


Figure 3.9 Reinhard Color Transformation Pipeline

3.2.4 Vahanade Color Transformation

Maintaining a consistent color representation in histological images is fundamental for accurate diagnosis and effective computational analysis in medical imaging. However, variations in staining protocols, differences in scanner configurations, and other technical factors frequently introduce inconsistencies that can compromise both manual assessments and algorithmic interpretations. To address these challenges,

Vahanade Color Transformation presents an innovative approach for stain separation and color normalization. This framework not only mitigates undesirable color variations, but also preserves the intricate biological structure of tissue samples, thereby enhancing the reliability and reproducibility of histopathological evaluations[80].

3.2.4.1 Sparse Non-Negative Matrix Factorization (SNMF)

A pivotal component of Vahanade Color Transformation is Sparse Non-Negative Matrix Factorization (SNMF), a sophisticated mathematical technique for deconstructing histological images into biologically meaningful elements. By incorporating non-negativity and sparseness constraints, SNMF ensures that stain densities are not only mathematically valid, but also biologically accurate, capturing the presence or absence of light absorption without introducing artificial or negative values. This precision enables SNMF to effectively isolate individual stain densities, producing a faithful structural representation of tissues that reflects their natural composition[80].

3.2.4.2 Structure-Preserving Color Normalization (SPCN)

To complement the capabilities of SNMF, Structure-Preserving Color Normalization (SPCN) focuses on harmonizing the color appearance of histological images. By aligning the color characteristics of a source image to match the target image preferred by experts, SPCN maintains the structural integrity of the tissue while achieving standardization. In contrast to conventional methods that distort the tissue architecture, SPCN preserves the biological information encapsulated in the stain density maps. This is accomplished by substituting the color basis of the source image with that of the target image, enabling consistent color representation without altering fundamental tissue properties. Such adaptability renders the SPCN particularly valuable for multi-institutional studies and computational pathology workflows that rely on uniform and accurate input data[80].

By seamlessly integrating SNMF and SPCN, the Vahanade Color Transformation presents a comprehensive solution to the persistent challenge of color inconsistencies

in histological imaging. This approach achieves an optimal balance between maintaining structural authenticity and achieving color uniformity, thereby facilitating reliable and interpretable analyses in both clinical and research settings.

3.2.4.3 Applying Vahadane-Based Stain Normalization of Histopathological TIFF Images

Vahadane's method provides an advanced approach for stain normalization in histopathological images, ensuring consistent color representation across samples. This procedure is particularly efficacious in medical imaging scenarios where varying staining protocols can introduce challenges to quantitative analysis.

The StainNormVahadane class constitutes the core of the implementation. It was designed to adjust the input ("source") image to correspond with the staining profile of a reference ("target") image. The process relies on optical density (OD) transformations and Non-Negative Matrix Factorization (NMF) to factorize strains into distinct components and recombine them based on the characteristics of the target.

Within this class, a fit function is utilized to load and analyze the target image. During this process, the target image is converted to RGB and transformed into an optical density domain. Low-density pixels that could introduce noise were filtered out. NMF was subsequently employed to extract the specific stain signatures of the target image, resulting in a "stain matrix" that characterized the major staining sources. A transform function then applies the previously learned target-stain matrix to each source image. Similar transformations (RGB to OD) and pixel filtering were performed on the source image, followed by the computation of its own stain matrix. By aligning these source stain contributions with the target's matrix, the method reconstructs the source image in a manner that accurately reflects the color distribution of the target.

EXIF and ICC profile metadata were preserved by loading and saving images through the PIL library, ensuring compatibility with medical imaging standards that often require metadata retention.

Furthermore, batch processing capabilities allow multiple TIFF images to be sequentially normalized. Folders containing source images can be iterated using the transform function applied to each file. Progress indicators (such as those from the "tqdm" library) provide real-time feedback, whereas periodic garbage collection further optimizes memory usage. This design is intended to streamline workflows in which numerous images must be standardized in an automated fashion, which is essential for large-scale digital pathology projects.

The main execution block demonstrates how users can specify source and target directories, instantiate the StainNormVahadane class, and batch process entire sets of TIFF images. This ensured that all images adhered to a consistent staining profile, facilitating subsequent analysis tasks and enhancing the reproducibility of the results.

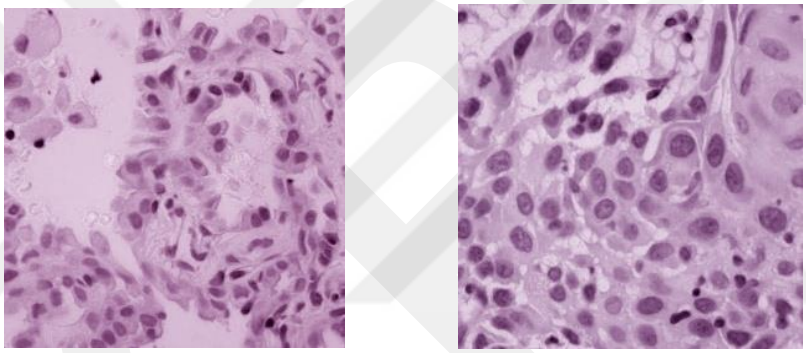


Figure 3.10 The Sample Images lungaca4991.tiff and lungscc542.tiff are the Vahanade Color Transformation-applied versions.

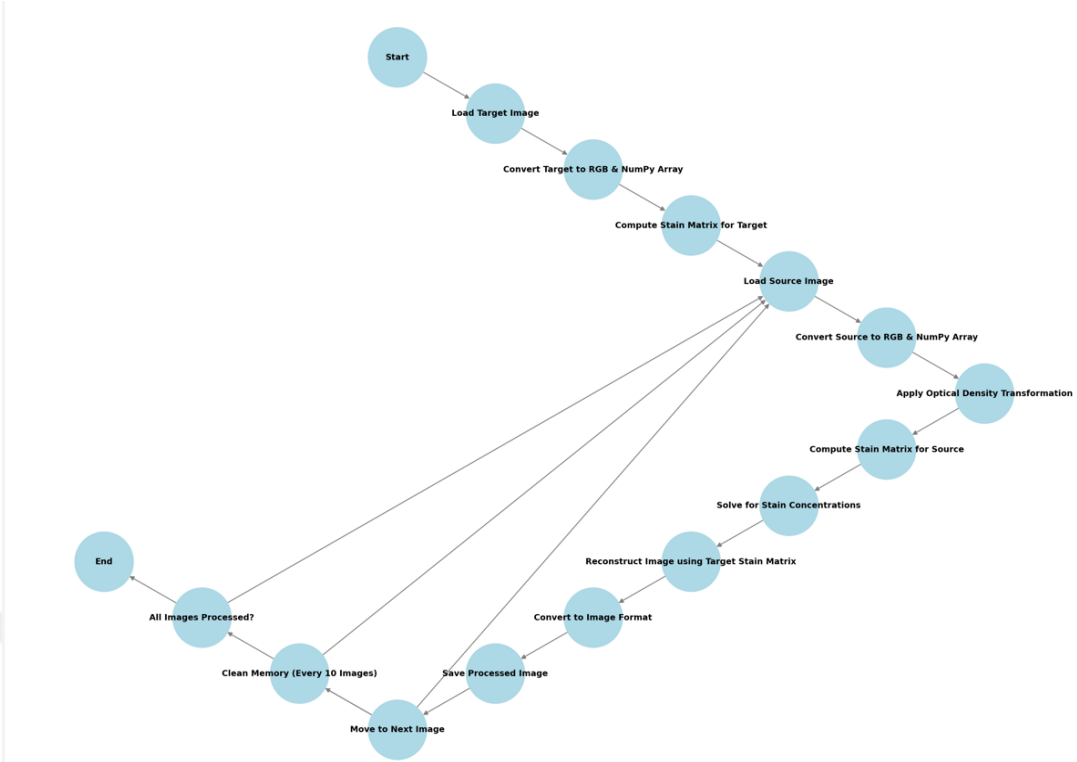


Figure 3.11 Vahanade Color Transformation Pipeline

3.2.5 Histogram Equalization

Histogram equalization is a fundamental nonlinear contrast-enhancement technique used in digital image processing to improve the contrast of an image. The method functions by redistributing the intensity values of an image such that they span the entire possible range, thereby enhancing the visibility of details in both dark and bright regions. It is particularly efficacious in situations where images exhibit poor contrast owing to suboptimal lighting conditions or specific characteristics of the imaging devices. This process involves transforming the intensity values such that the output histogram of the image is uniformly distributed and consequently improves the overall appearance of an image by increasing its dynamic range, thus rendering the features in the image more discernible [49].

3.2.5.1 Global Histogram Equalization (GHE)

There are various types of histogram equalization (HE) exist. Global Histogram Equalization (GHE) is the most basic and widely utilized method of histogram

equalization. It functions by uniformly applying the equalization process across the entire image. The transformation function employed in GHE is derived from the cumulative distribution function (CDF) of the complete image histogram, which subsequently redistributes the pixel intensity values to enhance the overall image contrast[47]–[50]–[51].

Nevertheless, GHE exhibits certain limitations, including a mean shift, over-enhancement, and lack of adaptability. Owing to the application of the same transformation across the entire image, it fails to account for local image characteristics [47].

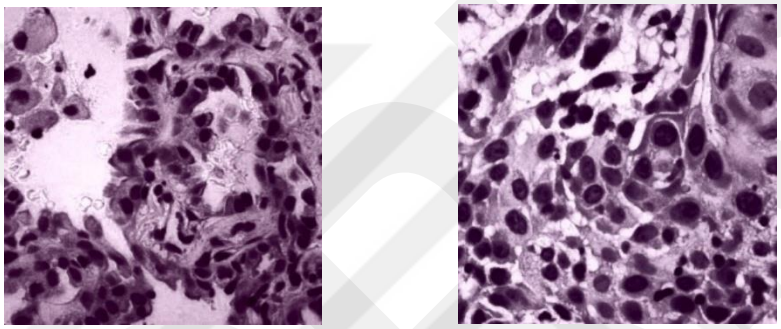


Figure 3.12 GHE was applied to the sample images lungaca4991.tiff and lungscc542.tiff, which had previously undergone the Vahanade color transformation.

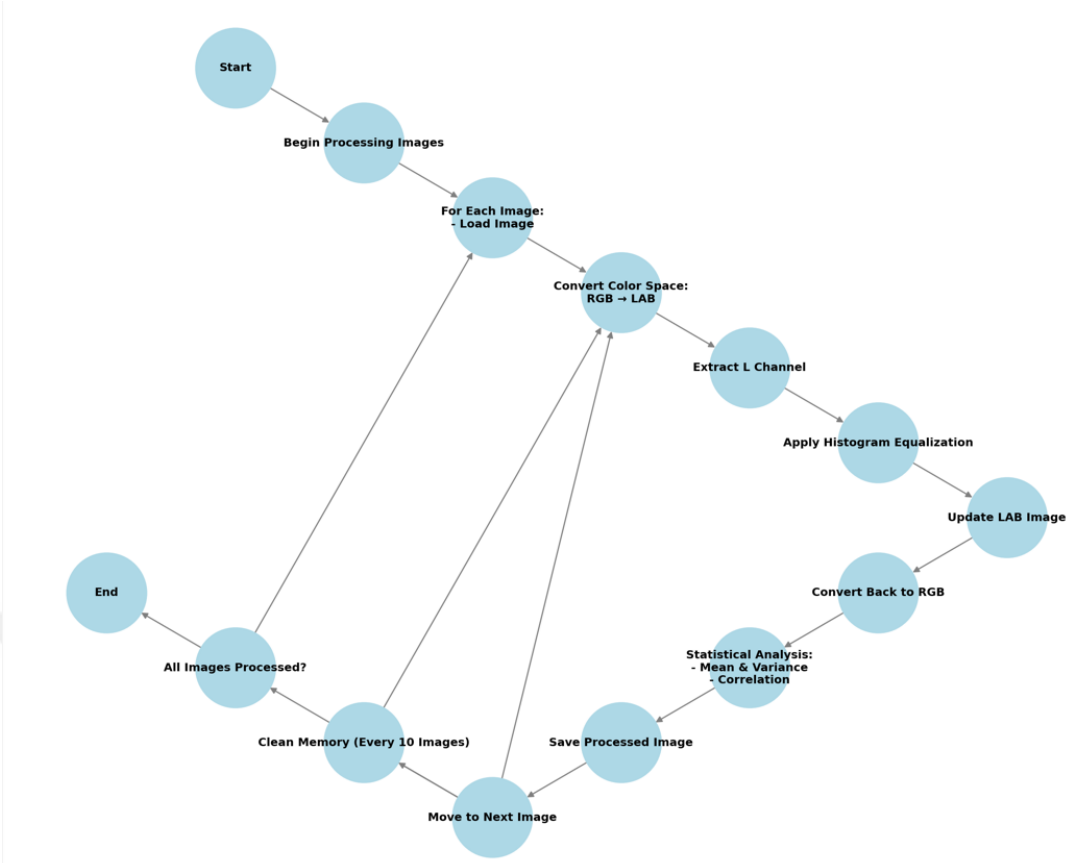


Figure 3.13 Global Histogram Equalization (GHE) Pipeline

3.2.5.2 Introduction to Adaptive Histogram Equalization (AHE) and Contrast Limited Adaptive Histogram Equalization (CLAHE)

Adaptive Histogram Equalization (AHE) and contrast-limited adaptive histogram equalization (CLAHE) are two variants of histogram equalization that have been extensively utilized in medical image processing to address the limitations of conventional global Histogram Equalization (GHE). These methodologies are designed to enhance the local contrast, rendering them suitable for images with varied brightness or intricate textures, such as those encountered in medical diagnostics. AHE operates by segmenting an image into small, non-overlapping regions and subsequently applying histogram equalization to each region independently. This localized processing facilitates contrast enhancement on a smaller scale, thereby ensuring that even subtle differences in texture and brightness are amplified. CLAHE was developed as an improvement on AHE to address its limitations, particularly with respect to noise amplification. In CLAHE, the contrast enhancement is constrained by

introducing a "clipping limit" to the local histograms. This means that pixel values exceeding a certain frequency are redistributed uniformly across the histogram, mitigating excessive contrast enhancement, and thereby reducing the risk of noise amplification. The clipping limit can be adjusted to regulate the degree of contrast enhancement, offering flexibility based on the requirements of a specific medical application[47]-[52].

3.2.5.3 Applying CLAHE to Vahanade-Standardized TIFF Images

An advanced implementation of contrast-limited adaptive histogram equalization (CLAHE) was provided for processing Reinhard standardized medical images stored in TIFF format.

The ImprovedCLAHEEqualizer class was initialized with two primary parameters: clip_limit and tile_grid_size. The clip_limit parameter restricts contrast amplification to prevent excessively bright areas with a default value of 2.0. The tile_grid_size parameter, a tuple, specifies the grid size for local contrast adjustments and provides control over local regions affected by CLAHE.

The "apply_clahe" method applies CLAHE to both color and grayscale images utilizing OpenCV's LAB color space, which separates lightness from chromaticity components. For images with three channels, the RGB image was converted into the LAB color space to isolate the luminance channel (L). CLAHE is applied exclusively to this L channel to avoid altering the color balance, and the processed channels are subsequently merged and converted back into RGB. For single-channel images, CLAHE was applied directly to enhance the contrast in grayscale formats.

The `process_image` method manages the complete CLAHE processing for each image. Initially, images are loaded via the PIL library, which allows the retention of metadata, such as EXIF and ICC profiles. Subsequently, the "apply_clahe" function is utilized to enhance the loaded image. Following enhancement, the image is saved in its original format, and metadata (EXIF data and ICC profiles) are maintained, ensuring compatibility and consistency with medical imaging standards. Additionally,

the processing time was recorded to monitor the performance and optimize for larger datasets.

Function process_all_images enables the batch processing of TIFF images within specified directories. The function first verifies and creates output directories, if they do not exist. It iterates through each image in the directory, invoking "process_image" to apply CLAHE and periodically calls garbage collection after every 50 images to prevent memory overflow. A progress bar from "tqdm" visually tracks the number of processed images, enhancing the usability in extensive datasets.

The main execution section specifies the input and output directories for the different categories of medical images. An instance of "ImprovedCLAHEEqualizer" was initialized with a default configuration, and images from each category were processed using the batch processing method.

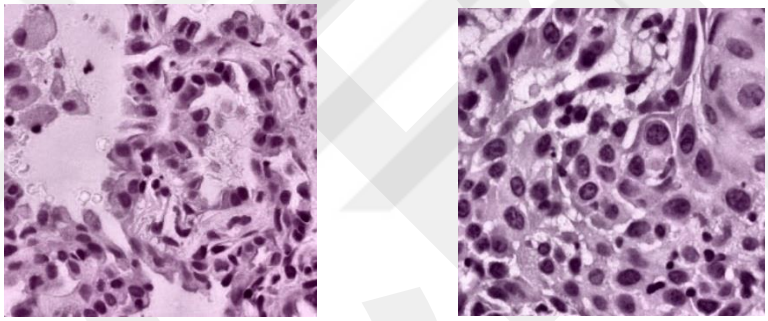


Figure 3.14 CLAHE with sharpening was applied to the sample images lungaca4991.tiff and lungscc542.tiff, which had previously undergone the Vahanade color transformation.

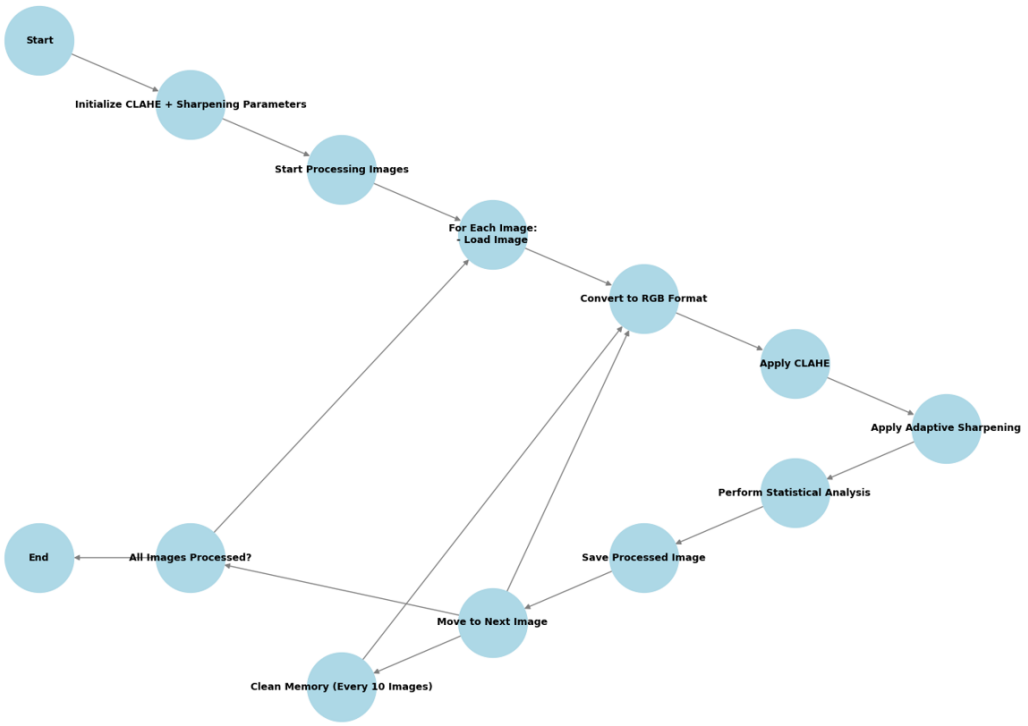


Figure 3.15 Contrast Limited Adaptive Histogram Equalization (CLAHE) with Adaptive Sharpening Pipeline

3.2.6 Super-Resolution

Single-image super-resolution (SISR) is an advanced area of computer vision that aims to reconstruct high-resolution (HR) images from their low-resolution (LR) counterparts. It has extensive applications in fields, such as medical imaging, remote sensing, and video processing. SISR presents a significant challenge owing to the inherent loss of detail during the degradation process from HR to LR images. Recent advancements, particularly in deep learning-based methods, have substantially improved SISR performance compared to traditional approaches [54]–[56].

The objective of SISR is to predict an HR image from a single LR image, a task that presents considerable challenges, particularly when addressing high upscaling factors (such as 4x or more). The reconstruction of finer textures becomes particularly challenging as the scaling factor increases, potentially leading to artifacts or excessively smooth textures [54]–[56].

Historically, SISR was approached using interpolation-based methods, such as bilinear or bicubic interpolation, which are computationally efficient but result in overly smooth outputs and loss of high-frequency details. Subsequently, more sophisticated methods, including reconstruction- and learning-based approaches, were introduced. In particular, deep learning has transformed this field, enabling significantly improved performance in recovering complex textures and details that earlier techniques struggled to address [54]–[56].

3.2.6.1 Classification of SISR Methods and Detailed Explanation of Figure

In a review of single-image super-resolution reconstruction using deep learning, three main categories of SISR models are discussed: Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), and transformers.

Convolutional Neural Networks establish mapping between low-resolution (LR) and high-resolution (HR) images through multiple convolutional layers. The incorporation of residual connections and attention mechanisms has facilitated improved feature extraction and image reconstruction [54]-[56].

GANs comprise two competing networks: the generator produces super-resolved images, whereas the discriminator assesses quality by comparing them to authentic HR images. SRGAN and its enhanced iterations, such as ESRGAN, are notable examples [54]–[56].

Transformer architecture, which demonstrates proficiency in capturing long-range dependencies and features in images, has also been applied to SISR tasks. Models such as CWT-Net utilize transformers to amalgamate features from different image scales, thereby yielding superior detail enhancement [57].

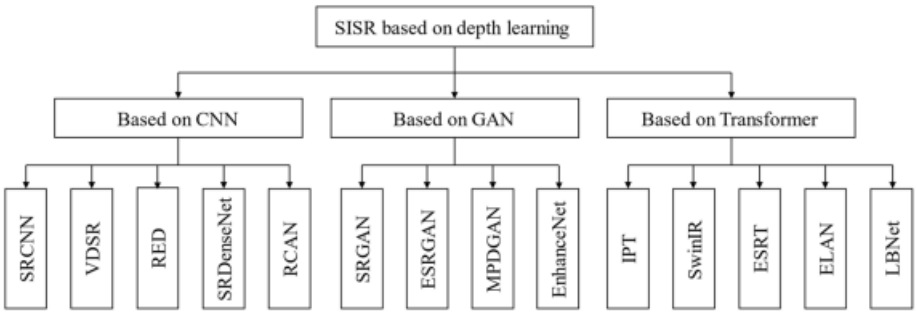


Figure 3.16 The classification of SISR algorithms in deep learning showcases a structured breakdown of methods based on underlying architectures [54].

Regression-based models employ end-to-end learning wherein the network directly maps LR images to their HR counterparts using regression functions. The primary objective is to minimize the discrepancy (or loss) between the output and ground truth by frequently utilizing the mean squared error (MSE). However, regression models may produce excessively smooth results [53].

Generative adversarial networks utilize adversarial loss, wherein the generator attempts to deceive a discriminator to classify the generated images as authentic. SRGAN and ESRGAN are prominent models that enhance perceptual quality by introducing losses focused on texture recovery [53]–[56].

Flow-based models employ a series of reversible transformations to map the LR image to the HR space by learning the probability distributions of the pixel values. Although computationally intensive, these models maintain greater diversity in the generated images but can produce excessively smooth textures [53].

VAE-based methods sample latent distributions to generate images. Although they offer greater diversity than regression models, they encounter difficulties with fine textures and details [53].

Denoising Diffusion Probabilistic Models (DDPM) utilize a diffusion process to progressively add and remove noise from images, enhancing both diversity and quality in the generated images [53].

3.2.6.2 ESRGAN and RRDB_PSNR_x4.pth

Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) represent an advancement over SRGAN, aiming to generate more realistic textures while mitigating the artifacts commonly observed in previous models. ESRGAN accomplishes this objective through the implementation of several key enhancements [55]: dual-in-residual dense blocks (RRDB), Perceptual Loss, and Relativistic GAN.

The Residual-in-Residual Dense Blocks (RRDB) architecture omits batch normalization layers, thereby facilitating the construction of deeper networks without encountering the vanishing gradient problem. RRDB emphasizes the learning of both residuals and dense features, which substantially enhances the capacity of the model to reconstruct fine details. The RRDB_PSNR_x4.pth model specifically improves the performance of 4x scaling factors by incorporating these advancements [55].

In lieu of employing pixel-wise loss, which can yield excessively smooth results, ESRGAN utilizes perceptual loss based on features extracted from a pretrained VGG network. This approach enables the model to prioritize high-frequency textures and perceptual similarity rather than pixel-level accuracy [55].

ESRGAN modifies the standard GAN methodology by incorporating relativistic GAN (RaGAN), wherein the discriminator predicts relative "realness" between real and generated images rather than providing a binary classification. This modification results in the generation of more natural textures in the final output [55].

3.2.6.3 Super-Resolution in Histopathology

Super-resolution techniques have been extensively applied in medical imaging, particularly in the field of histopathology. In this domain, the acquisition of high-resolution images of tissue samples is critical for diagnosis and research but is often resource-intensive and time-consuming. Super-resolution methods offer an efficient alternative for reconstructing high-quality images from low-resolution samples [58].

CWT-Net represents a significant advancement in the application of super-resolution to histopathological images. This approach integrates wavelet transforms with

transformer architectures to extract high-frequency features from multiple scales of medical images, thereby enhancing both image resolution and diagnostic accuracy. The multibranch methodology employed by CWT-Net facilitates the improved capture of texture information across various image scales, which is essential for the identification of small or subtle pathological features [57].

In the context of breast cancer histopathology, super-resolution techniques have been employed to enhance the resolution of tissue images, enabling more precise segmentation and classification of tumor regions. These advancements not only improve the quality of visualization but also enhance the performance of diagnostic algorithms [57].

3.2.6.4 Latest Advancements in Super-Resolution

Recent research has focused on improving super-resolution techniques, particularly in terms of efficiency and image diversity. The Latent Feature-oriented Diffusion Probability Model (LDDPM) is a novel approach that utilizes diffusion processes to generate high-resolution images from low-resolution inputs. This model employs a conditional encoder to constrain the solution space and multimodal adversarial training to enhance the capture of high-frequency details, thereby addressing the common issue of smooth textures in traditional SISR models [53].

Additional state-of-the-art developments include transformers specifically designed for SR tasks, which leverage their capacity to capture global image features more effectively than conventional CNNs do. These transformers have demonstrated particular efficacy in applications requiring multiscale detailed recovery, such as medical imaging and remote sensing [53].

3.2.6.5 Methods of Super-Resolution in this study

This section delineates the methodologies used for model loading, image processing, and batch processing of directories to generate super-resolved medical images.

The code employs the Peak Signal-to-Noise Ratio (PSNR) model based on the RRDBNet architecture, which specializes in super-resolution tasks:

The "load_model" function initializes the RRDBNet model with essential parameters, including input/output channels and the number of layers, and configures it to operate on either the CPU or GPU, contingent upon availability. The function loads the model parameters via "state_dict" and assigns them to the specified device. Furthermore, the model is set to "eval" mode to optimize computation and facilitate inference. This stage prepares a model for optimizing the PSNR values of the images, thus enhancing the image resolution.

The "process_image" function executes the super-resolution process on a single image through the following steps:

The input image is opened in the RGB format and normalized before being converted into a tensor. The model processes the image tensor utilizing "torch.inference_mode" and "torch.amp.autocast". These modes enhance the processing speed and memory optimization, particularly on GPUs. The tensor output from the model was converted back into a NumPy array and saved as an image. The saved image retains its original format (JPEG, PNG, etc.) and metadata. This stage meticulously configures each image to produce a high-resolution output, while preserving the original format and metadata.

The process_directory function applies the super-resolution model to all valid image files (JPEG, PNG, TIFF, etc.) in a directory. Each image is passed to the "process_image" function, which the model processes and saves in the output folder.

The main function orchestrates model loading, input and output directory setup, and batch processing for each category. It displays the progress of each directory's processing, informs the user upon completion, and performs the final memory cleanup. This modular configuration allows the model to process images in different folders efficiently, while supporting optimal system resource usage.

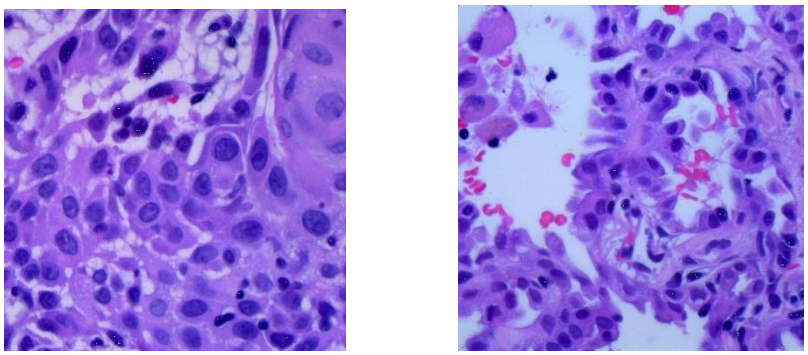


Figure 3.17 PSNR with RRDBNet architecture was applied to the sample original images lungaca4991.tiff and lungscc542.tiff.

Utilizing the PSNR with the RRDBNet architecture, the original 768 × 768 images were upscaled to 3072 × 3072 pixels, increasing their size by 16 times. Owing to the large file sizes, only the sample images were upscaled and were not used for classification. However, it can be employed as the final step in any digital image processing pipeline to enhance details.

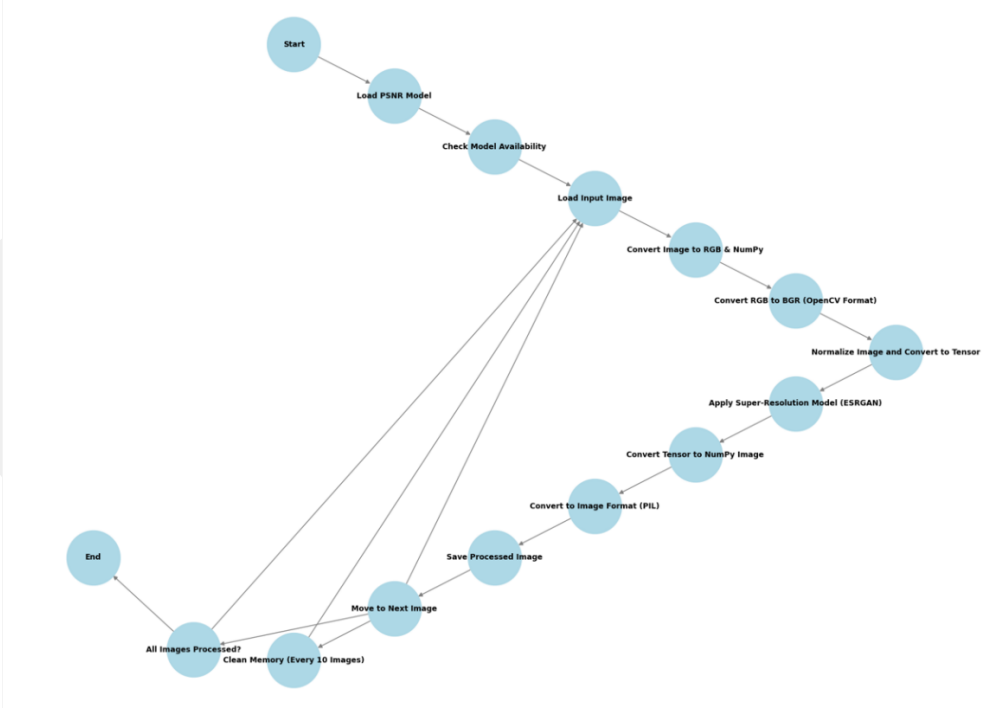


Figure 3.18 Super-Resolution with PSNR with the RRDBNet Architecture Pipeline

3.2.7 DCT-DWT

In the digital era, the protection of intellectual property and assurance of multimedia content authenticity have become critical concerns. Digital watermarking techniques provide a robust solution by embedding imperceptible signals into media files such as images to safeguard against unauthorized use or tampering. Among the numerous techniques available, the Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) are particularly noteworthy because of their capacity to represent image data in the frequency domain and their resilience against common attacks such as compression and noise [59] - [60] - [62].

3.2.7.1 Discrete Cosine Transform (DCT)

Discrete Cosine Transform (DCT) is a widely utilized transformation method that converts spatial domain data into frequency domain coefficients, thereby enabling the representation of image information based on various frequency components. It primarily decomposes an image into low, middle, and high-frequency components, wherein the watermark can be embedded in the middle frequencies to achieve an optimal balance between imperceptibility and robustness. The low-frequency components represent the homogeneous areas of the image, whereas the high-frequency components encompass the edges and fine details. The capacity of DCT to concentrate the energy of an image into a few significant coefficients renders it highly suitable for watermarking applications, particularly in compressed formats such as JPEG [59] - [60].

3.2.7.2 Discrete Wavelet Transform (DWT)

In contrast to DCT, DWT provides both frequency and spatial information by decomposing an image into different frequency sub-bands at multiple resolutions. DWT divides an image into low-low (LL), low-high (LH), high-low (HL), and high-high (HH) sub-bands, thereby facilitating a hierarchical representation of the image. The LL sub-band represents the approximation of the image and contains low-frequency components, whereas the LH, HL, and HH sub-bands capture horizontal, vertical, and diagonal details, respectively. This multiresolution capability renders

DWT highly effective for capturing localized features, which are crucial for watermark robustness against geometric distortions [59] - [60].

3.2.7.3 Methods of DCT-DWT in this study

This approach integrates Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) for noise reduction in histopathological images, ensuring computational efficiency and memory optimization even when processing high-resolution images in large quantities[62]. This section details the updated methodology applied to lung cancer histopathological images, focusing on advanced noise suppression techniques combined with efficient image processing workflows.

The pipeline applies DCT and DWT-based noise reduction to histopathological images, with the following key steps:

Image loading and preparation: Input images in formats such as TIFF, JPEG, and PNG are acquired and converted to the RGB format. Subsequently, the images were transformed into LAB color space, isolating the L channel for processing. This channel primarily represents luminance, rendering it suitable for noise-reduction.

DCT Application: The L channel is segmented into non-overlapping blocks (default size of 8 × 8), and DCT is applied to each block to transform the spatial domain into the frequency domain. Padding with reflection is employed when the image dimensions are not visible by the block size, thus preserving the border information.

DWT Decomposition: Haar wavelet decomposition was applied to the DCT coefficients of the L channel. This step decomposes the image data into multiple frequency components (e.g., LL, LH, HL, and HH sub-bands) for localized frequency analysis. Typically, a two-level decomposition is performed.

Noise Estimation and Thresholding: Noise was estimated from the HH sub-band using a median-based standard deviation calculation. A universal threshold (based on Donoho's method) was computed to suppress high-frequency noise. Soft thresholding

is then applied to the wavelet coefficients to reduce noise while preserving image details.

Reconstruction: The processed coefficients undergo inverse DWT, reconstructing the DCT-transformed L-channel. Subsequently, an inverse DCT was applied to reconstruct the processed L channel in the spatial domain.

Final Image Reconstruction: The processed L channel is combined with the unaltered A and B channels in the LAB color space. The results were converted back to RGB format. The final image is saved in the TIFF format with LZW compression, preserving the original metadata, such as ICC profiles and EXIF data.

Logging mechanisms capture the errors encountered during image processing, which are saved to an external log file (error_log.txt) for review.

The processing pipeline supports batch operations, allowing simultaneous handling of multiple images stored in specified directories. TIFF format was used for the output files to ensure compatibility and metadata preservation.

Key Advantages of the Approach

Enhanced Noise Suppression: By combining DCT and DWT, the updated method effectively suppresses noise, particularly in the luminance channel (L channel), while retaining the image details critical for histopathological analysis[60] - [61].

Computational Efficiency: The block-based DCT approach reduces the computational overhead, and the use of universal thresholding minimizes the need for manual parameter tuning.

Flexibility and Scalability: The modular design allows customization for various image resolutions and noise levels, rendering it applicable across diverse histopathological datasets[60] - [61].

This methodology significantly enhances the quality of histopathological images and facilitates accurate downstream analyses and interpretations in medical research[61].

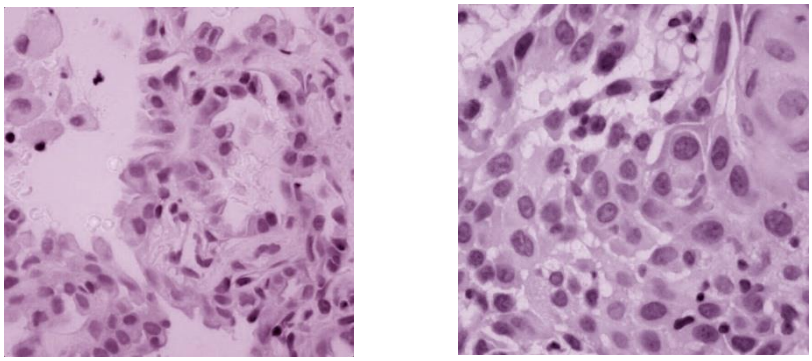


Figure 3.19 DCT-DWT was applied to the sample images lungaca4991.tiff and lungscc542.tiff, which had previously undergone the Vahanade color transformation.

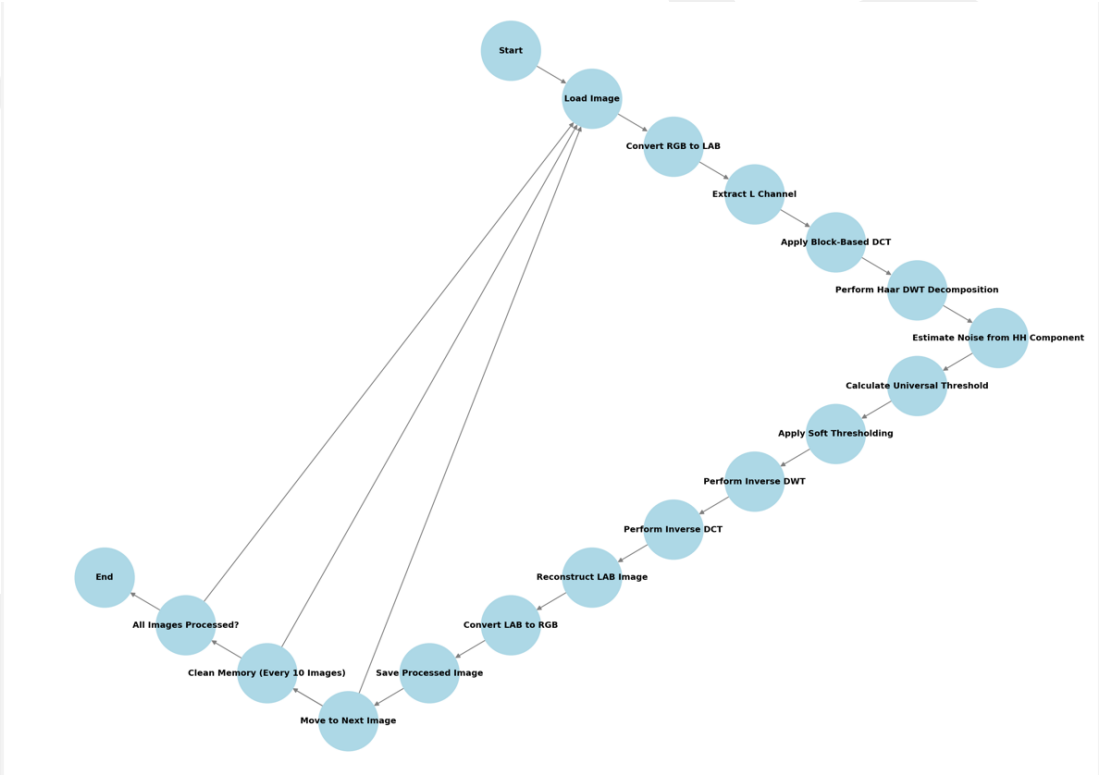


Figure 3.20 Discrete Cosine Transform(DCT) – Discrete Wavelet Transform (DWT) Pipeline

CHAPTER 4

METHODS

4.1 Transfer Learning Approach in Deep Learning and Classification

This method has been effectively utilized across diverse domains including computer vision, natural language processing, and healthcare, and has shown significant versatility and efficacy. As the intricacy of deep learning models escalates, transfer learning presents a promising direction for enhancing model performance and scalability within classification frameworks [63] – [64].

A notable theoretical model in the realm of transfer learning, particularly within deep learning, highlights the notion of domain adaptation, which pertains to the capacity of a model to extend its generalization from the source domain to the target domain by diminishing the distributional discrepancies between these two domains. This process is fundamentally significant for affirming that the knowledge acquired from the source domain can be proficiently conveyed to the target domain. An additional critical theoretical component is the fine-tuning procedure, which involves the modification of weights only in the uppermost layers of the neural architecture while the lower layers remain static. This method facilitates the adjustment of the model to a new task, while preserving the insights gained from the antecedent task. Moreover, the decision regarding the layers to be transferred is of paramount importance in the field of transfer learning. Empirical studies have indicated that features extracted from the lower layers of deep neural networks display greater generalizability across various tasks in contrast to the features derived from the higher layers, which are often more aligned with specific tasks. Consequently, the selection of appropriate layers for knowledge transfer is a pivotal factor that influences the effectiveness of transfer learning methodologies within deep-learning frameworks [63]- [64].

A significant benefit associated with transfer learning within classification endeavors is the capacity to utilize insights from analogous domains or undertakings to bolster the efficacy of a primary task characterized by a scarcity of labeled data. This is

particularly advantageous in contexts where acquiring extensive labeled data is both expensive and time consuming. By effectively migrating learned representations from a source domain to a target domain, transfer learning permits proficient training of deep learning models, negating the need to commence from a blank slate. Furthermore, the evidence suggests that transfer learning enhances the generalization capabilities and adaptability of models to novel data distributions, rendering it a useful instrument for practical implementation. Additionally, it may alleviate complications related to overfitting and augment the overall resilience of classification models through the integration of knowledge derived from pre-trained networks. Hence, the integration of transfer learning methodologies in classification tasks presents considerable potential for improving model efficacy and operational efficiency[63] – [64].

4.2 Fully Connected Layers in Image Classification

Within the domain of image classification, the role of fully connected layers is crucial for elevating both the accuracy and feature representation. As articulated by Wagh [65], adjustments made to Convolutional Neural Network (CNN) architectures that integrate specialized fully connected layers materially elevate classification outcomes in the context of satellite-based geographic area classification. The fusion of these layers with pooling and traditional layers enables the efficient capture and analysis of complex spatial patterns, thereby yielding more precise classification results across various geographical terrains. Moreover, the application of a pre-trained CNN model that incorporates altered fully connected layers, as indicated by Li et al.[66], serves not only enhances the extraction of features and arranges image feature vectors within a homomorphic bitmap index structure. This methodology boosts search efficiency while affirming the robustness and representativeness of the extracted features, thereby highlighting the importance of fully connected layers in the optimization of image classification methodologies.

Deep learning techniques, with particular emphasis on convolutional neural networks (CNNs), have revealed considerable achievements in the enhancement of image classification. Regarding the classification of lung cancer, alterations made to fully connected layers within CNN architectures have shown promise in bolstering the

performance. As noted by Benamara et al. [67], the deployment of deep modified CNN networks, especially those that employ pre-trained models, such as Densenet169, has resulted in augmented accuracy rates for lung cancer classification, thereby demonstrating the potential advantages of adapting network configurations for particular tasks. Additionally, Fei et al.[68] underscored the criticality of efficient and lightweight frameworks in the realm of image classification, emphasizing techniques for dimensionality reduction and adjustments to established architectures, such as VGG, in the context of hyperspectral image classification. This notion is congruent with the idea that specifically tailored modifications to fully connected layers can proficiently optimize classification results across varying image analysis contexts, underscoring the importance of customization and creativity in the design of networks to achieve enhanced outcomes.

4.3 K Fold Cross Validation in Image Classification

Within the domain pertaining to image classification, K Fold Cross Validation remains paramount for evaluating the efficacy of machine learning models. This methodological approach involves partitioning the dataset into K distinct subsets, which are commonly referred to as folds. In this schema, K-1 of these subsets is employed for the training phase, while the remaining subset is utilized for validation purposes. This cyclical process was performed K times, ensuring that each fold was utilized as the validation set on a singular occasion, thus facilitating a more comprehensive and robust approximation of the performance of the models and mitigating the probabilities of overfitting or underfitting phenomena. The application of K Fold Cross Validation enables us to scrutinize the generalization capacity of the models and pinpoint potential complications such as variance or bias. Furthermore, this technique provides more dependable metrics for the comparative analysis of diverse models, aiding in the selection of the most suitable model for subsequent deployment. In summary, K Fold Cross Validation is indispensable for enhancing the accuracy and dependability of image classification models.

Furthermore, it is pertinent to acknowledge that K Fold Cross Validation presents an array of advantages in the classification domain. A primary benefit resides in its

capacity to optimize data utilization, as it employs each data point for both training and testing purposes, which consequently mitigates the potential for overfitting. In addition, K Fold Cross Validation yields a more precise approximation of model performance when juxtaposed with a solitary train-test split, as it aggregates the evaluation metrics across several iterations. This methodology culminates in results that are deemed more dependable, and assists in discerning the optimal deployment model. Nonetheless, it is imperative to consider certain limitations that are endemic to the application of the K-fold cross-validation. One disadvantage pertains to augmented computational demands given that the model must be fitted on multiple occasions. Furthermore, K Fold Cross Validation may exhibit suboptimal performance with smaller datasets, wherein the partitioning of data into folds could result in a scenario in which each individual fold fails to adequately represent the overarching dataset. Despite these constraints, the advantages attributed to K Fold Cross Validation frequently eclipse its disadvantages, thereby rendering it an invaluable instrument for classification tasks [69].

4.4 Transfer Learning Between Folds in K Fold Cross Validation

Comprehending the transfer learning part in K-fold cross-validation requires examining how knowledge is used across separate data portions. The fundamental process is to apply pre-trained models that possess a wide-ranging comprehension of the target task, which is subject to adjustment during the training segment of each fold. Surprisingly, engaging in K-fold cross-validation not only helps generalize these models by testing them on different data subsets but also aids in honing their prediction abilities. When a repeated approach is used, the information acquired from one fold can appropriately bolster the training of the following folds, initiating a loop of ongoing learning that strengthens the overall strength of the model. This knowledge sharing guarantees that the model modifies itself to fit variations and detailed features present in the various training segments, thereby reinforcing its usefulness and efficiency in real-world situations [70].

The effectiveness of transfer learning when applied across folds in the K-fold cross-validation has significant consequences for both the generalizability and efficiency of

the models. By utilizing the knowledge gained from one fold to bolster learning in another, models can attain increased performance while necessitating a diminished total amount of data, which is particularly important in contexts where labeled datasets are sparse. This methodology not only speeds up the training process but also assists in reducing overfitting because the model acquires a more expansive viewpoint from various data subsets, which results in a more resilient comprehension. In addition, the incorporation of transfer learning frameworks enables researchers to pursue novel applications spanning a range of domains, including but not limited to natural language processing and computer vision, where specific knowledge related to tasks can be effectively amalgamated. This method contributes to a learning environment that adapts more proficiently to contextual subtleties, subsequently uncovering opportunities to improve performance metrics across diverse folds. Consequently, the implementation of transfer learning represents a pivotal transition towards more effective strategies for model training within the realm of machine learning research [70].

4.5 Data Augmentation in Histopathological Image Classification

Within the realm of histopathology, a number of data augmentation methods are utilized to amplify the diversity present in training datasets, which can, in turn, lead to enhancements in model performance. A commonly employed technique involves resizing images to predetermined dimensions to ensure that the input sizes remain consistent across various samples. This uniformity facilitates a more effective training and inference. Moreover, tactics such as RandomHorizontalFlip and RandomRotation apply crucial variations by altering the orientations of images, thus enabling models to achieve better generalization to different anatomical orientations and perspectives. In addition, the application of random affine transformations can mimic minor shifts in the positioning or scaling of tissue sections, consequently enriching the dataset with a wide array of representations of the same pathological traits. Furthermore, ColorJitter may be used to introduce inconsistencies in the staining intensity and color of tissues, reflecting the inherent variations in histopathological staining methodologies. When combined, these augmentation strategies work collaboratively to bolster the robustness and reliability of classification paradigms within histopathology [73] – [74].

The utilization of methodologies such as random horizontal flipping and rotation notably boosts the resilience of histopathological image classification systems. The process of data augmentation by artificially expanding the training dataset permits models to assimilate invariant features that exhibit improved generalizability to data that have not been previously encountered. An illustrative example involves random affine transformations that alter aspects, such as image scale, perspective, and skew, which may enable the model to develop increased resistance to fluctuations in image capture conditions, thereby mitigating instances of overfitting. Furthermore, the incorporation of color jitter has the potential to enhance model efficacy by adjusting for disparities in staining and lighting that could manifest within histological slides. Collectively, these strategies contributed to the formation of a more heterogeneous dataset, thereby endowing the model with an extensive array of scenarios to confront during the training phase, ultimately culminating in enhanced performance metrics, including accuracy and precision, when evaluated on the test datasets. The aggregate results of these methodologies not only amplify performance, but also bolster model stability in clinical contexts [74] – [75].

Data augmentation techniques play a significant role in the improvement of model precision and resilience in the classification of histopathological images. By implementing transformations such as resizing, random horizontal flips, and random rotations, researchers can substantially enhance the variability of training datasets, which subsequently diminishes overfitting and fosters better generalization to data that has not been seen before. For example, in various studies, such as the one that discussed U-net usage for segmentation and SVM for the classification of features, using augmented data allowed the model to reach a noteworthy accuracy of 98.5% in image segmentation in biomedical studies [71]. Moreover, the adoption of more sophisticated augmentation techniques can enhance the predictive performance of deep learning models, as evidenced by studies evaluating microvascular invasion in HCC cases of hepatocellular carcinoma. In this context, a comprehensive supervised deep learning method demonstrated considerable predictive capabilities, highlighting the advantages of diverse training datasets [72]. These results demonstrate that proficient data augmentation is crucial for establishing models that can deliver reliable and precise diagnostics in the field of histopathology [76].

4.6 Evaluation Metrics in Deep Learning

Within the domain of metrics for evaluation pertaining to deep learning, a significant apparatus for the appraisal of model performance is referred to as the confusion matrix. This matrix provides a thorough dissection of classifications that include true-positives, false-positives, true-negatives, and false-negatives. Each facet encapsulated within the matrix delineates particular features of model precision, thereby empowering researchers to extract crucial metrics such as precision and recall. To elucidate further, precision signifies the fraction of true positive outcomes in relation to the totality of positive predictions, thereby illustrating the fidelity of the positive classifications. On the other hand, recall reflects the aptitude of a model in discerning all pertinent instances residing in a dataset. The interplay between these two metrics is frequently a matter of tradeoff, which prompts the need for a certain equilibrium that can be proficiently assessed using the F1 Score. This score integrates the precision and recall into a cohesive singular metric. Furthermore, the application of the Kappa Score facilitates a deeper comprehension of model dependability by factoring in the agreement that transpires purely by chance, thereby rendering the confusion matrix an indispensable element in the quest to optimize outcomes in deep learning [78] – [79].

Within the domain of model evaluation, a detailed comprehension of metrics such as accuracy, precision, and recall is important for differentiating the effectiveness of predictive algorithms. Accuracy functions as a basic metric, indicating the ratio of true results, consisting of both true positives and true negatives, to the total number of assessed cases. Nevertheless, depending solely on accuracy can lead to misconceptions, particularly when addressing imbalanced datasets where a high level of accuracy may obscure inadequate performance in minor classes. Precision assesses the correctness of the positive predictions rendered by the model, whereas recall measures the model's capability to recognize all pertinent instances within the dataset. The relationship between these metrics is frequently encapsulated by the F1 Score, which integrates precision and recall into a cohesive measurement for equitable evaluation, along with the Kappa Score, which considers the concordance between predicted and actual values, thus providing more in-depth insights into the model performance [78] – [79].

Within the domain of assessment pertaining to the performance of models in machine learning, metrics such as the F1 Score and Kappa Score exist, which provide essential insights that transcend basic accuracy. Accuracy: Although it may provide a somewhat superficial overview of model efficacy, it does not adequately address performance in instances of imbalanced datasets where particular classes may skew the outcomes. The F1 Score, which amalgamates precision and recall, counters this limitation by underscoring the model's competence in correctly spotting pertinent instances without allowing them to be misled by the occurrence of false positives. For example, in investigations related to the detection of emotions in languages, including Bengali and Banglish, a reported F1 Score of 71.30% indicates a nuanced classification proficiency that goes beyond straightforward accuracy metrics [77]. Furthermore, the Kappa Score is an instrumental metric for assessing the degree of inter-rater agreement, yielding a robust analysis of the reliability of the model. A case in point is found in the examination of neonatal sleep patterns, where a Kappa Score of 0.65 suggests a significant degree of concordance between classifications made by automated systems and those rendered by experts, thereby affirming the model's effectiveness [78]. Consequently, the amalgamation of these metrics cultivates a well-rounded comprehension of model performance, thus enabling better decision-making in deep learning applications.

The formulas of accuracy, precision, recall, F1-Score, Cohen's Kappa, Matthew Correlation Coefficient(MCC) are given as follows [78] – [79]:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{4.1}$$

$$Recall = \frac{TP}{TP+FN} \tag{4.2}$$

$$Precision = \frac{TP}{TP+FP} \tag{4.3}$$

$$F1 = \frac{2xPrecisionxRecall}{Precision+Recall} \tag{4.4}$$

$$Kappa = \frac{Po-Pe}{1-Pe} \quad (4.5)$$

$$MCC = \frac{TPxTN-FPxFN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (4.6)$$

TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative

P_o: Observed Agreement, P_e: Expected Agreement

4.7 Proposed Model

This section presents a comprehensive analysis of the classification methodology employed for multiclass lung histopathological images to ensure that the code and its explanation align with the conventions and structure of a formal thesis. This approach integrates Vahadane-based stain normalization, ConViT architecture, and K-fold cross-validation to achieve robust and reliable classification results across three distinct classes of lung images: adenocarcinoma (lung_aca), normal cells (lung_n), and Squamous Cell Carcinoma (lung_scc).

4.7.1 Data Usage and Preparation

The dataset utilized in this study comprised three distinct directories, each corresponding to a specific lung class. To maximize the data utilization, the entire set of images within each directory was employed for training and validation purposes. To ensure consistency across folds, each image was assigned a label according to its class and was accompanied by a group identifier for cross-validation. This methodological approach ensures that images from the same group do not appear simultaneously in either the training or validation subsets, thereby mitigating the risk of data leakage.

4.7.2 Group Assignment for K-Fold Cross-Validation

A K-fold (in this case, k=5) cross-validation strategy was implemented to evaluate the model's performance systematically. The entire dataset was randomly shuffled and divided into five folds of approximately equal sizes. In the cross-validation process,

each fold was systematically employed as the validation dataset, while the amalgamation of the remaining folds functioned as the training corpus. To maintain independence between the training and validation samples, images belonging to the same group were not distributed across the folds. This approach ensures a more robust assessment of the generalization capabilities of the model.

4.7.3 Augmentation Techniques

To enhance the robustness of the model and avoid overfitting, data augmentation was applied to the original images:

Resizing: All the input TIFF images were resized to 224×224224 \times 224 pixels, matching the input requirement of the ConViT architecture.

Random Horizontal/Vertical Flips: Introduce reflections of tissue structures, mimicking variations in how a slide might be placed or viewed.

Color Jitter: Adjustments in brightness, contrast, and saturation simulate realistic stains or lighting changes and improve generalization.

Normalization: Pixel values were normalized using the standard ImageNet mean and standard deviation, aligned with the pre-trained weights used in the ConViT model.

These augmentation steps were selectively applied during training. For validation, only a resizing and normalization step was applied to maintain consistency when comparing the predicted results to the ground truths.

4.7.4 Model Architecture

The Convolution-Free Transformer (ConViT) model, obtained through the PyTorch Image Models (TIMM) library, is at the core of this classification pipeline. This architecture employs attention mechanisms, rather than traditional convolutions, allowing it to capture complex textures and features in histopathological images.

Pre-trained Backbone: ConViT is initialized with ImageNet weights, granting the model a strong representational foundation even when the lung image dataset is smaller than typical ImageNet-scale data.

Custom Classification Head: The final layer of the ConViT model was replaced with a fully connected (linear) layer for the three-class output. This adjustment enabled the network to output separate logits for lung_aca, lung_n, and lung_scc.

Such a transfer learning approach accelerates convergence and promotes more effective feature extraction, particularly for specialized medical imaging tasks.

4.7.5 Training and Validation

A standardized pipeline governs how data are fed into the model and how the model parameters are updated.

Training Loop:

- **Optimizer:** AdamW was employed, with a low learning rate (3×10−53 \times 10^{-5}). This optimizer provides stable updates for fine-tuning pre-trained networks and balances the adaptive learning rates with weight decay.
- **Loss Function:** Cross-entropy loss is used for multiclass classification, which measures the discrepancy between predicted logits and ground-truth labels.
- **Mixed Precision:** Automatic mixed precision (AMP) scales numerical values during backpropagation, reducing GPU memory usage and the computational overhead.
- **Learning Rate Scheduler:** A ReduceLROnPlateau schedule cuts the learning rate if the validation loss plateaus, mitigating overfitting, and refining convergence.

Validation Loop:

- **Optimizer:** AdamW is employed with a low learning rate (3×10−53 \times 10^{-5}). This optimizer provides stable updates for fine-tuning pre-trained networks, balancing adaptive learning rates with weight decay.

- **Loss Function:** Cross-entropy loss is used for multi-class classification, measuring the discrepancy between predicted logits and ground-truth labels.
- **Mixed Precision:** Automatic mixed precision (AMP) scales numerical values during backpropagation, reducing GPU memory usage and computational overhead.
- **Learning Rate Scheduler:** A ReduceLROnPlateau schedule cuts the learning rate if the validation loss plateaus, mitigating overfitting and refining convergence.
- **Accuracy and Loss:** After epoch, the model was evaluated on the unseen validation fold. The mean validation loss and accuracy were tracked and used to determine whether the model had improved.
- **Cohen's Kappa and Matthews Correlation Coefficient (MCC):** These metrics offer insights beyond simple accuracy, providing a better understanding of performance especially when class imbalances or multi-class nuances are present.
- **Best Epoch Selection:** If multiple epochs achieve similar validation accuracy, the epoch with the lower validation loss is favored. This two-tier criterion ensures both a high accuracy and stable performance.

By iterating this procedure over all kk folds, the consistency and reliability of the model were thoroughly examined across different subsets of the dataset.

4.7.6 Evaluation Metrics

Performance is quantified using various metrics that collectively provide a well-rounded assessment of model effectiveness.

Classification Report: Displays precision, recall, and F1-score for each class, highlighting the model's ability to correctly classify each lung condition.

Confusion matrix: A visual comparison of predicted labels versus ground truths is useful for identifying any significant misclassifications or confusion between classes.

ROC AUC Analysis: ROC curves were plotted for each class by comparing true-positive rates against false-positive rates across threshold variations. The Area Under the Curve (AUC) quantitatively measures the separability of each class.

These metrics were computed at both the per-and overall levels. Fold-by-fold results identify how the model performs under different training validation splits, while aggregated statistics present a comprehensive view of its generalization. In addition, the entire Vahadane-based standardization, ConViT architecture, and K-fold cross-validation form a robust framework for reliably classifying histopathological lung images.

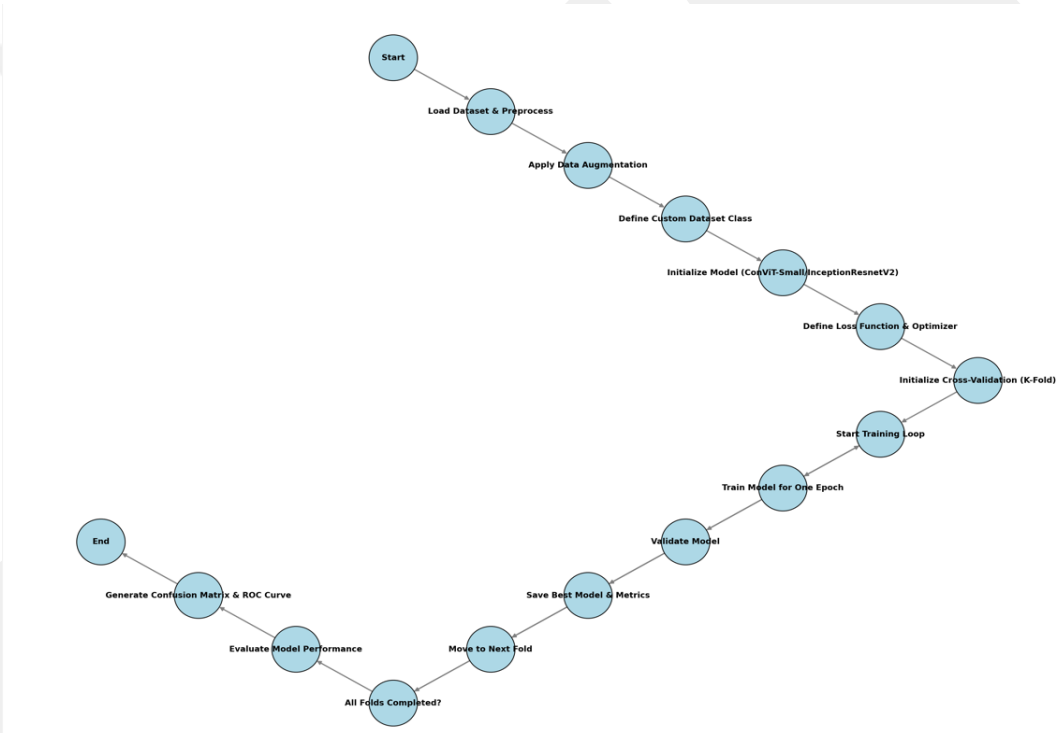


Figure 4.1 K-Fold Image Classification Pipeline

CHAPTER 5

RESULTS AND DISCUSSION

In this chapter, we present the results obtained by our model by comparing them with other methods in the literature, and then discuss our model and its outcomes.

5.1 Evaluation of Classification Metrics

Based on the results of our model, we used the loss and accuracy metric, F1 score, precision, recall, Cohen's Kappa, Matthew Correlation coefficient as evaluation metrics. The multi-classification results in the model are listed in Table 5.1. We can observe that our model achieved outstanding performance in the LUAD(lung adenocarcinoma) vs. LUSC(squamous cell carcinoma) vs. NC(normal cell) classification tasks, attaining accuracies of 99.84 %, 100 %, and 100% for the training, validation, and testing datasets, respectively. Additionally, the Kappa and MCC metrics reached a maximum value of 1.0000, indicating strong agreement and robust correlation between the predicted and actual labels. These results demonstrate the reliability of the model in differentiating between these classes with high precision.

--- Overall Classification Report ---

	precision	recall	f1-score	support
lung_aca	1.00	1.00	1.00	4727
lung_n	1.00	1.00	1.00	4744
lung_scc	1.00	1.00	1.00	4724
accuracy			1.00	14195
macro avg	1.00	1.00	1.00	14195
weighted avg	1.00	1.00	1.00	14195

--- Overall Classification Report ---

	precision	recall	f1-score	support
lung_aca	1.00	1.00	1.00	4727
lung_n	1.00	1.00	1.00	4744
lung_scc	1.00	1.00	1.00	4724
accuracy			1.00	14195
macro avg	1.00	1.00	1.00	14195
weighted avg	1.00	1.00	1.00	14195

Figure 5.1 Results of the Overall Classification Report for both InceptionResnetV2 and Convit_Small with Original Images

At the end of all folds, flawless overall classification report values were obtained for the classification performed on the original dataset images using the InceptionResNetV2 and ConViT_Small models.

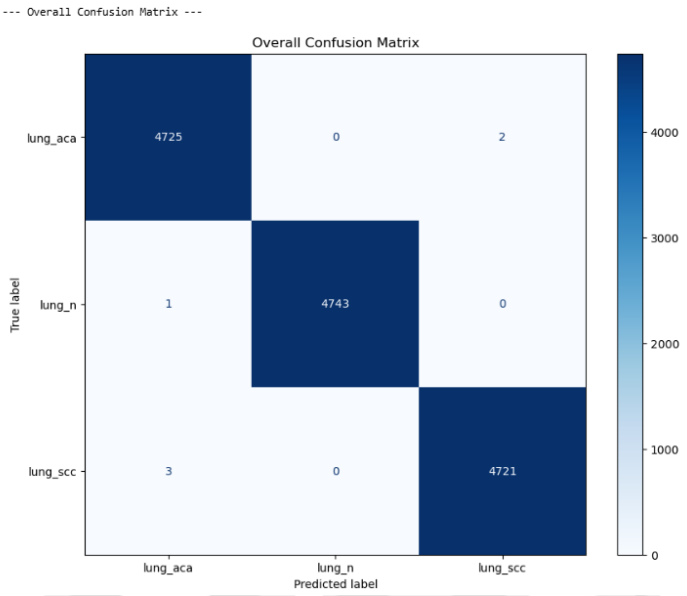


Figure 5.2 Overall Confusion Matrix of InceptionResnetV2 with Original Images

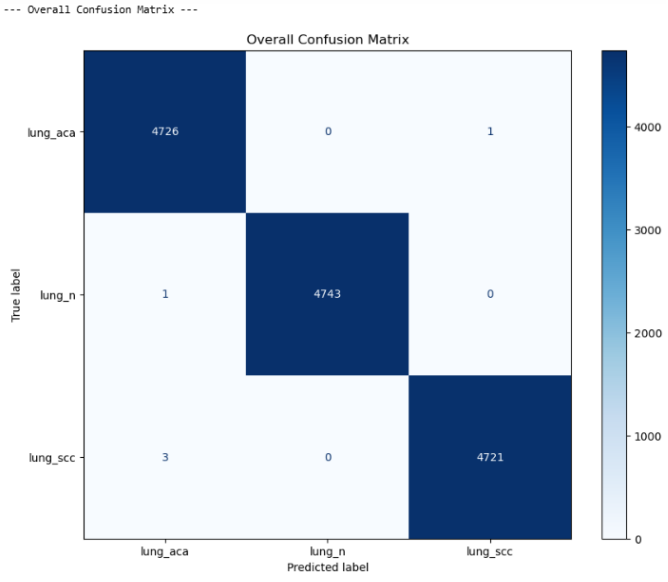


Figure 5.3 Overall Confusion Matrix of Convit_ Small with Original Images

Based on the overall confusion matrices, the InceptionResNetV2 model resulted in only 6 misclassifications at the end of all folds, while the ConViT_Small model had only 5 misclassifications.

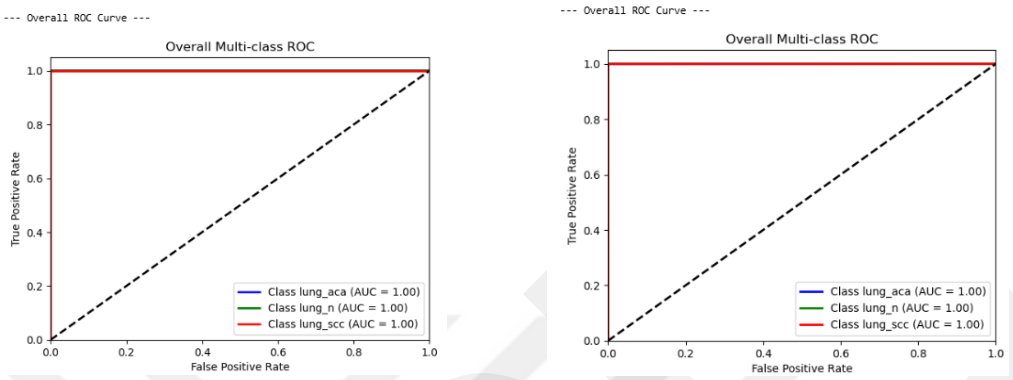


Figure 5.4 Results of the Overall Multi-Class ROC AUC Graph for both InceptionResnetV2 and Convit_ Small with Original Images

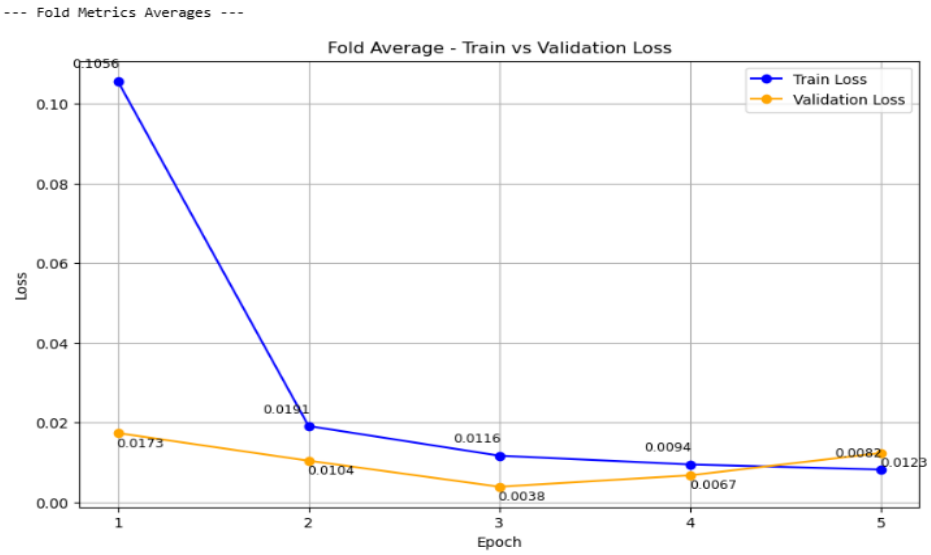


Figure 5.5 Overall Fold Average of Train vs. Validation Loss of InceptionResnetV2 with Original Images

The Multi-Class ROC AUC graphs also appear to be flawless, similar to the classification report values. However, based on the average loss and accuracy graphs, although the values seem good, they are not flawless and do not exhibit the desired balance.

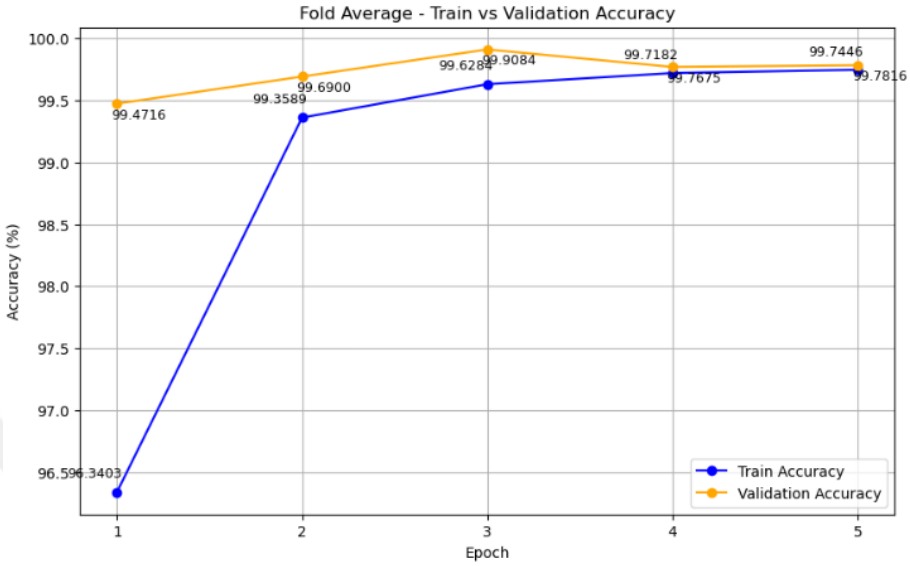


Figure 5.6 Overall Fold Average of Train vs. Validation Accuracy of InceptionResnetV2 with Original Images

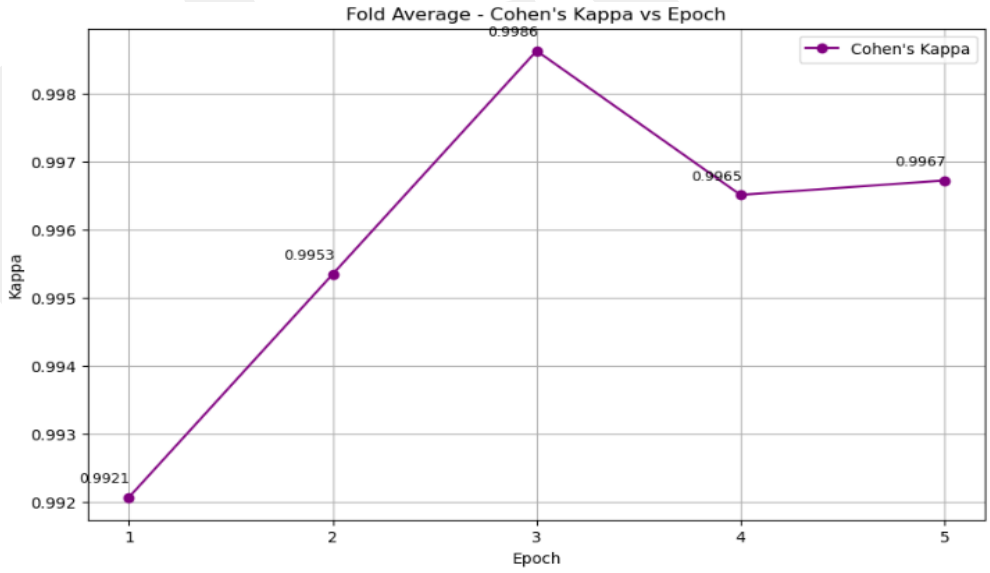


Figure 5.7 Overall Fold Average of Cohen's Kappa of InceptionResnetV2 with Original Images

This imbalance can also be observed in the average graphs of Cohen's Kappa and the Matthews Correlation Coefficient, despite their very high values.

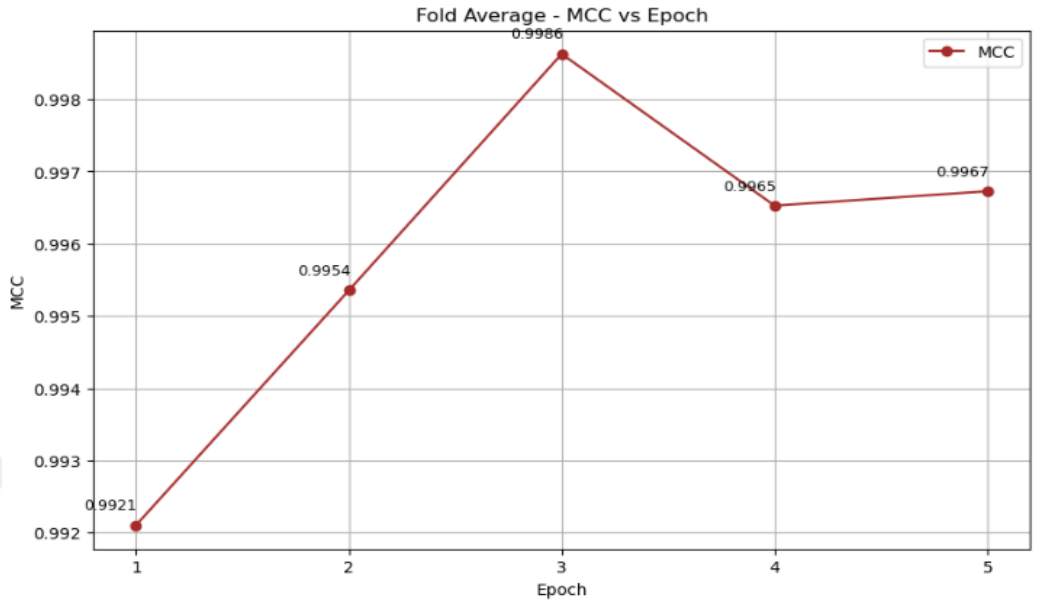


Figure 5.8 Overall Fold Average of Matthew Correlation Coefficient(MCC) of InceptionResnetV2 with Original Images

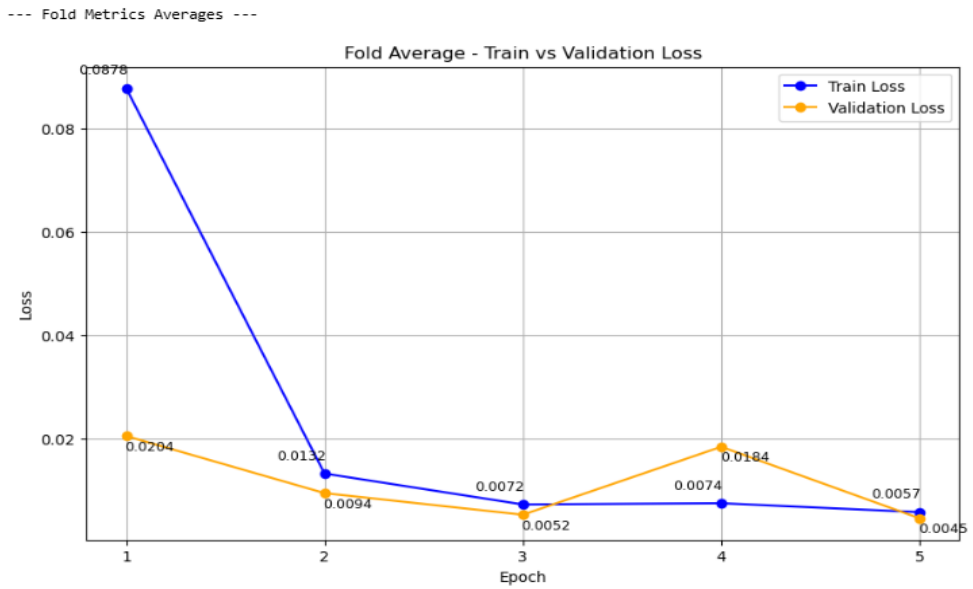


Figure 5.9 Overall Fold Average of Train vs. Validation Loss of Convit_Small with Original Images

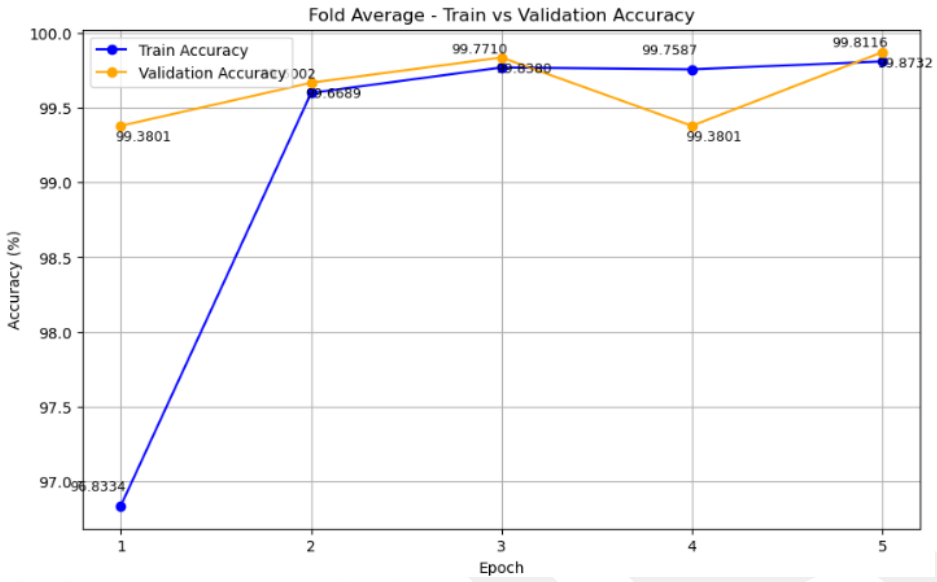


Figure 5.10 Overall Fold Average of Train vs. Validation Accuracy of Convit_Small with Original Images

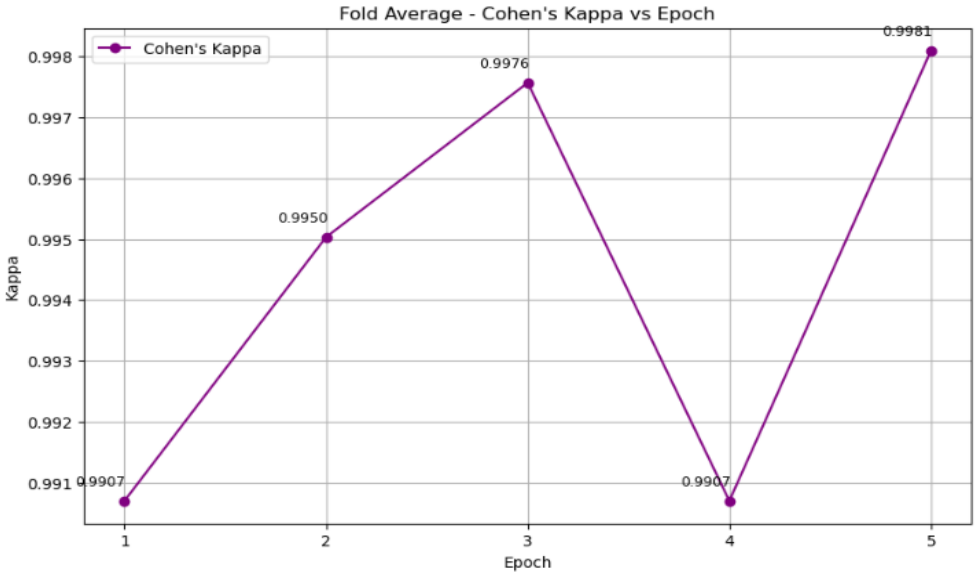


Figure 5.11 Overall Fold Average of Cohen's Kappa of Convit_Small with Original Images

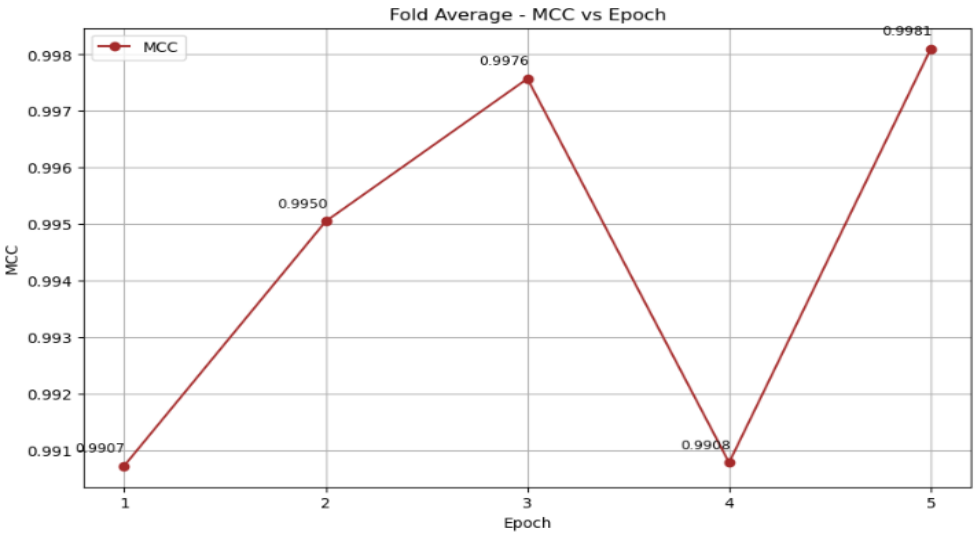


Figure 5.12 Overall Fold Average of Matthew Correlation Coefficient(MCC) of Convit_Small with Original Images

As seen, the undesired imbalances present in the graphs of InceptionResNetV2 are also observed in the loss, accuracy, Cohen's Kappa, and MCC graphs of the ConViT_Small model.

```
--- Overall Classification Report ---
                                precision    recall  f1-score   support

vahanade_lung_aca_tiff_pil_+_cv2      1.00      1.00      1.00      4726
                          lung_n      1.00      1.00      1.00      4744
vahanade_lung_scc_tiff_pil_+_cv2      1.00      1.00      1.00      4722

                        accuracy                           1.00     14192
                       macro avg      1.00      1.00      1.00     14192
                    weighted avg      1.00      1.00      1.00     14192
```

Figure 5.13 Results of the Overall Classification Report for Convit_Small with Vahanade Color Transformation Images

Since it demonstrated a similar performance to the remaining classification processes and was more cost-effective compared to InceptionResNetV2, the ConViT_Small model was chosen for further analysis.

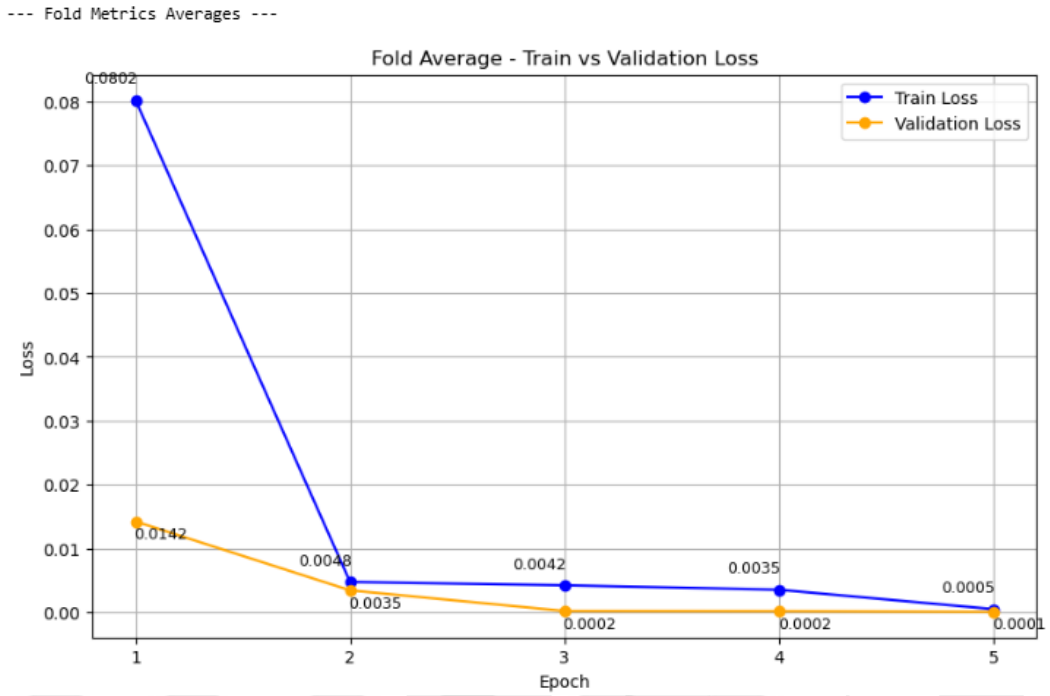


Figure 5.14 Overall Fold Average of Train vs. Validation Loss of Convit_Small with Vahanade Color Transformation Images

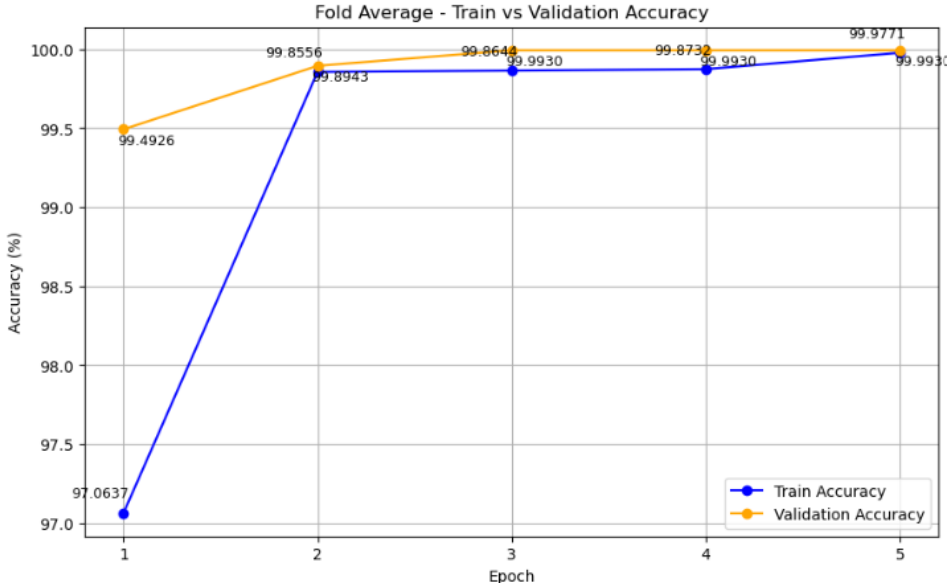


Figure 5.15 Overall Fold Average of Train vs. Validation Accuracy of Convit_Small with Vahanade Color Transformation Images

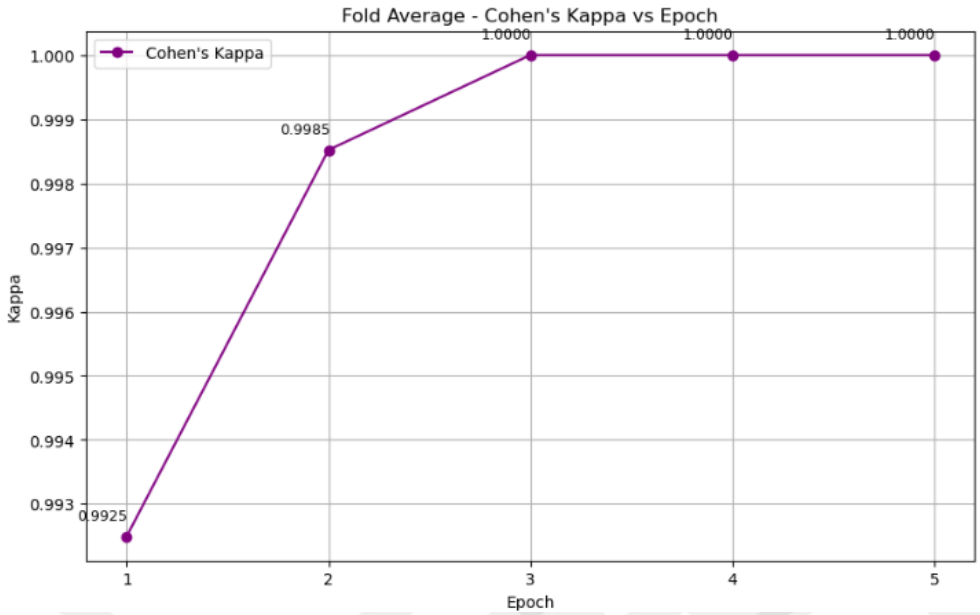


Figure 5.16 Overall Fold Average of Cohen's Kappa of Convit_Small with Vahanade Color Transformation Images

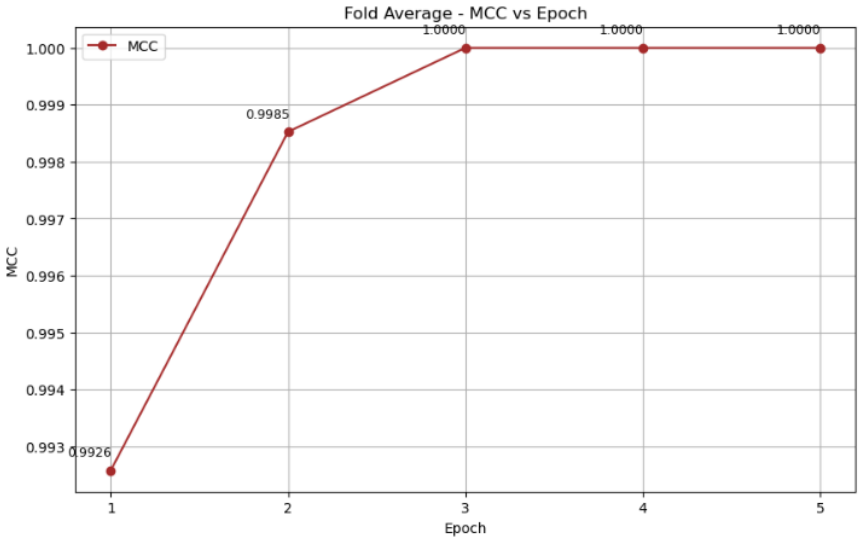


Figure 5.17 Overall Fold Average of Matthew Correlation Coefficient(MCC) of Convit_Small with Vahanade Color Transformation Images

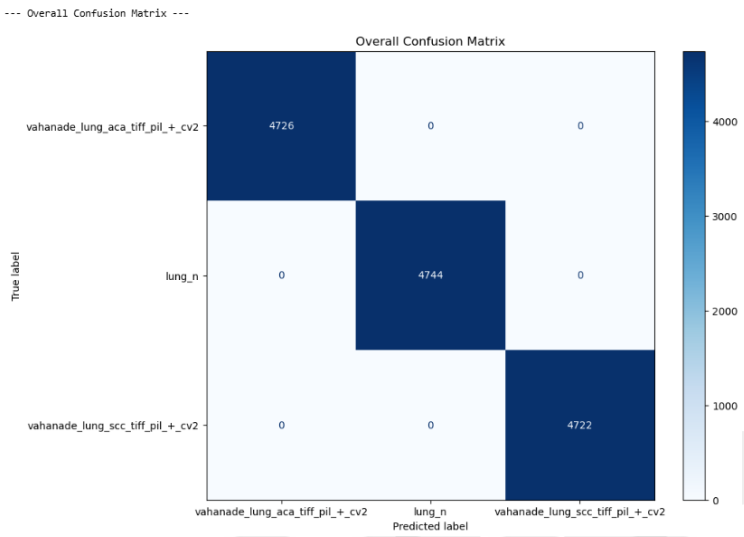


Figure 5.18 Overall Confusion Matrix of Convit_ Small with Vahanade Color Transformation Images

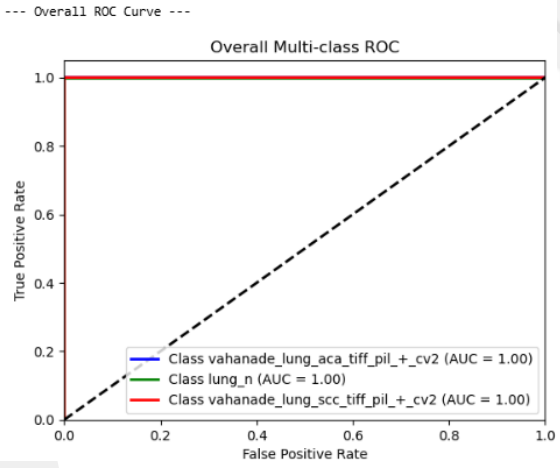


Figure 5.19 Results of the Overall Multi-Class ROC AUC Graph for Convit_ Small with Vahanade Color Transformation Images

When the Vahanade color transformation, a type of staining normalization, was applied to the original dataset images, the results improved surprisingly well. As expected, the classification report and Multi-Class ROC AUC graph remained flawless. However, the unique improvement is evident in the other metrics and graphs.

The Confusion Matrix shows zero misclassifications. The loss and accuracy graphs are extremely balanced and smooth, and despite representing fold average values, the loss reached a perfect value of 0.0001, while the accuracy reached an almost flawless value of 99.9930. Additionally, in the Cohen's Kappa and MCC graphs, the last 3-epoch average ended at 1.00, reflecting a highly stable and flawless result.

Fold Average Metrics	inceptionresnetv2 with original images	convit small with original images	only reinhard applied images	only vahanade applied images	vahanade+ dct-dwt	dct-dwt+vahanade	vahanade+ ghe	vahanade +clahe with a. sharpenin	dct-dwt+clahe with a. sharpening+vahanade
Train Loss (Epoch 1)	0,1056	0,0878	0,0932	0,0802	0,0882	0,0806	0,0956	0,0879	0,0881
Train Loss (Epoch 2)	0,0191	0,0132	0,0146	0,0048	0,0053	0,0074	0,0167	0,0133	0,0124
Train Loss (Epoch 3)	0,0116	0,0072	0,0084	0,0042	0,0095	0,0093	0,0060	0,0139	0,0121
Train Loss (Epoch 4)	0,0094	0,0074	0,0045	0,0035	0,0010	0,0045	0,0081	0,0048	0,0031
Train Loss (Epoch 5)	0,0123	0,0057	0,0017	0,0005	0,0010	0,0004	0,0161	0,0026	0,0003
Validation Loss (Epoch 1)	0,0173	0,0204	0,0240	0,0142	0,0087	0,0078	0,0254	0,0090	0,0185
Validation Loss (Epoch 2)	0,0104	0,0094	0,0106	0,0035	0,0005	0,0013	0,0197	0,0040	0,0072
Validation Loss (Epoch 3)	0,0038	0,0052	0,0092	0,0002	0,0010	0,0307	0,0170	0,0042	0,0031
Validation Loss (Epoch 4)	0,0067	0,0184	0,0107	0,0002	0,0002	0,0007	0,0778	0,0055	0,0007
Validation Loss (Epoch 5)	0,0082	0,0045	0,0053	0,0001	0,0001	0,0001	0,0048	0,0005	0,0011
Train Accuracy (Epoch 1)	96,3403	96,8334	96,5727	97,0637	96,8347	96,9402	96,5579	96,7552	96,7724
Train Accuracy (Epoch 2)	99,3589	99,6002	99,5210	99,8556	99,8150	99,7551	99,4381	99,5702	99,5689
Train Accuracy (Epoch 3)	99,6284	99,7710	99,7059	99,8664	99,6583	99,7710	99,8186	99,5508	99,5613
Train Accuracy (Epoch 4)	99,7675	99,7587	99,8503	99,8732	99,9560	99,8362	99,7129	99,8238	99,8943
Train Accuracy (Epoch 5)	99,7816	99,8116	99,9313	99,9771	99,9542	99,9894	99,5314	99,9154	99,9806
Validation Accuracy (Epoch 1)	99,4716	99,3801	99,1828	99,4926	99,7041	99,7815	99,1121	99,7181	99,3517
Validation Accuracy (Epoch 2)	99,6900	99,6689	99,6830	99,8943	99,9789	99,9859	99,2813	99,8591	99,7815
Validation Accuracy (Epoch 3)	99,9084	99,8380	99,7112	99,9930	99,9507	98,9006	99,3871	99,9084	99,8802
Validation Accuracy (Epoch 4)	99,7182	99,3801	99,6055	99,9930	99,9718	99,9648	97,8083	99,8362	99,9713
Validation Accuracy (Epoch 5)	99,7446	99,8732	99,8168	99,9930	99,9789	100,0000	99,8379	99,9859	99,9366
Cohen's Kappa Score (Epoch 1)	0,9921	0,9907	0,9879	0,9925	0,9959	0,9967	0,9868	0,9958	0,9905
Cohen's Kappa Score (Epoch 2)	0,9953	0,9950	0,9955	0,9985	1,0000	0,9998	0,9893	0,9979	0,9969
Cohen's Kappa Score (Epoch 3)	0,9986	0,9976	0,9959	1,0000	0,9996	0,9835	0,9909	0,9986	0,9984
Cohen's Kappa Score (Epoch 4)	0,9965	0,9907	0,9943	1,0000	0,9999	0,9995	0,9671	0,9974	0,9998
Cohen's Kappa Score (Epoch 5)	0,9967	0,9981	0,9975	1,0000	1,0000	1,0000	0,9977	0,9998	0,9993
MCC Score (Epoch 1)	0,9921	0,9907	0,9881	0,9926	0,9959	0,9967	0,9869	0,9958	0,9906
MCC Score (Epoch 2)	0,9954	0,9950	0,9955	0,9985	1,0000	0,9998	0,9895	0,9979	0,9969
MCC Score (Epoch 3)	0,9986	0,9976	0,9959	1,0000	0,9996	0,9843	0,9910	0,9986	0,9984
MCC Score (Epoch 4)	0,9965	0,9908	0,9943	1,0000	0,9999	0,9995	0,9699	0,9974	0,9998
MCC Score (Epoch 5)	0,9967	0,9981	0,9975	1,0000	1,0000	1,0000	0,9977	0,9998	0,9993
ROC AUC Score (All Classes)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Precision (All Classes)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Recall (All Classes)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
F1 Score (All Classes)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Number of misclassified images (CM)	6	5	2	0	0	0	13	1	2

Figure 5.20 All 5 Folds Average Results for each Epochs

Inceptionresnetv2 with original images: The original images of the dataset were classified using InceptionResNetV2.

Convit_Small with original images: The original images of the dataset were classified using Convit_Small.

Only reinhard applied images: Only reinhard colour transformation was applied to the original dataset and classified using Convit_Small.

Only vahanade applied images: Only vahanade colour transformation was applied to the original dataset and classified using Convit_Small.

Vahanade+dct-dwt: After dct-dwt digital image processing was applied to the original dataset images, vahanade colour transformation was applied and classified using Convit_Small.

Dct-dwt+vahanade: After vahanade colour transformation was applied to the original dataset images, dct-dwt was applied and classified using Convit_Small.

Vahanade+ghe: After applying global histogram equalization to the original dataset images, Vahanade colour transformation was applied and classified using Convit_Small.

Vahanade+clahe with a. sharpening: After applying contrast limited adaptive histogram equalization with adaptive sharpening to the original dataset images, Vahanade colour transformation was applied and classified using Convit_Small.

Dct-dwt+clahe with a. sharpening+vahanade: Firstly, vahanade colour transformation process, secondly clahe with adaptive sharpening process and finally dct-dwt process were applied to the original dataset images and classified using Convit_Small.

CM: Confusion Matrix

The values in the columns “only reinhard applied images”, “vahanade+clahe with a. sharpening”, and “dct-dwt+clahe with a. sharpening+vahanade”, as well as the CM values in the last row, can be seen to yield better results compared to the classification performed on the original images of the dataset.

The values in the “vahanade+ghe” process column can be seen to produce significantly worse results compared to the classification performed on the original images of the dataset.

The best results can be seen in the “only vahanade applied images”, “vahanade+dct-dwt”, and “dct-dwt+vahanade” columns. Additionally, based on the CM value, only these three columns contain processes that yielded error-free results. Moreover, only the “dct-dwt+vahanade” process achieved 100% validation accuracy both in the fold average and across all folds. You can observe this in the tables provided in Appendix E, which contain the metric results for the other folds.

However, based on the table for the 5th fold in Appendix E, an unexpected drop can be seen in the validation accuracy values in the “dct-dwt+vahanade” column, despite reaching 100% accuracy. Therefore, in terms of both reliability and lower processing cost, applying only vahanade colour transformation to the original dataset images appears to be sufficient.

5.2 Comparison with Literature Studies

In this section, we collate our method with a previous study. As presented in Table, our model demonstrates comparable performances against previous studies using the LC25000 dataset for classification of LUAD vs. LUSC vs. NC. As can be seen in Table, by using different modules, our model reached the highest accuracy rate for classification of lung cancer.

Table 5.1 Comparison with other studies

Author [reference]	**Year**	**Dataset**	**Cancer class**	**Method**	**Performance Accuracy (%)**
Pradhan and Sahu [82]	2023	LC25000	Lung cancer	Enhanced grasshopper optimization algorithm (EGOA)	98.50
Sethy et al. [81]	2023	LC25000	Lung cancer	AlexNet, wavelet, and support vector machines	99.30
Halder and Dey[83]	2023	LC25000	Lung cancer	*Morph*ology-based *Atten*tion *N et*work(MorphAttnNet)	98.33
Hatuwal and Thapa [84]	2020	LC25000	Lung cancer	CNN	97.2
Baranwal et al.[85]	2022	LC25000	Lung cancer	Inception-ResNetv2	99.70%
Civit-Masot et al.[86]	2022	LC25000	Lung cancer	Custom CNN	97.11
Noaman et al. [87]	2024	LC25000	Lung cancer	DenseNet-121 + Color Histogram + KNN	99.68
Tian et al.[88]	2024	LC25000	Lung cancer	FPN + SE + ResNet-18	98.84
Mercaldo et al.[89]	2024	LC25000	Lung cancer	VGG-16	99.20
Hamed et al.[90]	2023	LC25000	Lung cancer	Custom CNN + LightGBM	99.60
Wadekar and Singh [91]	2023	LC25000	Lung cancer	VGG-19	97.73
Hadiyoso et al.[92]	2023	LC25000	Lung+ Colon	VGG-16+CLAHE	98.96

Table 5.1 (continued)

Ibrahim and Talaat [93]	2022	LC25000	Lung+ Colon	EfficientNetB7+ DWT with Double-CLAHE	99.5
Chhillar and Singh [94]	2023	LC25000	Lung+ Colon	LightGBM	97.72%, 99.92%, and 100%
Farhadip our[95]	2023	LC25000	Lung+ Colon	DarkNet19+ GCLAHE, VGG19 + GCLAHE	99.71, 99.84
Proposed model	2025	LC25000	Lung cancer	Convit_Small, Inception ResnetV2	100

5.3 Discussion

In this study, the effects of digital image processing techniques and deep learning methods on the classification of lung cancer histopathological images were evaluated. The LC25000 Lung Cancer dataset used in the study was processed using advanced preprocessing techniques such as Vahanade color transfer, DCT-DWT and CLAHE using reference images obtained from the Cancer Imaging Archive. These methods significantly improved the classification accuracy.

Vahanade color transfer provided a consistent analysis environment by normalizing the color changes in the images. This method successfully reduced the variability caused by different staining protocols in the dataset while preserving the structural integrity of the images. In particular, the color transfer was optimized using Non-Negative Matrix Factorization (NMF). In addition, CLAHE increased the local contrast and made histopathological details more distinct.

The transfer learning approach used in this study effectively adapted the pre-trained networks to the LC25000 dataset. Successful results were obtained using InceptionResNetV2 and a visual transform model called ConVit. ConVit Small is a smaller model than InceptionResNetV2, but it provides similar results with a lower computational cost, making it the preferred choice for classification tasks.Compared to the studies in the literature, especially Vahanede color transformation and DCT-DWT digital image processing techniques were not applied to the LC25000 dataset, and only the CLAHE method was the most applied method in the literature.

For example, in the study of Hadiyoso et al. [91], 98.96% accuracy was obtained from 25000 histopathological images using the VGG16 model and CLAHE technique together. In the study of Ibrahim and Talaat [92], 99.5% accuracy was obtained using Double-CLAHE with DWT, EfficientNetB7 Deep learning technique 85:15 hold-out method.

In another study by Chhillar & Singh [93], by applying both hold-out and 10-fold Cross Validation with LightGBM, 97.72%, 99.92%, and 100% accuracy was achieved in color, texture, and combination, respectively, without eliminating duplicate images. In the study by Farhadipour [94], 99.71, 99.84 accuracy was achieved with DarkNet19 + GCLAHE and VGG19 + GCLAHE, respectively, without eliminating duplicate images. As can be seen from the studies in the literature, in many studies, duplicate images were not eliminated and the CLAHE method was frequently used. However, in this study, duplicate images were eliminated, 5-fold Cross Validation was applied, and additionally, different image processing methods such as Vahanade color transformation and DCT-DWT were applied, and higher accuracy than other studies, 100% accuracy in each fold with only 5 epochs, was achieved.

In conclusion, the combination of different image processing techniques and transfer learning increased the generalization ability of the model and provided more consistent and reliable results through this study.

CHAPTER 6

CONCLUSION

This study addresses critical challenges in lung cancer image classification by integrating novel preprocessing methodologies, lightweight deep learning architectures, and advanced image processing techniques. The LC25000 dataset was refined by removing duplicate images, a previously undocumented preprocessing step, resulting in a balanced dataset of 14195 images in three classes (4727, 4744, and 4724). This refinement not only improved the dataset quality but also aimed to contribute to the literature by introducing a reproducible framework for future medical imaging studies.

As a result of the experiments conducted with the InceptionResNetV2 and ConViT-Small models, 6-5 misclassifications were achieved out of 14195 images with fine tuning on and with a single linear layer in the frozen state, respectively, without even using hyperparameter optimization. Among the digital image processing techniques used, especially the Vahanade color transformation technique increased the efficiency and performance of the models and filled the gap in the literature with 0 misclassifications. As can be seen from the results of the study, the Vahanade technique, which was developed to prevent different staining procedures in hospitals, achieved more stable and reliable classification in graphics. This technique was the first study to apply a method that was not previously done by studies using the same dataset in the literature. In addition, it was proven to be better at less cost than studies with 10 and more epochs in the literature, using only 5-fold cross-validation and 5 epochs.

There are certain limitations of the study. First of all, due to the insufficient GPU capacity and hard disk capacity, digital image processing techniques such as super-resolution, which increase the image size and make the details more distinct, could not be applied. Second, XAI methods such as Grad-Cam could not be added to the study.

Third, the models and image processing techniques used in the study could not be applied to data sets other than LC25000.

Future studies will focus on expanding the dataset to further improve model accuracy and applicability, integrating clinical metadata to increase diagnostic sensitivity, and applying multimodal imaging techniques. In addition, incorporating explainable artificial intelligence (XAI) methods will increase the interpretability of model predictions and provide greater confidence among clinicians.

In conclusion, this study highlights the transformative potential of advanced image processing and deep learning for lung cancer histopathological image classification. The findings provide a foundation for further research aimed at improving these techniques and expanding their applications to other areas of cancer diagnosis.

REFERENCES

[1] M. B. Schabath and M. L. Cote, "Cancer progress and priorities: Lung cancer," *Cancer Epidemiology, Biomarkers & Prevention*, vol. 28, no. 10, pp. 1563–1579, Oct. 2019.

[2] J. R. Molina, P. Yang, S. D. Cassivi, S. E. Schild, and A. A. Adjei, "Non-small cell lung cancer: Epidemiology, risk factors, treatment, and survivorship," *Mayo Clinic Proceedings*, vol. 83, no. 5, pp. 584–594, May 2008.

[3] D. Amicizia, M. F. Piazza, F. Marchini, M. Astengo, F. Grammatico, A. Battaglini, I. Schenone, C. Sticchi, R. Lavieri, B. Di Silverio, G. B. Andreoli, and F. Ansaldi, "Systematic review of lung cancer screening: Advancements and strategies for implementation," *Healthcare*, vol. 11, no. 14, p. 2085, Jul. 2023.

[4] R. Javed, T. Abbas, A. H. Khan, A. Daud, A. Bukhari, and R. Alharbey, "Deep learning for lung cancer detection: A review," *Artificial Intelligence Review*, vol. 57, no. 8, 2024.

[5] V. Kumar, C. Prabha, P. Sharma, N. Mittal, S. S. Askar, and M. Abouhawwash, "Unified deep learning models for enhanced lung cancer prediction with ResNet-50–101 and EfficientNet-B3 using DICOM images," *BMC Medical Imaging*, vol. 24, no. 1, 2024.

[6] J. Kim, S. Kim, S.-Y. Park, G. K. Lee, K. Y. Lim, J. Y. Kim, J.-A. Hwang, N. Yu, E. H. Kang, M. Hwang, B. R. Song, C. Park, and J.-Y. Han, "Molecular subtypes and tumor microenvironment characteristics of small-cell lung cancer associated with platinum-resistance," *Cancers*, vol. 15, no. 14, p. 3568, Jul. 2023.

[7] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, Mar. 2021.

[8] J. W. Yang, D. H. Song, H. J. An, and S. B. Seo, "Classification of subtypes including LCNEC in lung cancer biopsy slides using convolutional neural network from scratch," *Scientific Reports*, vol. 12, no. 1, 2022.

[9] L. Pantanowitz, N. Farahani, and A. Parwani, "Whole slide imaging in pathology: Advantages, limitations, and emerging perspectives," *Pathology and Laboratory Medicine International*, vol. 23, 2015.

[10] W. De Wever, J. Coolen, and J. A. Verschakelen, "Imaging techniques in lung cancer," *Breathe*, vol. 7, no. 4, pp. 338–346, 2011.

[11] Á. D. Tárnoki, D. L. Tárnoki, M. Dąbrowska, M. Knetki-Wróblewska, A. Frille, H. Stubbs, K. G. Blyth, and A. D. Juul, "New developments in the imaging of lung cancer," *Breathe (Sheffield, England)*, vol. 20, no. 1, p. 230176, 2024.

[12] M. Aharonu and R. L. Kumar, "Systematic review of deep learning techniques for lung cancer detection," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 3, 2023.

[13] A. E. Elnakib, M. H. Amer, and F. E. Z. Abou-Chadi, "Early lung cancer detection using deep learning optimization," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 16, no. 6, pp. 82–94, 2020.

[14] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[15] P. Schaböck, "Deep Learning in Medical Image Analysis." M.A. thesis, TU Wien, Vienna, Austria, 2015.

[16] A. Truini, P. Santos Pereira, A. Cavazza, P. Spagnolo, S. Nosseir, L. Longo, A. Jukna, F. Lococo, G. Vincenzi, G. Bogina, M. Tiseo, and G. Rossi, "Classification of different patterns of pulmonary adenocarcinomas," *Expert Review of Respiratory Medicine*, vol. 9, no. 5, pp. 571–586, 2015.

[17] E. Kuhn, P. Morbini, A. Cancellieri, S. Damiani, A. Cavazza, and C. E. Comin, "Adenocarcinoma classification: Patterns and prognosis," *Pathologica*, vol. 110, no. 1, pp. 5–11, 2018.

[18] W. D. Travis, "Pathology of lung cancer," *Clinics in Chest Medicine*, vol. 32, no. 4, pp. 669–692, 2011.

[19] X. Luo, S. Yin, L. Yang, J. Fujimoto, Y. Yang, C. Moran, N. Kalhor, A. Weissferdt, Y. Xie, A. Gazdar, J. Minna, I. I. Wistuba, Y. Mao, and G. Xiao, "Development and validation of a pathology image analysis-based predictive model for lung adenocarcinoma prognosis - a multi-cohort study," *Scientific Reports*, vol. 9, no. 1, 2019.

[20] Y. Wang, X. Pan, H. Lin, C. Han, Y. An, B. Qiu, Z. Feng, X. Huang, Z. Xu, Z. Shi, X. Chen, B. Li, L. Yan, C. Lu, Z. Li, Y. Cui, Z. Liu, and Z. Liu, "Multi-scale pathology image texture signature is a prognostic factor for resectable lung adenocarcinoma: A multi-center, retrospective study," *Journal of Translational Medicine*, vol. 20, no. 1, 2022.

[21] A. Sadhwani, H.-W. Chang, A. Behrooz, T. Brown, I. Auvigne-Flament, H. Patel, R. Findlater, V. Velez, F. Tan, K. Tekiela, E. Wulczyn, E. S. Yi, C. H. Mermel, D. Hanks, P.-H. C. Chen, K. Kulig, C. Batenchuk, D. F. Steiner, and P. Cimermancic, "Comparative analysis of machine learning approaches to classify tumor mutation burden in lung adenocarcinoma using histopathology images," *Scientific Reports*, vol. 11, no. 1, 2021.

[22] L. F. Talib, J. Amin, M. Sharif, M. Raza, "Transformer based semantic segmentation and CNN network for detection of histopathological lung cancer." *Biomedical Signal Processing and Control*, vol. 92, p. 106106, 2024.

[23] N. Coudray, P. S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyö, A. L. Moreira, N. Razavian, A. Tsirigos, "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning." *Nature Medicine*, vol. 24, no. 10, pp. 1559–1567, 2018.

[24] J. Ji, J. Li, W. Zhang, Y. Geng, Y. Dong, J. Huang, L. Hong, "Automated lung and colon cancer classification using histopathological images." *Biomedical Engineering and Computational Biology*, vol. 15, 2024.

[25] P. Afshar, A. Mohammadi, P. N. Tyrrell, P. Cheung, A. Sigiuk, K. N. Plataniotis, E. T. Nguyen, A. Oikonomou, "DRTOP: Deep learning based radiomics for the time to event outcome prediction in lung cancer." *Scientific Reports*, vol. 10, no. 1, 2020.

[26] S. Zheng, J. Guo, J. A. Langendijk, S. Both, R. N. J. Veldhuis, M. Oudkerk, P. M. A. van Ooijen, R. Wijsman, N. M. Sijtsema, "Survival prediction for stage I–IIIA non small cell lung cancer using deep learning." *Radiotherapy and Oncology*, vol. 180, p. 109483, 2023.

[27] Y. Xu, A. Hosny, R. Zeleznik, C. Parmar, T. Coroller, I. Franco, R. H. Mak, H. J. W. L. Aerts, "Deep learning predicts lung cancer treatment response from serial medical imaging." *Clinical Cancer Research*, vol. 25, no. 11, pp. 3266–3275, 2019.

[28] X. Wang, H. Duan, X. Li, X. Ye, G. Huang, S. Nie, "A prognostic analysis method for non small cell lung cancer based on the computed tomography radiomics." *Physics in Medicine & Biology*, vol. 65, no. 4, p. 045006, 2020.

[29] J. W. Yang, D. H. Song, H. J. An, S. B. Seo, "Classification of subtypes including LCNEC in lung cancer biopsy slides using convolutional neural network from scratch." *Scientific Reports*, vol. 12, no. 1, 2022.

[30] P. Aonpong, Y. Iwamoto, X.-H. Han, L. Lin, Y.-W. Chen, "Genotype guided radiomics signatures for recurrence prediction of non small cell lung cancer." *IEEE Access*, vol. 9, pp. 90244–90254, 2021.

[31] P. Huang, P. B. Illei, W. Franklin, P.-H. Wu, P. M. Forde, S. Ashrafinia, C. Hu, H. Khan, H. V. Vadvala, I.-M. Shih, R. J. Battafarano, M. A. Jacobs, X. Kong, J. Lewis, R. Yan, Y. Chen, F. Housseau, A. Rahmim, E. K. Fishman, and E. Gabrielson, "Lung cancer recurrence risk prediction through integrated deep learning evaluation." *Cancers*, vol. 14, no. 17, p. 4150, 2022.

[32] H. H. Pham, M. Futakuchi, A. Bychkov, T. Furukawa, K. Kuroda, J. Fukuoka, "Detection of lung cancer lymph node metastases from whole slide histopathologic images using a two step deep learning approach." *The American Journal of Pathology*, vol. 189, no. 12, pp. 2428–2439, 2019.

[33] R. Ding, P. Prasanna, G. Corredor, C. Barrera, P. Zens, C. Lu, P. Velu, P. Leo, N. Beig, H. Li, P. Toro, S. Berezowska, V. Baxi, D. Balli, M. Belete, D. L. Rimm, V. Velcheti, K. Schalper, A. Madabhushi, "Image analysis reveals molecularly distinct patterns of TILs in NSCLC associated with treatment outcome." *Npj Precision Oncology*, vol. 6, no. 1, 2022.

[34] L. Sha, B. L. Osinski, I. Y. Ho, T. L. Tan, C. Willis, H. Weiss, N. Beaubier, B. M. Mahon, T. J. Taxter, S. S. Yip, "Multi field of view deep learning model predicts non small cell lung cancer programmed death ligand 1 status from whole slide hematoxylin and eosin images." *Journal of Pathology Informatics*, vol. 10, no. 1, p. 24, 2019.

[35] F. Kanavati, G. Toyokawa, S. Momosaki, H. Takeoka, M. Okamoto, K. Yamazaki, S. Takeo, O. Iizuka, M. Tsuneki, "A deep learning model for the classification of indeterminate lung carcinoma in biopsy whole slide images." *Scientific Reports*, vol. 11, no. 1, 2021.

[36] C. Lu, K. Bera, X. Wang, P. Prasanna, J. Xu, A. Janowczyk, N. Beig, M. Yang, P. Fu, J. Lewis, H. Choi, R. A. Schmid, S. Berezowska, K. Schalper, D. Rimm, V. Velcheti, A. Madabhushi, "A prognostic model for overall survival of patients with early stage non small cell lung cancer: A multicentre, retrospective study." *The Lancet Digital Health*, vol. 2, no. 11, 2020.

[37] K.-H. Yu, C. Zhang, G. J. Berry, R. B. Altman, C. Ré, D. L. Rubin, M. Snyder, "Predicting non small cell lung cancer prognosis by fully automated microscopic pathology image features." *Nature Communications*, vol. 7, no. 1, 2016.

[38] Xiangxue Wang, K. Bera, C. Barrera, Y. Zhou, C. Lu, P. Vaidya, P. Fu, M. Yang, R. A. Schmid, S. Berezowska, H. Choi, V. Velcheti, A. Madabhushi, "A prognostic and predictive computational pathology image signature for added benefit of adjuvant chemotherapy in early stage non small cell lung cancer." *eBioMedicine*, vol. 69, p. 103481, 2021.

[39] S. Lin, J. P. Samsoondar, E. Bandari, S. Keow, B. Bikash, D. Tan, J. Martinez Acevedo, J. Loggie, M. Pham, N. J. Wu, T. Misra, V. H. K. Lam, I. Sansano, M. J. Cecchini, "Digital quantification of tumor cellularity as a novel prognostic feature of non–small cell lung carcinoma." *Modern Pathology*, vol. 36, no. 3, p. 100055, 2023.

[40] W. S. Shim, K. Yim, T.-J. Kim, Y. E. Sung, G. Lee, J. H. Hong, S. H. Chun, S. Kim, H. J. An, S. J. Na, J. J. Kim, M. H. Moon, S. W. Moon, S. Park, S. A. Hong, Y. H. Ko, "DeepRePath: Identifying the prognostic features of early stage lung adenocarcinoma using multi scale pathology images and deep convolutional neural networks." *Cancers*, vol. 13, no. 13, p. 3308, 2021.

[41] S. Wang, T. Wang, L. Yang, D. M. Yang, J. Fujimoto, F. Yi, X. Luo, Y. Yang, B. Yao, S. Lin, C. Moran, N. Kalhor, A. Weissferdt, J. Minna, Y. Xie, I. I. Wistuba, Y. Mao, G. Xiao, "ConvPath: A software tool for lung adenocarcinoma digital pathological image analysis aided by a convolutional neural network." *EBioMedicine*, vol. 50, pp. 103–110, 2019.

[42] Z. Wu, L. Wang, C. Li, Y. Cai, Y. Liang, X. Mo, Q. Lu, L. Dong, Y. Liu, "DeepLRHE: A deep convolutional neural network framework to evaluate the risk of lung cancer recurrence and metastasis from histopathology images." *Frontiers in Genetics*, vol. 11, 2020.

[43] A. A. Borkowski, M. M. Bui, L. B. Thomas, C. P. Wilson, L. A. DeLand, and S. M. Mastorides, "Lung and Colon Cancer Histopathological Image Dataset (LC25000)." Internet: https://arxiv.org/abs/1912.12142v1, 2019 [Accessed: May 15, 2023].

[44] F. Akmal Dzulkifli, M. Yusoff Mashor, R. A. A. Raof, H. Jaafar, "An introduction to double stain normalization technique for brain tumour histopathological images." *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 1, p. 375, 2024.

[45] Md. Z. Hoque, A. Keskinarkaus, P. Nyberg, T. Seppänen, "Stain normalization methods for histopathology image analysis: A comprehensive review and experimental comparison." *Information Fusion*, vol. 102, p. 101997, 2024.

[46] S. Anand, G. Sangeethapriya, "Visual improvement of breast histopathology images using color transfer techniques." *International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)*, vol. 7, no. 4, 2018.

[47] S. Roy, K. Bhalla, R. Patel, "Mathematical analysis of histogram equalization techniques for medical image enhancement: A tutorial from the perspective of data loss." *Multimedia Tools and Applications*, vol. 83, no. 5, pp. 14363–14392, 2023.

[48] Cancer Imaging Archive, "Eaglescope." Internet: https://pathdb.cancerimagingarchive.net/eaglescope/dist/?configurl=%2Fsystem%2Ffiles%2Fcollectionmetadata%2F202210%2Feaglescope_main_page_current.json&filterState=%5B%7B%22id%22%3A%22location%22%2C%22title%22%3A%22Cancer%2BLocation%22%2C%22field%22%3A%22location%22%2C%22operation%22%3A%22eq%22%2C%22values%22%3A%22Lung%22%7D%2C%7B%22id%22%3A%22protocol%22%2C%22title%22%3A%22Staining%2BProtocol%22%2C%22field%22%3A%22protocol%22%2C%22operation%22%3A%22eq%22%2C%22values%22%3A%22Hematoxylin%2Band%2Beosin%2Bstain%22%7D%5D, Jan. 23, 2021 [Accessed: May 1, 2023].

[49] S. S. Bagade, V. K. Shandilya, "Use of histogram equalization in image processing for image enhancement." *International Journal of Software Engineering Research and Practices*, 2011.

[50] P. Jagatheeswari, S. S. Kumar, and M. Rajaram, "Novel approach for contrast enhancement based on histogram equalization followed by median filter." *ARPN Journal of Engineering and Applied Sciences*, vol. 4, no. 7, 2009.

[51] O. Patel, P. S. Maravi, S. Sharma, "A comparative study of histogram equalization based image enhancement techniques for brightness preservation and contrast enhancement." *Signal & Image Processing: An International Journal*, vol. 4, no. 5, pp. 11–25, 2013.

[52] A. Farhadipour, "Lung and colon cancer detection with convolutional neural networks and adaptive histogram equalization." *Iran Journal of Computer Science*, vol. 7, no. 2, pp. 381–395, 2023.

[53] X. Wang, J.-K. Yan, J.-Y. Cai, J.-H. Deng, Q. Qin, Y. Cheng, "Super resolution reconstruction of single image for latent features." *Computational Visual Media*, vol. 10, no. 6, pp. 1219–1239, 2024.

[54] M. Yu, J. Shi, C. Xue, X. Hao, G. Yan, "A review of single image super resolution reconstruction based on deep learning." *Multimedia Tools and Applications*, vol. 83, no. 18, pp. 55921–55962, 2023.

[55] Xintao Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, C. C. Loy, "ESRGAN: Enhanced super resolution generative adversarial networks." *Lecture Notes in Computer Science*, pp. 63–79, 2019.

[56] S. Strban, "Super Resolution." Unpublished manuscript, Ljubljana University, Ljubljana, Slovenia, [n.d.]. Internet: https://www.sling.si/wp-content/uploads/2024/09/Sebastien-Strban-Super-resolution-ENG_small.pdf [Accessed: May 1, 2023].

[57] F. Jia, Z. Chen, Z. Song, L. Liu, and C. Jia, "CWT Net: Super resolution of histopathology images using a cross scale wavelet based transformer." Internet: https://arxiv.org/html/2409.07092v1, 2024 [Accessed: May 15, 2023].

[58] P. Ramamoorthy, B. R. Ramakantha, S. S. Askar, M. Abouhawwash, "Histopathology based breast cancer prediction using deep learning methods for healthcare applications." *Frontiers in Oncology*, vol. 14, 2024.

[59] A. K. Abdulrahman, S. Ozturk, "A novel hybrid DCT and DWT based robust watermarking algorithm for color images." *Multimedia Tools and Applications*, vol. 78, no. 12, pp. 17027–17049, 2019.

[60] K. J. Giri, S. M. Quadri, R. Bashir, J. I. Bhat, "DWT based color image watermarking: A review." *Multimedia Tools and Applications*, vol. 79, nos. 43–44, pp. 32881–32895, 2020.

[61] Y. Qi, X. Liu, H. Li, M. Liu, and W. Li, "Discrete Wavelet Transform-Based CNN for Breast Cancer Classification from Histopathology Images," *in Machine Learning for Cyber Security. ML4CS 2022*, eds. Y. Xu, H. Yan, H. Teng, J. Cai, and J. Li, *Lecture Notes in Computer Science*, vol. 13655, Cham: Springer, 2023.

[62] A. O. Mohammed, H. I. Hussein, R. J. Mstafa, A. M. Abdulazeez, "A blind and robust color image watermarking scheme based on DCT and DWT domains." *Multimedia Tools and Applications*, vol. 82, no. 21, pp. 32855–32881, 2023.

[63] H. Djouima, A. Zitouni, A. C. Megherbi, S. Sbaa, "Data augmentation by wavelet transform for breast cancer based on deep learning." *Revue d'Intelligence Artificielle*, vol. 38, no. 4, pp. 1097–1107, 2024.

[64] J. Wang and Y. Chen, *Introduction to transfer learning: Algorithms and practice*. Singapore: Springer Nature Singapore, 2023.

[65] J. Wagh, "Geographical area classification on satellite images using CNN architecture." *International Journal of Scientific Research in Engineering and Management*, vol. 8, no. 5, pp. 1–5, 2024.

[66] M. Li, Y. Zhu, R. Du, C. Jia, "Secure verification encrypted image retrieval scheme with addition homomorphic bitmap index." *Proceedings of the 2024 International Conference on Multimedia Retrieval*, pp. 193–201, 2024.

[67] Z. Benamara, S. Zehani, A. Zitouni, "The effect of fully connected layers in different pre trained CNN architectures on the enhancement of lung cancer classification." *In 2024 8th International Conference on Image and Signal Processing and Their Applications (ISPA)*, pp. 1–6, 2024.

[68] X. Fei, S. Wu, J. Miao, G. Wang, L. Sun, "Lightweight VGG: A fast deep learning architecture based on dimensionality reduction and nonlinear enhancement for hyperspectral image classification." *Remote Sensing*, vol. 16, no. 2, p. 259, 2024.

[69] J. Brownlee, (2018), *Statistical Methods for Machine Learning: Discover how to Transform Data into Knowledge with Python*, [n.p.]: [n.p.], [Online]. Available: Google Play ebook Accessed: May 15, 2023.

[70] J. Quiñonero-Candela, M. Sugiyama, A. Schwaighofer, and N. Lawrence, *Dataset shift in machine learning*. Cambridge, MA: The MIT Press, 2022.

[71] M. Çalışan, V. Gündüzalp, N. Olgun, "Evaluation of U net and ResNet architectures for biomedical image segmentation." *International Journal of 3D Printing Technologies and Digital Industry*, vol. 7, no. 3, pp. 561–570, 2023.

[72] R. He, C. Li, X. Yang, J. Yang, T. Jiang, M. Grzegorzek, H. Sun, "Application of graph curvature features in computer aided diagnosis for histopathological image identification of gastric cancer." *Intelligent Medicine*, vol. 4, no. 3, pp. 141–152, 2024.

[73] A. Laurent Bellue, A. Sadraoui, L. Claude, J. Calderaro, K. Posseme, E. Vibert, D. Cherqui, O. Rosmorduc, M. Lewin, J.-C. Pesquet, C. Guettier, "Deep learning classification and quantification of pejorative and nonpejorative architectures in resected hepatocellular carcinoma from digital histopathologic images." *The American Journal of Pathology*, vol. 194, no. 9, pp. 1684–1700, 2024.

[74] C. C. Reyes Aldasoro, A. Janowczyk, M. Veta, P. Bankhead, K. Sirinukunwattana, "Digital pathology: 15th European Congress, ECDP 2019, Warwick, UK, April 10–13, 2019, Proceedings." *Springer International Publishing*, 2019.

[75] A. A. Rizvanov, B. K. Singh, P. Ganasala, "Advances in biomedical engineering and technology: Select proceedings of ICBEST 2018." *Springer Singapore*, 2021.

[76] Y. Xue, C. Chen, C. Chen, L. Zuo, Y. Liu, "Data augmentation, labelling, and imperfections: Third MICCAI workshop, Dali 2023, held in conjunction with MICCAI 2023, Vancouver, BC, Canada, October 12, 2023, Proceedings." *Springer*, 2024.

[77] M. R. Faisal, A. M. Shifa, M. H. Rahman, M. A. Uddin, R. M. Rahman, "Bengali & Banglish: A monolingual dataset for emotion detection in linguistically diverse contexts." *Data in Brief*, vol. 55, p. 110760, 2024.

[78] R. Zhang, X. Zheng, L. Zhang, Y. Xu, X. Lin, X. Wang, C. Wu, F. Jiang, J. Wang, "Lanmao sleep recorder versus polysomnography in neonatal EEG recording and sleep analysis." *Journal of Neuroscience Methods*, vol. 410, p. 110222, 2024.

[79] H. DALIANIS, *Clinical text mining: Secondary use of electronic patient records*. New York: Springer, 2019.

[80] A. Vahadane, T. Peng, A. Sethi, S. Albarqouni, L. Wang, M. Baust, K. Steiger, A. M. Schlitter, I. Esposito, N. Navab, "Structure preserving color normalization and sparse stain separation for histological images." *IEEE Transactions on Medical Imaging*, vol. 35, no. 8, pp. 1962–1971, 2016.

[81] P. K. Sethy, A. Geetha Devi, B. Padhan, S. K. Behera, S. Sreedhar, K. Das, "Lung cancer histopathological image classification using wavelets and AlexNet." *Journal of X Ray Science and Technology*, vol. 31, no. 1, pp. 211–221, 2023.

[82] M. Pradhan, R. K. Sahu, "Automatic detection of lung cancer using the potential of artificial intelligence (AI)." *Advances in Medical Technologies and Clinical Practice*, pp. 106–123, 2022.

[83] A. Halder, D. Dey, "MorphAttnNet: An attention based morphology framework for lung cancer subtype classification." *Biomedical Signal Processing and Control*, vol. 86, p. 105149, 2023.

[84] B. K. Hatuwal, H. C. Thapa, "Lung cancer detection using convolutional neural network on histopathological images." *International Journal of Computer Trends & Technology*, vol. 68, no. 10, pp. 21–24, 2020.

[85] N. Baranwal, P. Doravari, R. Kachhoria, "Classification of histopathology images of lung cancer using convolutional neural network (CNN)." *Disruptive Developments in Biomedical Applications*, pp. 75–89, 2022.

[86] J. Civit Masot, A. Bañuls Beaterio, M. Domínguez Morales, M. Rivas Pérez, L. Muñoz Saavedra, J. M. Rodríguez Corral, "Non small cell lung cancer diagnosis aid with histopathological images using explainable deep learning techniques." *Computer Methods and Programs in Biomedicine*, vol. 226, p. 107108, 2022.

[87] N. F. Noaman, B. M. Kanber, A. A. Smadi, L. Jiao, M. K. Alsmadi, "Advancing oncology diagnostics: AI enabled early detection of lung cancer through hybrid histological image analysis." *IEEE Access*, vol. 12, pp. 64396–64415, 2024.

[88] L. Tian, J. Wu, W. Song, Q. Hong, D. Liu, F. Ye, F. Gao, Y. Hu, M. Wu, Y. Lan, L. Chen, "Precise and automated lung cancer cell classification using deep neural network with multiscale features and model distillation." *Scientific Reports*, vol. 14, no. 1, 2024.

[89] F. Mercaldo, M. G. Tibaldi, L. Lombardi, L. Brunese, A. Santone, M. Cesarelli, "An explainable method for lung cancer detection and localisation from tissue

images through convolutional neural networks." *Electronics*, vol. 13, no. 7, p. 1393, 2024.

[90] E. A.-R. Hamed, M. A.-M. Salem, N. L. Badr, M. F. Tolba, "An efficient combination of convolutional neural network and LIGHTGBM algorithm for lung cancer histopathology classification." *Diagnostics*, vol. 13, no. 15, p. 2469, 2023.

[91] S. Wadekar, D. K. Singh, "A modified convolutional neural network framework for categorizing lung cell histopathological image based on residual network." *Healthcare Analytics*, vol. 4, p. 100224, 2023.

[92] S. Hadiyoso, S. Aulia, I. D. Irawati, "Diagnosis of lung and colon cancer based on clinical pathology images using convolutional neural network and CLAHE framework." *International Journal of Applied Science and Engineering*, vol. 20, no. 1, pp. 1–7, 2023.

[93] N. Yahia Ibrahim, A. S. Talaat, "An enhancement technique to diagnose colon and lung cancer by using double CLAHE and deep learning." *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 8, 2022.

[94] I. Chhillar, A. Singh, "A feature engineering based machine learning technique to detect and classify lung and colon cancer from histopathological images." *Medical & Biological Engineering & Computing*, vol. 62, no. 3, pp. 913–924, 2023.

[95] A. Farhadipour, "Lung and colon cancer detection with convolutional neural networks and adaptive histogram equalization." *Iran Journal of Computer Science*, vol. 7, no. 2, pp. 381–395, 2023.

APPENDIX A

TRAINING AND VALIDATION METRICS

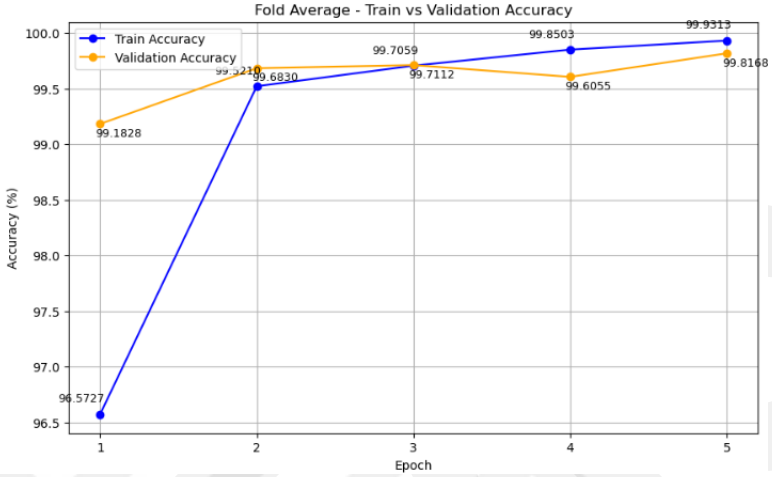


Figure A.1 Overall Fold Average of Train vs. Validation Accuracy of Convit_Small with Reinhard Color Transformation Images

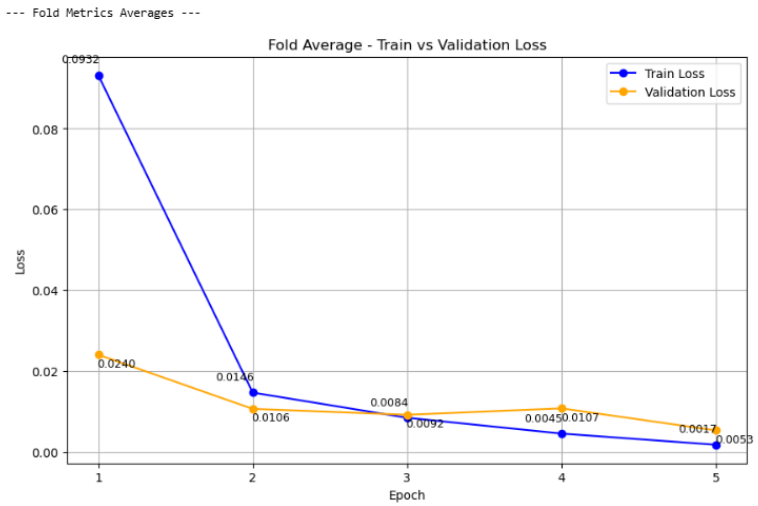


Figure A.2 Overall Fold Average of Train vs. Validation Loss of Convit_Small with Reinhard Color Transformation Images

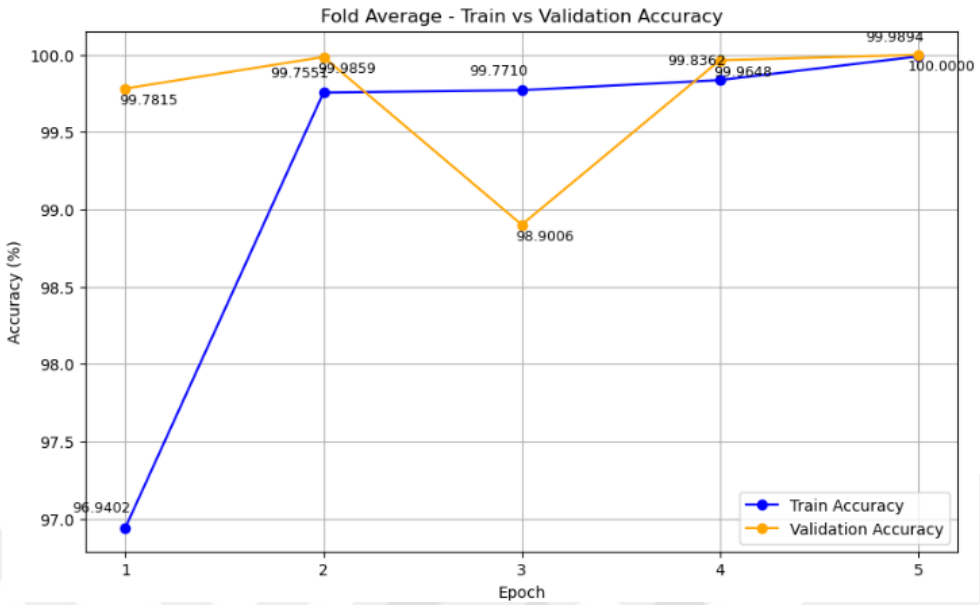


Figure A.3 Overall Fold Average of Train vs. Validation Accuracy of Convit_Small with DCT-DWT Applied Vahanade Color Transformation Images

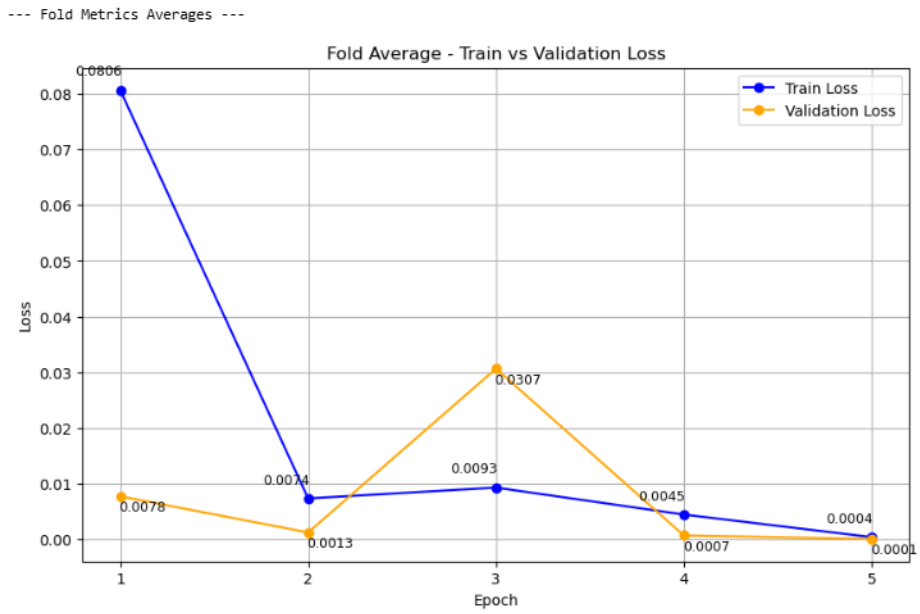


Figure A.4 Overall Fold Average of Train vs. Validation Loss of Convit_Small with DCT-DWT Applied Vahanade Color Transformation Images

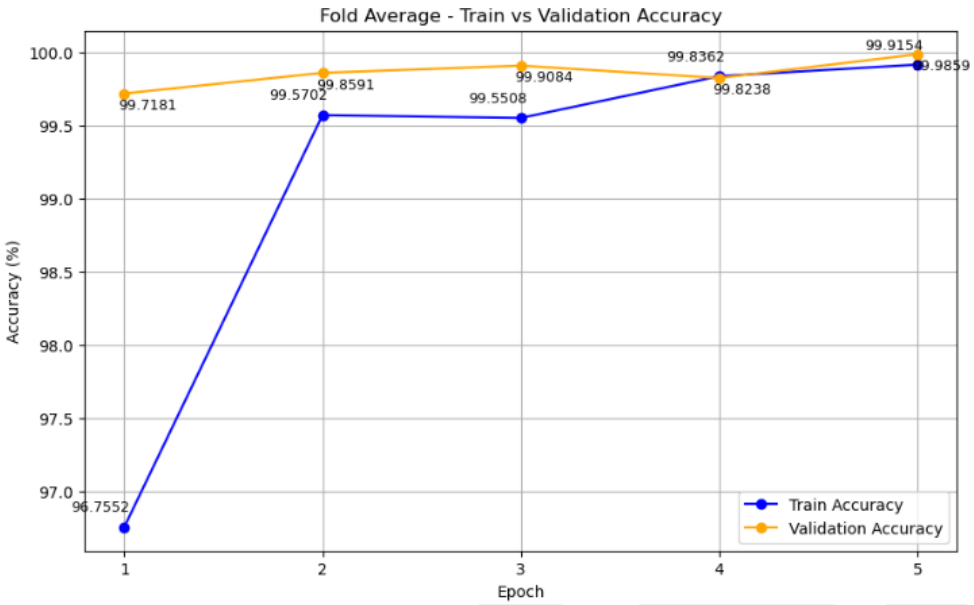


Figure A.5 Overall Fold Average of Train vs. Validation Accuracy of Convit_Small with CLAHE and Sharpening Applied Vahanade Color Transformation Images

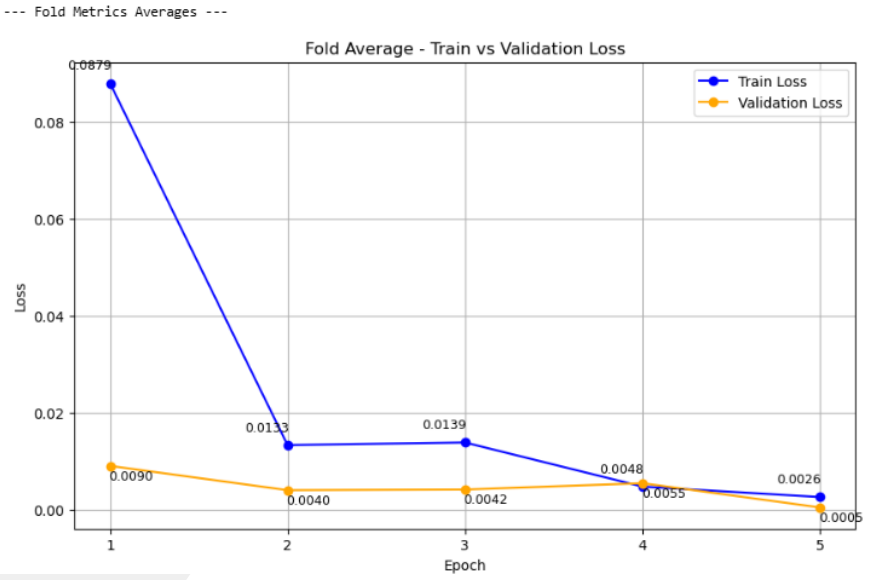


Figure A.6 Overall Fold Average of Train vs. Validation Loss of Convit_Small with CLAHE and Sharpening Applied Vahanade Color Transformation Images

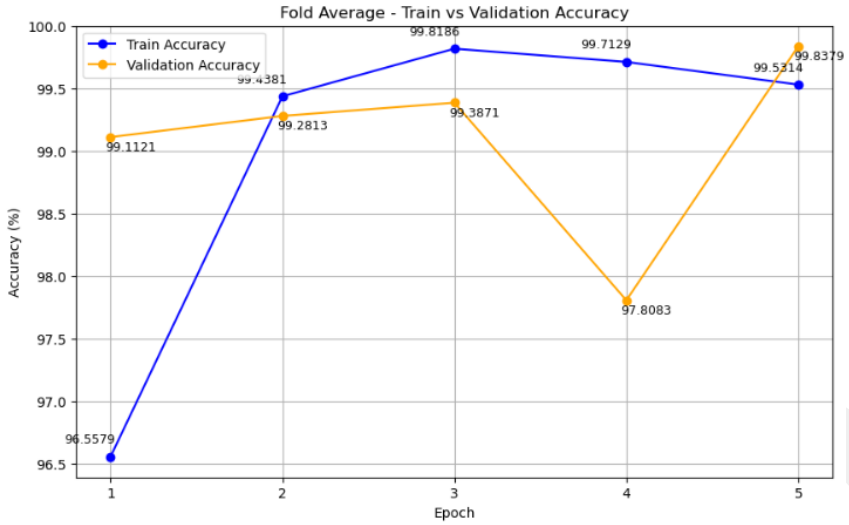


Figure A.7 Overall Fold Average of Train vs. Validation Accuracy of Convit_Small with GHE Applied Vahanade Color Transformation Images

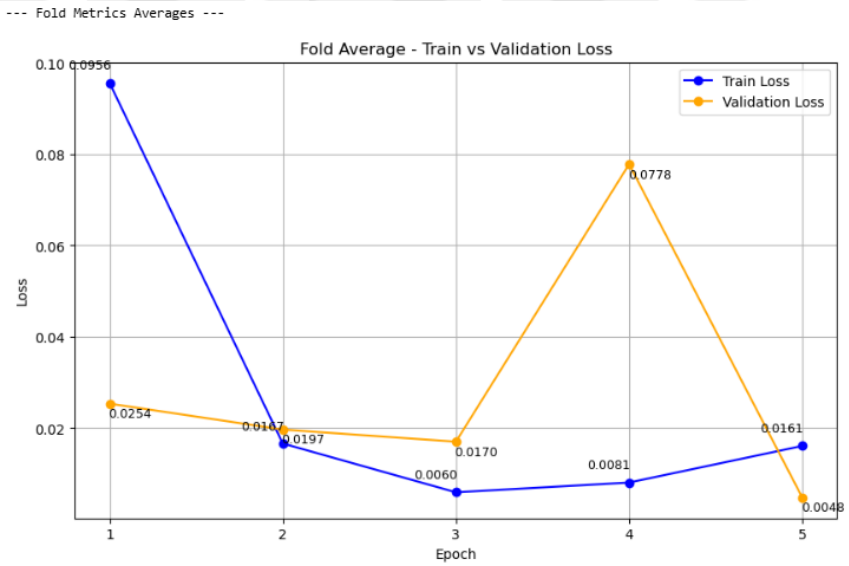


Figure A.8 Overall Fold Average of Train vs. Validation Loss of Convit_Small with GHE Applied Vahanade Color Transformation Images

APPENDIX B

CONFUSION MATRICES

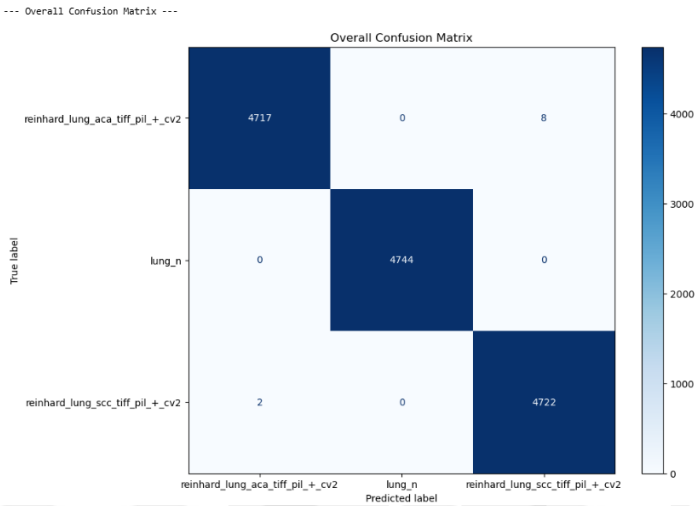


Figure B.1 Overall Confusion Matrix of Convit_ Small with Reinhard Color Transformation Images

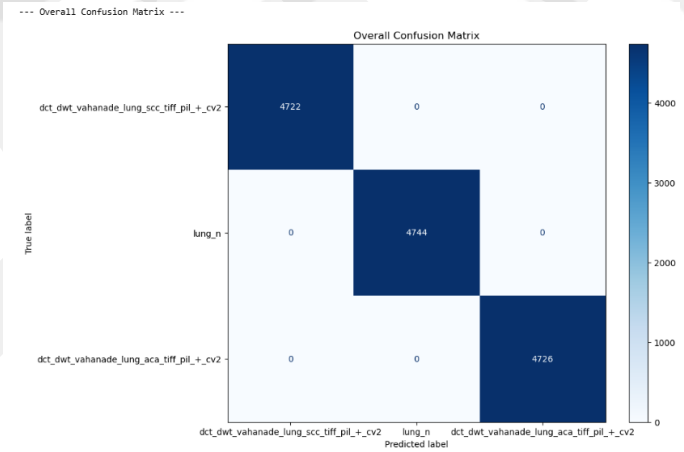


Figure B.2 Overall Confusion Matrix of Convit_ Small with DCT-DWT Applied Vahanade Color Transformation Images

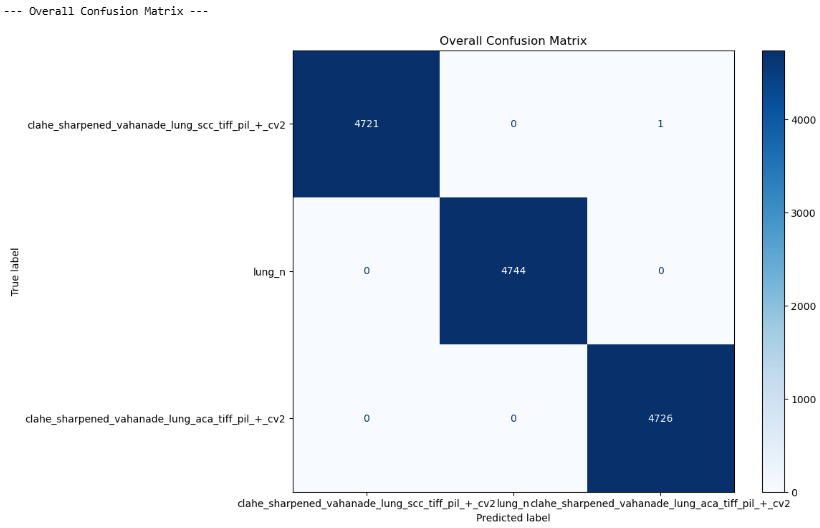


Figure B.3 Overall Confusion Matrix of Convit_ Small with CLAHE and Sharpening Applied Vahanade Color Transformation Images

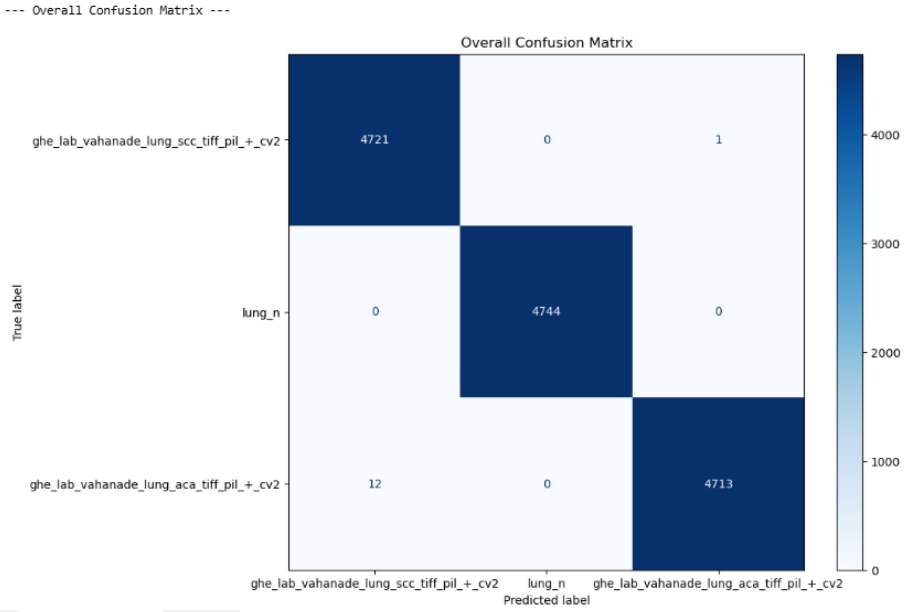


Figure B.4 Overall Confusion Matrix of Convit_ Small with GHE Applied Vahanade Color Transformation Images

APPENDIX C

CLASSIFICATION REPORTS

--- Overall Classification Report ---

	precision	recall	f1-score	support
reinhard_lung_aca_tiff_pil_+_cv2	1.00	1.00	1.00	4725
lung_n	1.00	1.00	1.00	4744
reinhard_lung_scc_tiff_pil_+_cv2	1.00	1.00	1.00	4724
accuracy			1.00	14193
macro avg	1.00	1.00	1.00	14193
weighted avg	1.00	1.00	1.00	14193

Figure C.1 Results of the Overall Classification Report for Convit_Small with Reinhard Color Transformation Images

--- Overall Classification Report ---

	precision	recall	f1-score	support
dct_dwt_vahanade_lung_scc_tiff_pil_+_cv2	1.00	1.00	1.00	4722
lung_n	1.00	1.00	1.00	4744
dct_dwt_vahanade_lung_aca_tiff_pil_+_cv2	1.00	1.00	1.00	4726
accuracy			1.00	14192
macro avg	1.00	1.00	1.00	14192
weighted avg	1.00	1.00	1.00	14192

Figure C.2 Results of the Overall Classification Report for Convit_Small with DCT-DWT Applied Vahanade Color Transformation Images

--- Overall Classification Report ---

	precision	recall	f1-score	support
clahe_sharpened_vahanade_lung_scc_tiff_pil_+_cv2	1.00	1.00	1.00	4722
lung_n	1.00	1.00	1.00	4744
clahe_sharpened_vahanade_lung_aca_tiff_pil_+_cv2	1.00	1.00	1.00	4726
accuracy			1.00	14192
macro avg	1.00	1.00	1.00	14192
weighted avg	1.00	1.00	1.00	14192

Figure C.3 Results of the Overall Classification Report for Convit_Small with CLAHE and Sharpening Applied Vahanade Color Transformation Images

--- Overall Classification Report ---

	precision	recall	f1-score	support
ghe_lab_vahanade_lung_scc_tiff_pil_+_cv2	1.00	1.00	1.00	4722
lung_n	1.00	1.00	1.00	4744
ghe_lab_vahanade_lung_aca_tiff_pil_+_cv2	1.00	1.00	1.00	4725
accuracy			1.00	14191
macro avg	1.00	1.00	1.00	14191
weighted avg	1.00	1.00	1.00	14191

Figure C.4 Results of the Overall Classification Report for Convit_Small with GHE Applied Vahanade Color Transformation Images

APPENDIX D

ROC AUC AND OTHER SCORES

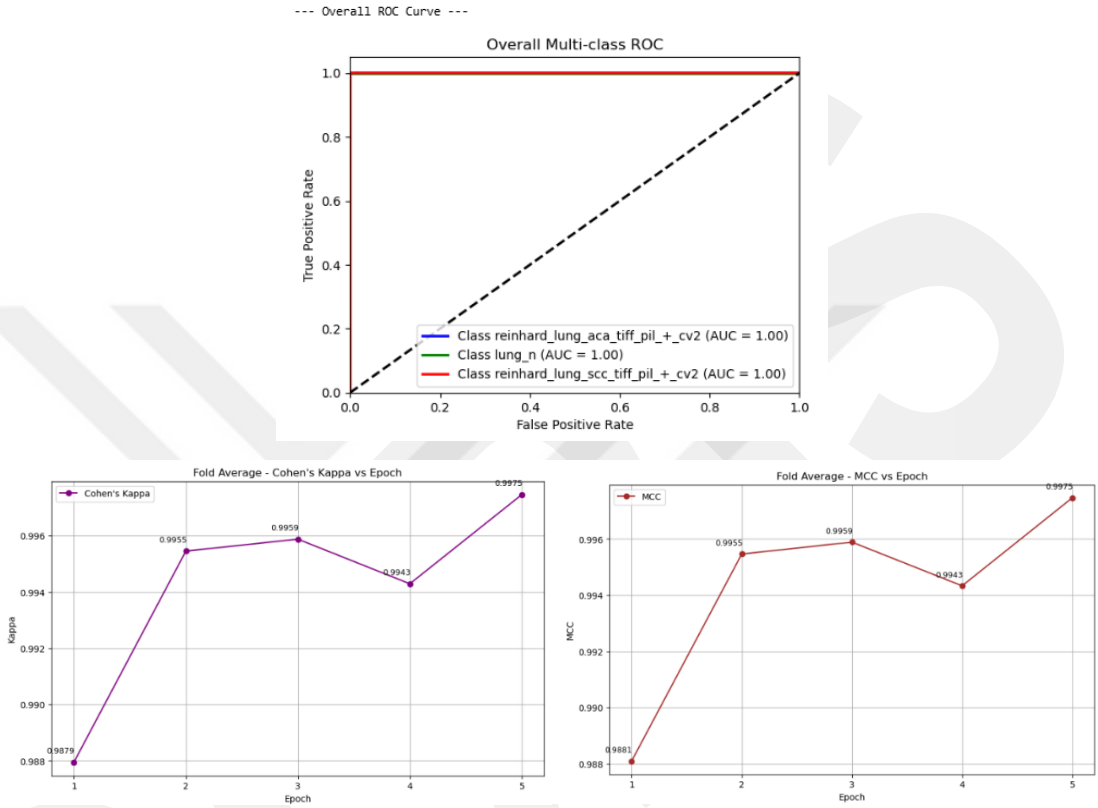


Figure D.1 Overall Fold Average of the Overall Multi-Class ROC AUC Graph, Cohen's Kappa and Matthew Correlation Coefficient(MCC) Scores of Convit_Small with Reinhard Color Transformation Images

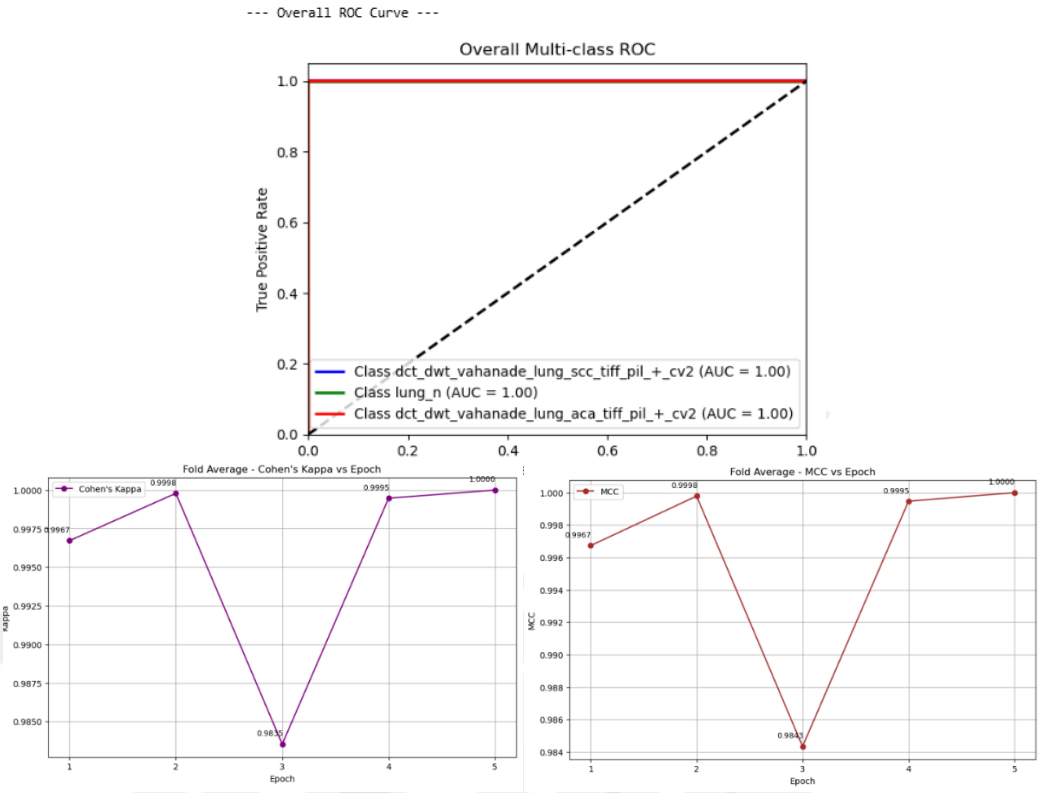


Figure D.2 Overall Fold Average of the Overall Multi-Class ROC AUC Graph, Cohen's Kappa and Matthew Correlation Coefficient(MCC) Scores of Convit_Small with DCT-DWT Applied Vahanade Color Transformation Images

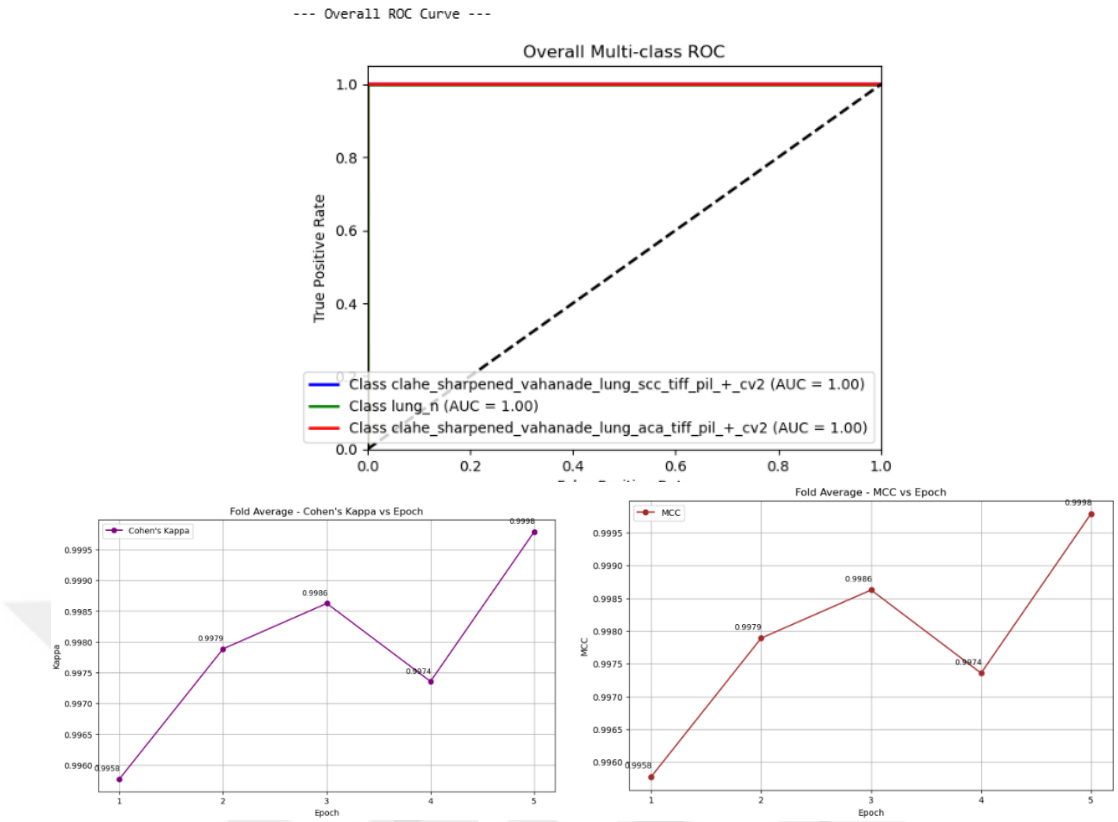


Figure D.3 Overall Fold Average of the Overall Multi-Class ROC AUC Graph, Cohen's Kappa and Matthew Correlation Coefficient(MCC) Scores of Convit_Small with CLAHE and Sharpening Applied Vahanade Color Transformation Images

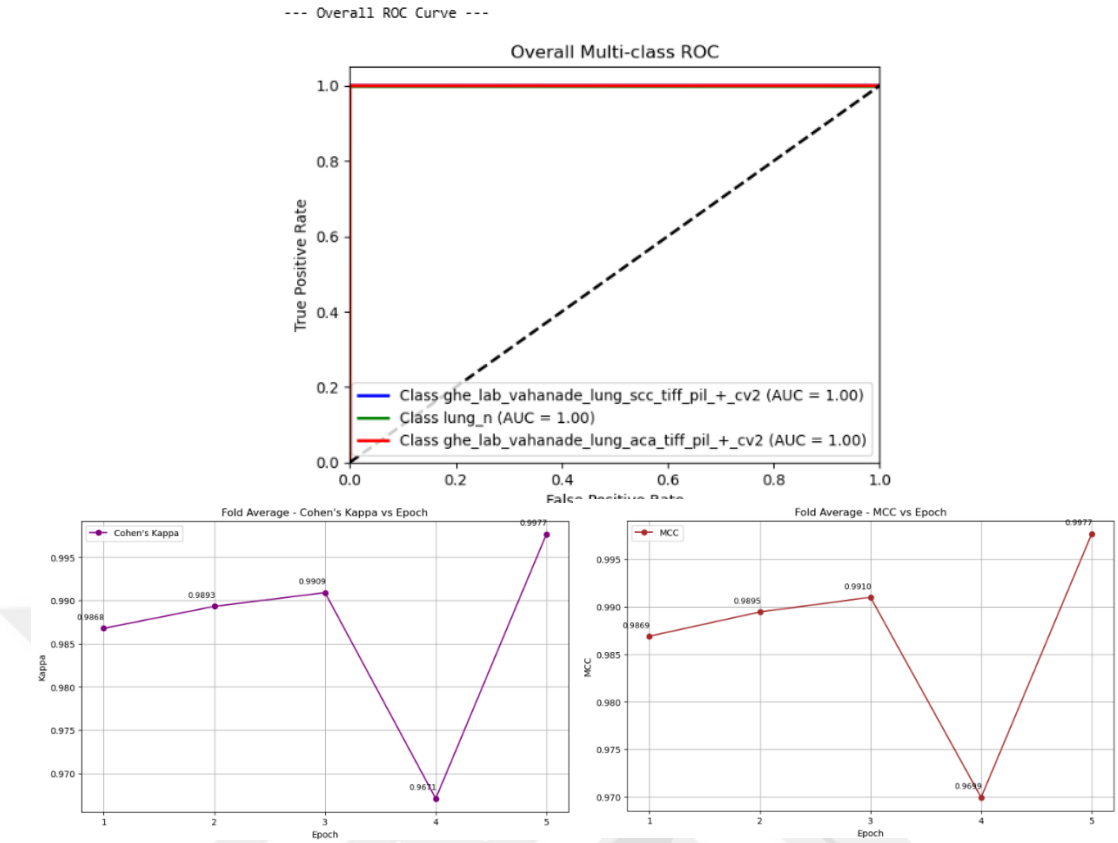


Figure D.4 Overall Fold Average of the Overall Multi-Class ROC AUC Graph, Cohen's Kappa and Matthew Correlation Coefficient(MCC) Scores of Convit_Small with GHE Applied Vahanade Color Transformation Images

APPENDIX E

RESULTS OF ALL FOLDS

Fold 1 Metrics	inceptionresnetv2 with original images	convit small with original images	only reinhard applied	only vahanade applied	vahanade+dct-dwt applied	dct-dwt+vahanade applied	vahanade+ghe applied	vahanade+clahe with a. sharpening applied	dct-dwt+clahe with a. sharpening+vahanade applied
Train Loss (Epoch 1)	0,1059	0,0938	0,0953	0,0884	0,1015	0,0829	0,0906	0,0950	0,0899
Train Loss (Epoch 2)	0,0197	0,0166	0,0148	0,0018	0,0149	0,0058	0,0146	0,0073	0,0101
Train Loss (Epoch 3)	0,0098	0,0111	0,0067	0,0002	0,0003	0,0115	0,0085	0,0063	0,0039
Train Loss (Epoch 4)	0,0090	0,0046	0,0082	0,0001	0,0002	0,0001	0,0237	0,0068	0,0077
Train Loss (Epoch 5)	0,0095	0,0055	0,0018	0,0001	0,0001	0,0001	0,0006	0,0054	0,0003
Validation Loss (Epoch 1)	0,0226	0,0225	0,0161	0,0042	0,0196	0,0070	0,0159	0,0101	0,0394
Validation Loss (Epoch 2)	0,0089	0,0144	0,0114	0,0003	0,0004	0,0004	0,0448	0,0011	0,0251
Validation Loss (Epoch 3)	0,0031	0,0030	0,0095	0,0002	0,0002	0,0002	0,0166	0,0013	0,0010
Validation Loss (Epoch 4)	0,0027	0,0299	0,0333	0,0001	0,0001	0,0001	0,0034	0,0018	0,0007
Validation Loss (Epoch 5)	0,0509	0,0023	0,0047	0,0001	0,0001	0,0001	0,0014	0,0005	0,0002
Train Accuracy (Epoch 1)	96,5921	96,6097	96,5569	96,7148	96,2040	96,8819	96,8026	96,4943	96,5469
Train Accuracy (Epoch 2)	99,3660	99,5157	99,4805	99,9560	99,8327	99,8150	99,5420	99,7886	99,6300
Train Accuracy (Epoch 3)	99,6654	99,6302	99,7887	99,9912	99,9736	99,6301	99,7534	99,7710	99,8679
Train Accuracy (Epoch 4)	99,7358	99,8415	99,7182	99,9912	99,9736	100,0000	99,1544	99,7181	99,7269
Train Accuracy (Epoch 5)	99,7182	99,8151	99,9207	99,9912	99,9736	100,0000	99,9912	99,8238	99,9912
Validation Accuracy (Epoch 1)	99,2603	99,2603	99,6125	99,9296	99,2603	99,8239	99,5069	99,7534	98,4848
Validation Accuracy (Epoch 2)	99,7887	99,4716	99,6830	100,0000	100,0000	100,0000	98,1684	100,0000	99,2248
Validation Accuracy (Epoch 3)	99,9296	99,8591	99,7534	100,0000	100,0000	100,0000	99,1899	100,0000	99,9295
Validation Accuracy (Epoch 4)	99,8943	98,9081	98,8376	100,0000	100,0000	100,0000	99,8591	99,9296	99,9648
Validation Accuracy (Epoch 5)	99,2603	99,9648	99,8591	100,0000	100,0000	100,0000	99,9296	100,0000	99,9648
Cohen's Kappa Score (Epoch 1)	0,9889	0,9889	0,9942	0,9989	0,9889	0,9974	0,9926	0,9963	0,9778
Cohen's Kappa Score (Epoch 2)	0,9968	0,9921	0,9952	1,0000	1,0000	1,0000	0,9730	1,0000	0,9889
Cohen's Kappa Score (Epoch 3)	0,9989	0,9979	0,9963	1,0000	1,0000	1,0000	0,9879	1,0000	0,9995
Cohen's Kappa Score (Epoch 4)	0,9984	0,9936	0,9826	1,0000	1,0000	1,0000	0,9979	0,9989	1,0000
Cohen's Kappa Score (Epoch 5)	0,9889	0,9995	0,9979	1,0000	1,0000	1,0000	0,9989	1,0000	1,0000
MCC Score (Epoch 1)	0,9889	0,9889	0,9942	0,9989	0,9889	0,9974	0,9926	0,9963	0,9781
MCC Score (Epoch 2)	0,9968	0,9921	0,9953	1,0000	1,0000	1,0000	0,9730	1,0000	0,9890
MCC Score (Epoch 3)	0,9989	0,9979	0,9963	1,0000	1,0000	1,0000	0,9879	1,0000	0,9995
MCC Score (Epoch 4)	0,9984	0,9938	0,9828	1,0000	1,0000	1,0000	0,9979	0,9989	1,0000
MCC Score (Epoch 5)	0,9889	0,9995	0,9979	1,0000	1,0000	1,0000	0,9989	1,0000	1,0000
ROC AUC Score (Adenocarcinoma)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
ROC AUC Score (Normal Cell)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
ROC AUC Score (Squamous Cell Carcinoma)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Precision (Adenocarcinoma)	0,9979	1,0000	0,9979	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Precision (Normal Cell)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Precision (Squamous Cell Carcinoma)	1,0000	0,9989	0,9979	1,0000	1,0000	1,0000	0,9979	1,0000	1,0000
Recall (Adenocarcinoma)	1,0000	0,9989	0,9979	1,0000	1,0000	1,0000	0,9980	1,0000	1,0000
Recall (Normal Cell)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Recall (Squamous Cell Carcinoma)	0,9979	1,0000	0,9979	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
F1 Score (Adenocarcinoma)	0,9989	0,9995	0,9979	1,0000	1,0000	1,0000	0,9990	1,0000	1,0000
F1 Score (Normal Cell)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
F1 Score (Squamous Cell Carcinoma)	0,9989	0,9995	0,9979	1,0000	1,0000	1,0000	0,9989	1,0000	1,0000
Best Validation Accuracy	0,9993	0,9996	0,9986	1,0000	1,0000	1,0000	0,9993	1,0000	0,9996
Number of misclassified images (CM)	2	1	4	0	0	0	2	0	0

Figure E.1 Fold 1 Results: Epoch-by-Epoch Evaluation of Performance Metrics Across Different Preprocessing Methods

Fold 2 Metrics	inceptionresnetv2 with original images	convit small with original images	only reinhard applied	only vahanade applied	vahanade+dct-dwt applied	dct-dwt+vahanade applied	vahanade+ghe applied	vahanade+clahe with a. sharpening applied	dct-dwt+clahe with a. sharpening+vahanade applied
Train Loss (Epoch 1)	0,1013	0,0932	0,1084	0,0807	0,0859	0,0759	0,1051	0,0857	0,9977
Train Loss (Epoch 2)	0,0172	0,0080	0,0117	0,0044	0,0040	0,0036	0,0172	0,0040	0,0051
Train Loss (Epoch 3)	0,0117	0,0058	0,0114	0,0186	0,0151	0,0002	0,0036	0,0125	0,0355
Train Loss (Epoch 4)	0,0098	0,0098	0,0024	0,0003	0,0004	0,0001	0,0047	0,0003	0,0006
Train Loss (Epoch 5)	0,0099	0,0062	0,0004	0,0001	0,0004	0,0001	0,0047	0,0001	0,0002
Validation Loss (Epoch 1)	0,0090	0,0290	0,0191	0,0070	0,0051	0,0023	0,0191	0,0081	0,0093
Validation Loss (Epoch 2)	0,0047	0,0078	0,0130	0,0004	0,0004	0,0003	0,0080	0,0099	0,0013
Validation Loss (Epoch 3)	0,0050	0,0085	0,0172	0,0005	0,0021	0,0001	0,0468	0,0016	0,0023
Validation Loss (Epoch 4)	0,0048	0,0292	0,0059	0,0002	0,0005	0,0001	0,0042	0,0002	0,0004
Validation Loss (Epoch 5)	0,0020	0,0007	0,0027	0,0001	0,0001	0,0001	0,0004	0,0001	0,0009
Train Accuracy (Epoch 1)	96,2487	96,6890	95,9493	97,0759	97,0847	97,2342	95,9394	96,9612	96,3178
Train Accuracy (Epoch 2)	99,3836	99,7887	99,6125	99,8767	99,8679	99,9119	99,4187	99,8855	99,8591
Train Accuracy (Epoch 3)	99,6125	99,8327	99,5773	99,3835	99,4716	100,0000	99,9119	99,6124	98,7403
Train Accuracy (Epoch 4)	99,7270	99,6478	99,9296	99,9912	99,9912	100,0000	99,8591	100,0000	99,9912
Train Accuracy (Epoch 5)	99,6742	99,8239	99,9824	99,9912	99,9824	100,0000	99,8150	100,0000	99,9912
Validation Accuracy (Epoch 1)	99,8591	98,8376	99,4716	99,7534	99,8239	100,0000	99,4364	99,7534	99,6829
Validation Accuracy (Epoch 2)	99,8591	99,7887	99,5773	100,0000	99,9296	100,0000	99,8239	99,6125	99,9295
Validation Accuracy (Epoch 3)	99,8943	99,7182	99,4716	100,0000	99,8591	100,0000	98,5558	99,9648	99,9648
Validation Accuracy (Epoch 4)	99,8591	98,9433	99,7534	100,0000	99,8943	100,0000	99,8239	100,0000	99,9648
Validation Accuracy (Epoch 5)	99,9296	100,0000	99,9648	100,0000	99,9296	100,0000	100,0000	100,0000	99,9295
Cohen's Kappa Score (Epoch 1)	0,9979	0,9826	0,9921	0,9963	0,9984	1,0000	0,9915	0,9963	0,9958
Cohen's Kappa Score (Epoch 2)	0,9979	0,9968	0,9937	1,0000	1,0000	1,0000	0,9974	0,9942	0,9995
Cohen's Kappa Score (Epoch 3)	0,9984	0,9958	0,9921	1,0000	0,9989	1,0000	0,9783	0,9995	1,0000
Cohen's Kappa Score (Epoch 4)	0,9979	0,9841	0,9963	1,0000	0,9995	1,0000	0,9974	1,0000	1,0000
Cohen's Kappa Score (Epoch 5)	0,9989	1,0000	0,9995	1,0000	1,0000	1,0000	1,0000	1,0000	0,9995
MCC Score (Epoch 1)	0,9979	0,9826	0,9921	0,9963	0,9984	1,0000	0,9916	0,9963	0,9958
MCC Score (Epoch 2)	0,9979	0,9968	0,9937	1,0000	1,0000	1,0000	0,9974	0,9942	0,9995
MCC Score (Epoch 3)	0,9984	0,9958	0,9921	1,0000	0,9989	1,0000	0,9786	0,9995	1,0000
MCC Score (Epoch 4)	0,9979	0,9843	0,9963	1,0000	0,9995	1,0000	0,9974	1,0000	1,0000
MCC Score (Epoch 5)	0,9989	1,0000	0,9995	1,0000	1,0000	1,0000	1,0000	1,0000	0,9995
ROC AUC Score (Adenocarcinoma)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
ROC AUC Score (Normal Cell)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
ROC AUC Score (Squamous Cell Carcinoma)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Precision (Adenocarcinoma)	0,9989	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Precision (Normal Cell)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Precision (Squamous Cell Carcinoma)	0,9989	1,0000	0,9989	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Recall (Adenocarcinoma)	0,9989	1,0000	0,9989	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Recall (Normal Cell)	0,9990	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Recall (Squamous Cell Carcinoma)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
F1 Score (Adenocarcinoma)	0,9989	1,0000	0,9995	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
F1 Score (Normal Cell)	0,9995	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
F1 Score (Squamous Cell Carcinoma)	0,9995	1,0000	0,9995	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Best Validation Accuracy	0,9993	1,0000	0,9996	1,0000	0,9993	1,0000	1,0000	1,0000	0,9996
Number of misclassified images (CM)	2	0	1	0	0	0	0	0	0

Figure E.2 Fold 2 Results: Epoch-by-Epoch Evaluation of Performance Metrics Across Different Preprocessing Methods

Fold 3 Metrics	inceptionresnetv2 with original images	convit small with original images	only reinhard applied	only vahanade applied	vahanade+dct-dwt applied	dct-dwt+vahanade applied	vahanade+ghe applied	vahanade+clahe with a. sharpening applied	dct-dwt+clahe with a. sharpening+vahanade applied
Train Loss (Epoch 1)	0,0114	0,0764	0,0841	0,0751	0,0850	0,0654	0,0901	0,0787	0,0842
Train Loss (Epoch 2)	0,0151	0,0103	0,0158	0,0051	0,0062	0,0202	0,0219	0,0080	0,0152
Train Loss (Epoch 3)	0,0129	0,0079	0,0075	0,0005	0,0022	0,0041	0,0017	0,0273	0,0004
Train Loss (Epoch 4)	0,0125	0,0095	0,0037	0,0170	0,0036	0,0008	0,0002	0,0046	0,0002
Train Loss (Epoch 5)	0,0058	0,0010	0,0030	0,0022	0,0041	0,0001	0,0506	0,0004	0,0001
Validation Loss (Epoch 1)	0,0096	0,0214	0,0546	0,0030	0,0051	0,0156	0,0164	0,0103	0,0062
Validation Loss (Epoch 2)	0,0100	0,0040	0,0179	0,0023	0,0003	0,0033	0,0034	0,0014	0,0058
Validation Loss (Epoch 3)	0,0048	0,0065	0,0053	0,0002	0,0017	0,0005	0,0005	0,0120	0,0009
Validation Loss (Epoch 4)	0,0053	0,0041	0,0059	0,0005	0,0002	0,0002	0,0002	0,0011	0,0014
Validation Loss (Epoch 5)	0,0046	0,0004	0,0031	0,0001	0,0001	0,0001	0,0066	0,0002	0,0010
Train Accuracy (Epoch 1)	96,0549	97,2349	96,9091	97,3049	96,6708	97,6748	96,7236	97,2961	97,1459
Train Accuracy (Epoch 2)	99,5509	99,7094	99,4452	99,8327	99,7446	99,2954	99,1721	99,7622	99,4626
Train Accuracy (Epoch 3)	99,6125	99,7446	99,7446	99,9824	99,9031	99,8943	99,9560	99,0928	99,9824
Train Accuracy (Epoch 4)	99,6037	99,7006	99,9119	99,3923	99,8679	100,0000	99,9912	99,8679	99,9824
Train Accuracy (Epoch 5)	99,8855	99,9648	99,8855	99,9119	99,8591	100,0000	98,6965	99,9912	99,9824
Validation Accuracy (Epoch 1)	99,7534	99,5421	97,8161	99,9296	99,7887	99,4362	99,4715	99,7181	99,8591
Validation Accuracy (Epoch 2)	99,6478	99,9296	99,4364	99,9296	100,0000	99,9295	99,9295	99,9648	99,8591
Validation Accuracy (Epoch 3)	99,8239	99,7534	99,7887	100,0000	99,9296	100,0000	100,0000	99,7181	99,9295
Validation Accuracy (Epoch 4)	99,9648	99,8239	99,7534	100,0000	100,0000	100,0000	100,0000	100,0000	99,9648
Validation Accuracy (Epoch 5)	99,8239	100,0000	99,8239	100,0000	100,0000	100,0000	99,8591	100,0000	99,9648
Cohen's Kappa Score (Epoch 1)	0,9963	0,9931	0,9678	0,9989	0,9968	0,9915	0,9921	0,9958	0,9979
Cohen's Kappa Score (Epoch 2)	0,9947	0,9989	0,9921	0,9989	1,0000	0,9989	0,9989	0,9995	0,9979
Cohen's Kappa Score (Epoch 3)	0,9974	0,9963	0,9974	1,0000	0,9989	1,0000	1,0000	0,9958	0,9989
Cohen's Kappa Score (Epoch 4)	0,9995	0,9974	0,9968	1,0000	1,0000	1,0000	1,0000	1,0000	0,9995
Cohen's Kappa Score (Epoch 5)	0,9974	1,0000	0,9979	1,0000	1,0000	1,0000	0,9979	1,0000	0,9995
MCC Score (Epoch 1)	0,9963	0,9931	0,9684	0,9989	0,9968	0,9916	0,9921	0,9958	0,9979
MCC Score (Epoch 2)	0,9947	0,9989	0,9921	0,9989	1,0000	0,9989	0,9989	0,9995	0,9979
MCC Score (Epoch 3)	0,9974	0,9963	0,9974	1,0000	0,9989	1,0000	1,0000	0,9958	0,9989
MCC Score (Epoch 4)	0,9995	0,9974	0,9968	1,0000	1,0000	1,0000	1,0000	1,0000	0,9995
MCC Score (Epoch 5)	0,9974	1,0000	0,9979	1,0000	1,0000	1,0000	0,9979	1,0000	0,9995
ROC AUC Score (Adenocarcinoma)	0,9999	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
ROC AUC Score (Normal Cell)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
ROC AUC Score (Squamous Cell Carcinoma)	0,9999	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Precision (Adenocarcinoma)	0,9990	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Precision (Normal Cell)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Precision (Squamous Cell Carcinoma)	1,0000	1,0000	0,9957	1,0000	1,0000	1,0000	1,0000	1,0000	0,9989
Recall (Adenocarcinoma)	1,0000	1,0000	0,9960	1,0000	1,0000	1,0000	1,0000	1,0000	0,9990
Recall (Normal Cell)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Recall (Squamous Cell Carcinoma)	0,9989	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
F1 Score (Adenocarcinoma)	0,9995	1,0000	0,9980	1,0000	1,0000	1,0000	1,0000	1,0000	0,9995
F1 Score (Normal Cell)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
F1 Score (Squamous Cell Carcinoma)	0,9995	1,0000	0,9979	1,0000	1,0000	1,0000	1,0000	1,0000	0,9995
Best Validation Accuracy	0,9996	1,0000	0,9982	1,0000	1,0000	1,0000	1,0000	1,0000	0,9996
Number of misclassified images (CM)	1	0	4	0	0	0	0	0	1

Figure E.3 Fold 3 Results: Epoch-by-Epoch Evaluation of Performance Metrics Across Different Preprocessing Methods

Fold 4 Metrics	inceptionresnetv2 with original images	convit small with original images	only reinhard applied	only vahanade applied	vahanade+dct-dwt applied	dct-dwt+vahanade applied	vahanade+ghe applied	vahanade+clahe with a. sharpening applied	dct-dwt+clahe with a. sharpening+vahanade applied
Train Loss (Epoch 1)	0,1033	0,0761	0,0962	0,0670	0,0850	0,0828	0,0951	0,0920	0,0917
Train Loss (Epoch 2)	0,0236	0,0134	0,0188	0,0083	0,0084	0,0056	0,0165	0,0227	0,0190
Train Loss (Epoch 3)	0,0081	0,0070	0,0066	0,0016	0,0004	0,0025	0,0047	0,0023	0,0048
Train Loss (Epoch 4)	0,0076	0,0046	0,0062	0,0001	0,0001	0,0002	0,0069	0,0097	0,0007
Train Loss (Epoch 5)	0,0104	0,0051	0,0005	0,0001	0,0001	0,0001	0,0085	0,0059	0,0001
Validation Loss (Epoch 1)	0,0215	0,0212	0,0223	0,0155	0,0082	0,0086	0,0176	0,0087	0,0192
Validation Loss (Epoch 2)	0,0110	0,0129	0,0073	0,0139	0,0007	0,0016	0,0307	0,0069	0,0021
Validation Loss (Epoch 3)	0,0042	0,0014	0,0033	0,0002	0,0002	0,0002	0,0164	0,0013	0,0046
Validation Loss (Epoch 4)	0,0173	0,0159	0,0062	0,0001	0,0001	0,0002	0,0059	0,0231	0,0002
Validation Loss (Epoch 5)	0,0014	0,0110	0,0153	0,0001	0,0001	0,0001	0,0073	0,0010	0,0001
Train Accuracy (Epoch 1)	96,5041	97,5784	96,3808	97,7015	97,0674	96,6091	96,5299	96,3449	96,6173
Train Accuracy (Epoch 2)	99,1018	99,6213	99,4276	99,7358	99,7006	99,7886	99,4716	99,2073	99,3041
Train Accuracy (Epoch 3)	99,7887	99,7622	99,7622	99,9648	99,9736	99,9295	99,8327	99,9560	99,7798
Train Accuracy (Epoch 4)	99,7975	99,8503	99,9384	99,9912	99,9736	100,0000	99,7358	99,6653	99,9736
Train Accuracy (Epoch 5)	99,6125	99,8327	99,9736	99,9912	99,9736	100,0000	99,6917	99,7886	99,9824
Validation Accuracy (Epoch 1)	99,4012	99,5421	99,2603	99,4362	99,7533	99,7886	99,4362	99,6829	99,3305
Validation Accuracy (Epoch 2)	99,6125	99,4716	99,7534	99,5772	100,0000	100,0000	98,8372	99,7181	99,9648
Validation Accuracy (Epoch 3)	99,9296	100,0000	99,9296	100,0000	100,0000	100,0000	99,4362	99,9648	99,7886
Validation Accuracy (Epoch 4)	99,2955	99,5069	99,7534	100,0000	100,0000	100,0000	99,7886	99,2248	100,0000
Validation Accuracy (Epoch 5)	100,0000	99,5773	99,4364	100,0000	100,0000	100,0000	99,6124	99,9648	100,0000
Cohen's Kappa Score (Epoch 1)	0,9910	0,9931	0,9894	0,9915	0,9963	0,9968	0,9921	0,9952	0,9900
Cohen's Kappa Score (Epoch 2)	0,9942	0,9921	0,9968	0,9937	1,0000	1,0000	0,9831	0,9958	0,9995
Cohen's Kappa Score (Epoch 3)	0,9989	1,0000	0,9995	1,0000	1,0000	1,0000	0,9921	0,9995	0,9968
Cohen's Kappa Score (Epoch 4)	0,9894	0,9926	0,9968	1,0000	1,0000	1,0000	0,9974	0,9884	1,0000
Cohen's Kappa Score (Epoch 5)	1,0000	0,9937	0,9921	1,0000	1,0000	1,0000	0,9947	0,9995	1,0000
MCC Score (Epoch 1)	0,9910	0,9931	0,9894	0,9916	0,9963	0,9968	0,9921	0,9952	0,9900
MCC Score (Epoch 2)	0,9942	0,9921	0,9968	0,9937	1,0000	1,0000	0,9833	0,9958	0,9995
MCC Score (Epoch 3)	0,9989	1,0000	0,9995	1,0000	1,0000	1,0000	0,9921	0,9995	0,9968
MCC Score (Epoch 4)	0,9895	0,9926	0,9968	1,0000	1,0000	1,0000	0,9974	0,9884	1,0000
MCC Score (Epoch 5)	1,0000	0,9937	0,9921	1,0000	1,0000	1,0000	0,9947	0,9995	1,0000
ROC AUC Score (Adenocarcinoma)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
ROC AUC Score (Normal Cell)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
ROC AUC Score (Squamous Cell Carcinoma)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Precision (Adenocarcinoma)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9947	0,9989	1,0000
Precision (Normal Cell)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Precision (Squamous Cell Carcinoma)	1,0000	1,0000	0,9990	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Recall (Adenocarcinoma)	1,0000	1,0000	0,9989	1,0000	1,0000	1,0000	0,9946	1,0000	1,0000
Recall (Normal Cell)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Recall (Squamous Cell Carcinoma)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9989	1,0000
F1 Score (Adenocarcinoma)	1,0000	1,0000	0,9994	1,0000	1,0000	1,0000	0,9973	0,9995	1,0000
F1 Score (Normal Cell)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
F1 Score (Squamous Cell Carcinoma)	1,0000	1,0000	0,9995	1,0000	1,0000	1,0000	0,9974	0,9995	1,0000
Best Validation Accuracy	1,0000	1,0000	0,9993	1,0000	1,0000	1,0000	0,9979	0,9996	1,0000
Number of misclassified images (CM)	0	0	1	0	0	0	5	1	0

Figure E.4 Fold 4 Results: Epoch-by-Epoch Evaluation of Performance Metrics Across Different Preprocessing Methods

Fold 5 Metrics	inceptionresnetv2 with original images	convit small with original images	only reinhard applied	only vahanade applied	vahanade+dct-dwt applied	dct-dwt+vahanade applied	vahanade+ghe applied	vahanade+clahe with a. sharpening applied	dct-dwt+clahe with a. sharpening+vahanade applied
Train Loss (Epoch 1)	0,1064	0,0996	0,0822	0,0896	0,0834	0,0961	0,0971	0,0882	0,0767
Train Loss (Epoch 2)	0,0198	0,0176	0,0121	0,0043	0,0031	0,0018	0,0130	0,0246	0,0124
Train Loss (Epoch 3)	0,0157	0,0042	0,0100	0,0003	0,0295	0,0284	0,0113	0,0210	0,0160
Train Loss (Epoch 4)	0,0082	0,0086	0,0067	0,0001	0,0006	0,0213	0,0049	0,0024	0,0064
Train Loss (Epoch 5)	0,0052	0,0107	0,0029	0,0001	0,0003	0,0019	0,0162	0,0012	0,0008
Validation Loss (Epoch 1)	0,0239	0,0082	0,0082	0,0414	0,0052	0,0053	0,0583	0,0080	0,0186
Validation Loss (Epoch 2)	0,0173	0,0078	0,0035	0,0003	0,0005	0,0008	0,0117	0,0009	0,0017
Validation Loss (Epoch 3)	0,0020	0,0066	0,0104	0,0002	0,0008	0,1524	0,0050	0,0046	0,0069
Validation Loss (Epoch 4)	0,0035	0,0128	0,0024	0,0001	0,0003	0,0032	0,3754	0,0012	0,0010
Validation Loss (Epoch 5)	0,0025	0,0081	0,0006	0,0001	0,0002	0,0003	0,0084	0,0006	0,0034
Train Accuracy (Epoch 1)	96,3015	96,0549	97,0676	96,5214	97,1416	96,3009	96,7941	96,6796	97,2340
Train Accuracy (Epoch 2)	99,3924	99,3660	99,6390	99,8767	99,9295	99,9648	99,5860	99,2073	99,5860
Train Accuracy (Epoch 3)	99,4628	99,8855	99,6566	100,0000	98,9696	99,4011	99,6389	99,3218	99,4362
Train Accuracy (Epoch 4)	99,7270	99,7534	99,7534	100,0000	99,9736	99,2073	99,8239	99,9295	99,7974
Train Accuracy (Epoch 5)	99,8327	99,6213	99,8943	100,0000	99,9824	99,9472	99,4627	99,9736	99,9560
Validation Accuracy (Epoch 1)	99,0842	99,7182	99,7534	98,4144	99,8943	99,8591	97,7097	99,6829	99,4010
Validation Accuracy (Epoch 2)	99,5421	99,6830	99,9648	99,9648	99,9648	100,0000	99,6476	100,0000	99,9295
Validation Accuracy (Epoch 3)	99,9648	99,8591	99,6125	99,9648	99,9648	94,5032	99,7533	99,8943	99,7886
Validation Accuracy (Epoch 4)	99,8239	99,7182	99,9296	99,9648	99,9648	99,8238	89,5701	99,9648	99,9648
Validation Accuracy (Epoch 5)	99,8943	99,8239	100,0000	99,9648	99,9648	100,0000	99,7886	99,9648	99,8238
Cohen's Kappa Score (Epoch 1)	0,9863	0,9958	0,9963	0,9767	0,9989	0,9979	0,9656	0,9952	0,9910
Cohen's Kappa Score (Epoch 2)	0,9931	0,9952	0,9995	1,0000	1,0000	1,0000	0,9947	1,0000	0,9989
Cohen's Kappa Score (Epoch 3)	0,9995	0,9979	0,9942	1,0000	1,0000	0,9177	0,9963	0,9984	0,9968
Cohen's Kappa Score (Epoch 4)	0,9974	0,9958	0,9989	1,0000	1,0000	0,9974	0,8429	0,9995	0,9995
Cohen's Kappa Score (Epoch 5)	0,9984	0,9974	1,0000	1,0000	1,0000	1,0000	0,9968	0,9995	0,9974
MCC Score (Epoch 1)	0,9864	0,9958	0,9963	0,9771	0,9989	0,9979	0,9661	0,9953	0,9910
MCC Score (Epoch 2)	0,9931	0,9953	0,9995	1,0000	1,0000	1,0000	0,9947	1,0000	0,9989
MCC Score (Epoch 3)	0,9995	0,9979	0,9942	1,0000	1,0000	0,9217	0,9963	0,9984	0,9968
MCC Score (Epoch 4)	0,9974	0,9958	0,9989	1,0000	1,0000	0,9974	0,8571	0,9995	0,9995
MCC Score (Epoch 5)	0,9984	0,9974	1,0000	1,0000	1,0000	1,0000	0,9968	0,9995	0,9974
ROC AUC Score (Adenocarcinoma)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
ROC AUC Score (Normal Cell)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
ROC AUC Score (Squamous Cell Carcinoma)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	1,0000	1,0000
Precision (Adenocarcinoma)	1,0000	0,9958	1,0000	1,0000	1,0000	1,0000	0,9990	1,0000	1,0000
Precision (Normal Cell)	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Precision (Squamous Cell Carcinoma)	0,9989	1,0000	1,0000	1,0000	1,0000	1,0000	0,9945	1,0000	0,9989
Recall (Adenocarcinoma)	0,9990	1,0000	1,0000	1,0000	1,0000	1,0000	0,9949	1,0000	0,9989
Recall (Normal Cell)	1,0000	0,9989	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Recall (Squamous Cell Carcinoma)	1,0000	0,9968	1,0000	1,0000	1,0000	1,0000	0,9989	1,0000	1,0000
F1 Score (Adenocarcinoma)	0,9995	0,9979	1,0000	1,0000	1,0000	1,0000	0,9969	1,0000	0,9995
F1 Score (Normal Cell)	1,0000	0,9995	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
F1 Score (Squamous Cell Carcinoma)	0,9995	0,9984	1,0000	1,0000	1,0000	1,0000	0,9967	1,0000	0,9995
Best Validation Accuracy	0,9996	0,9986	1,0000	0,9996	0,9996	1,0000	0,9979	1,0000	0,9996
Number of misclassified images (CM)	1	4	0	0	0	0	6	0	1

Figure E.5 Fold 5 Results: Epoch-by-Epoch Evaluation of Performance Metrics Across Different Preprocessing Methods