”Weakly Supervised and Deep Learning Methods for Histopathological Image Classification in Neurological and Renal Disorders”

A thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science
in
Computer Science and Engineering by Research

by

Anirudh Reddy
20161196
anirudhreddy.r@research.iiit.ac.in

Advisor: C. V. Jawahar

International Institute of Information Technology Hyderabad
500 032, India

Jun 2024

International Institute of Information Technology Hyderabad
Hyderabad, India

CERTIFICATE

This is to certify that work presented in this thesis proposal titled ***Weakly Supervised and Deep Learning Methods for Histopathological Image Classification in Neurological and Renal Disorders*** by *Anirudh Reddy* has been carried out under my supervision and is not submitted elsewhere for a degree.

_____ _____

Date Advisor: C. V. Jawahar

To my Mother and Father.

For everything.

Acknowledgments

Abstract

The analysis of digital histopathology slides or Whole Slide Images (WSIs) is critical for several diagnoses. Recent advancements in computational techniques, particularly in the field of digital pathology, have shown promise in automating the classification process. Whole Slide Imaging (WSI), combined with deep learning and modern computer vision techniques, has emerged as a powerful tool in this domain. This thesis addresses two major medical challenges using deep learning and computer vision techniques: the classification of Lupus Nephritis (LN) and low-grade gliomas into their respective subtypes.

Systemic lupus erythematosus (SLE) is an autoimmune disease wherein the patient's immune system attacks healthy tissues, leading to Lupus Nephritis (LN), a severe condition causing renal failure. Traditional methods for diagnosing LN require meticulous pathological assessment of renal biopsies, which is time-consuming. **In the first architecture**(chapter 3), We propose a novel pipeline that automates this process by: 1) detecting various glomerular patterns in WSIs using Periodic Acid-Schiff (PAS) stained images, and 2) classifying each image based on these extracted glomerular features. This approach leverages deep learning to improve the accuracy and efficiency of LN classification.

Low-grade glioma, a type of brain tumor originating from glial cells, also presents significant diagnostic challenges due to the large size and complexity of WSIs. **In the second architecture**(chapter 4), our work involves the classification of low-grade gliomas into Astrocytoma and Oligodendroglioma. Given the computational infeasibility of training deep learning models on gigapixel images, we adopt a weakly supervised method to extract discriminative patches from WSIs, which represent the tumor regions. A Convolutional Neural Network (CNN) is then trained on these discriminative patches, and the results are aggregated to determine the WSI label. Evaluated on a dataset of 581,616 patches from

286 WSIs obtained from The Cancer Genome Atlas (TCGA) portal, our method achieved a slide-wise accuracy of 79.31%, which increased to 89.65% when trained only on discriminative patches.

The methodologies presented in this thesis not only demonstrate significant improvements in classification accuracy but also offer scalable and efficient solutions for enhancing the diagnostic processes in pathology, ultimately contributing to better patient outcomes and more efficient healthcare delivery.

Contents

List of Figures

List of Tables

Symbols

α	path loss exponent
B	available bandwidth
β_{AF}	AF gain factor

List of Related Publications

[P1] Akash Gupta†*, **Anirudh Reddy***, CV Jawahar and PK Vinod, **"Classification of histopathology images using ConvNets to detect Lupus Nephritis"**, in proceedings of *Medical Imaging meets NeurIPS*, 2021.

* : denotes equal contribution

†: Work done partially during an internship at IIIT Hyderabad

Chapter 1

Introduction

Histopathology, the microscopic examination of biopsied tissue, plays a crucial role in diagnosing a wide range of diseases, from cancers to autoimmune disorders. With advancements in imaging technology, whole slide images (WSIs) have become a standard in digitizing tissue samples, providing pathologists with detailed views of cellular structures. However, the sheer size and complexity of WSIs pose significant challenges for manual analysis, necessitating the development of automated methods to assist in the diagnostic process. In recent years, machine learning, particularly deep learning, has emerged as a powerful tool for analyzing medical images. This thesis focuses on applying advanced machine learning techniques to two distinct but related problems in histopathology: the classification of Lupus Nephritis (LN) and the differentiation of glioma subtypes—Astrocytoma and Oligodendroglioma.

1.1 Problem Addressed

The accurate classification of histopathology images is critical for effective diagnosis and treatment planning. This thesis addresses two major problems within this domain, each presenting unique challenges and requiring specialized approaches.

Lupus Nephritis Classification: Systemic Lupus Erythematosus (SLE) is a chronic autoimmune disease that can affect multiple organs, including the kidneys, leading to a condition known as Lupus Nephritis (LN). LN can cause significant kidney damage and, if untreated, can progress to renal failure. Accurate classification of LN is essential for guiding patient treatment and improving outcomes.

The International Society of Nephrology/Renal Pathology Society (ISN/RPS) has classified LN into six classes based on various glomerular features observed in histopathology images. However, existing classification methods face several challenges:

- Label Noise: Traditional approaches often use bulk labeling schemes where the same label is applied to all glomeruli within a kidney, irrespective of individual phenotypical differences. This introduces significant label noise, reducing classification accuracy.

- Limited Class Consideration: Many methods only consider a limited number of glomerular classes, which is insufficient for comprehensive LN classification.

- Scalability: High-resolution WSIs require robust and scalable computational methods to process and analyze the vast amount of data effectively.

Glioma Subtype Classification: Gliomas, which originate from glial cells in the brain, are a diverse group of tumors categorized into High-Grade Glioma (HGG) and Low-Grade Glioma (LGG). Among LGGs, differentiating between Astrocytoma and Oligodendroglioma is particularly challenging due to the heterogeneous nature of these tumors. The classification of glioma subtypes using histopathology images faces several obstacles:

- Gigapixel Image Analysis: WSIs are extremely large, with resolutions often reaching $100,000 \times 100,000$ pixels, making it computationally infeasible to process the entire image directly.

- Weakly Supervised Data: Most available datasets are coarsely annotated, providing slide-level labels without detailed pixel-level annotations. This weakly supervised scenario complicates the training of deep learning models.

- Patch Selection: Tumor regions within WSIs are interspersed with normal tissue, necessitating the selection of discriminative patches that accurately represent the tumor features to improve classification performance.

1.2 Contributions

This thesis makes significant contributions to the field of histopathology image analysis by developing and applying innovative deep-learning techniques tailored to the specific challenges of LN and glioma subtype classification.

Lupus Nephritis Classification:

- Automated Pipeline Development: We propose an end-to-end automated pipeline that begins with the extraction of glomeruli from PAS-stained histopathology images and proceeds to classify these glomeruli into the six LN stages as defined by the ISN/RPS.

- Addressing Label Noise: Our method mitigates the issue of label noise by incorporating phenotypical features of glomeruli, ensuring that individual glomeruli are correctly classified based on their specific characteristics rather than relying on bulk labels.

- Improved Classification Accuracy: Through rigorous experimentation, we demonstrate that our approach significantly enhances classification accuracy compared to existing methods, providing a more reliable tool for LN diagnosis.

Glioma Subtype Classification:

- Weakly Supervised Learning: We introduce a weakly supervised deep learning method that selects discriminative patches from WSIs to differentiate between Astrocytoma and Oligodendroglioma. This approach leverages slide-level labels to train the model while focusing on the most relevant tumor regions.

- Patch-Based Analysis: By analyzing high-resolution patches extracted from WSIs, we overcome the computational challenges posed by gigapixel images. This method preserves the fine-grained details necessary for accurate classification.

- Enhanced Model Performance: Our experiments show a substantial increase in classification accuracy when using discriminative patches, validated by pathologists, indicating the method's potential for broader application in histopathology.

1.3 Thesis Organisation

The thesis is structured to provide a comprehensive and detailed exploration of the methodologies and findings related to the classification of histopathology images in the context of LN and glioma subtypes. Each chapter is designed to build on the previous ones, offering a coherent and logical progression from problem definition to solution implementation and evaluation.

1. In chapter 1, **Introduction**: This chapter introduces the overarching goals of the thesis, outlines the specific problems addressed, highlights the contributions made, and provides an overview of the thesis structure.

2. In chapter 2, **Background**: This chapter gives a background on Histopathology Imaging and the related work in the field of digital histopathology.

3. In chapter 3, **Classification of Histopathology Images Using ConvNets to Detect Lupus Nephritis**: This chapter delves into the methodologies employed for LN classification. It includes a detailed description of the dataset, the preprocessing steps, the convolutional neural network architecture used, and the experimental results obtained. The chapter also discusses the challenges encountered and the solutions implemented to address them.

4. In chapter 4, **Weakly Supervised Method for the Classification of Astrocytoma and Oligodendroglioma Using Histopathology Images**: Focusing on glioma subtype classification, this chapter presents the weakly supervised learning approach developed. It covers the process of patch extraction, the selection of discriminative patches, the training and validation of the deep learning model, and the evaluation of its performance. The chapter also explores potential extensions of the method to other cancer types and multiclass classification tasks.

5. In chapter 5, **Conclusion**: The final chapter summarizes the key findings of the research, discusses the implications of the results for the field of histopathology, and suggests directions for future work. It reflects on the limitations of the current approaches and proposes areas for further investigation and improvement.

By addressing the specific challenges associated with the classification of LN and glioma subtypes, this thesis contributes to the advancement of automated histopathology image analysis, providing valuable tools for improving diagnostic accuracy and patient care.

Chapter 2

Background

2.1 Introduction to Histopathology Imaging

Histopathology imaging is a pivotal technique in medical diagnostics that involves the microscopic examination of tissue samples to study the manifestations of diseases at the cellular level. This process typically starts with the collection of a biopsy or surgical specimen, which is then processed and fixed onto glass slides. These slides are stained with specific dyes to highlight various cellular components and structures, enabling pathologists to observe and analyze tissue morphology in detail.

Staining Techniques: The most common staining technique used in histopathology is Hematoxylin and Eosin (H&E) staining. Hematoxylin stains the cell nuclei blue, while Eosin stains the cytoplasm and extracellular matrix pink. This contrast allows for a clear distinction between different cellular components, aiding in the identification of pathological changes. Other staining methods, such as Periodic Acid-Schiff (PAS) staining, are used for specific applications. PAS staining, for instance, highlights polysaccharides like glycogen and mucosubstances, which are important in diagnosing kidney diseases like Lupus Nephritis (LN).

Digital Histopathology and Whole Slide Imaging Advancements in digital imaging technologies have revolutionized histopathology. Whole Slide Imaging (WSI) involves scanning entire glass slides to create high-resolution digital images. These digital slides can then be viewed, analyzed, and shared electronically, providing numerous advantages over traditional microscopy. WSIs are typically gigapixel

images with resolutions around $100,000 \times 100,000$ pixels, allowing for the detailed visualization of large tissue areas.

However, the immense size and high resolution of WSIs present computational challenges, particularly when applying machine learning algorithms. Processing such large images in their entirety is computationally infeasible, necessitating techniques that can handle this scale effectively.

Applications of Deep Learning in Histopathology Recent developments in deep learning have shown great promise in automating various tasks in histopathology imaging, thereby assisting pathologists in diagnostic processes. Convolutional Neural Networks (CNNs), a class of deep learning algorithms, are particularly well-suited for image analysis tasks due to their ability to learn hierarchical feature representations from raw pixel data.

In the context of Lupus Nephritis, deep learning models can be trained to detect and classify glomerular patterns in WSIs. Traditional methods for LN diagnosis are labor-intensive and time-consuming, requiring expert pathologists to manually assess renal biopsies. By contrast, deep learning approaches can automate this process, providing faster and potentially more accurate diagnoses. Our proposed pipeline for LN classification involves detecting glomeruli in PAS-stained images and then classifying these images based on the glomerular features. This method leverages the power of CNNs to implicitly learn relevant features, enhancing diagnostic accuracy and efficiency.

For glioma classification, particularly distinguishing between Astrocytoma and Oligodendroglioma, deep learning models face the challenge of identifying relevant tumor regions within large WSIs. Due to the computational infeasibility of processing entire gigapixel images, the standard approach is to divide WSIs into smaller patches. However, not all patches contain tumor tissue, and training models on irrelevant patches can degrade performance. To address this, we use a weakly supervised method to extract discriminative patches that represent tumor regions. These patches are then used to train CNNs, and their predictions are aggregated to classify the entire WSI. This approach has demonstrated significant improvements in classification accuracy, highlighting the potential of deep learning in histopathology.

Challenges and Future Directions While deep learning offers powerful tools for histopathology imaging, several challenges remain. The high resolution and large size of WSIs demand substantial computational resources. Additionally, most histopathology datasets provide only slide-level labels, necessitating methods that can handle weakly supervised learning scenarios. Another challenge is the variability in tissue staining and imaging conditions, which can affect model performance.

Future research in histopathology imaging aims to overcome these challenges by developing more efficient algorithms and exploring novel techniques for data augmentation and normalization. Expanding the range of detectable features and improving model robustness will be critical for broader clinical adoption. Moreover, integrating these advanced computational tools into clinical workflows will require user-friendly interfaces and validation in real-world settings.

In summary, histopathology imaging is a cornerstone of disease diagnosis, and the integration of digital imaging and deep learning techniques holds tremendous potential for enhancing diagnostic accuracy and efficiency. As these technologies continue to evolve, they promise to transform the field of pathology, leading to better patient outcomes and more effective healthcare delivery.

2.2 Related Work

The analysis of histopathological images has evolved significantly over the years, with a growing body of research focusing on improving the detection, classification, and interpretation of complex tissue structures, particularly with the advent of machine learning and deep learning techniques. This section provides a comprehensive overview of related work in the field, emphasizing approaches for glomeruli detection, classification in Lupus Nephritis (LN), and the classification of brain tumors using Whole Slide Images (WSIs).

Glomeruli Detection and Classification:

The detection and classification of glomeruli in histopathological images have been widely studied due to their importance in diagnosing and grading various kidney diseases, including Lupus Nephritis. Several works have been published that focus on detecting and classifying glomeruli using traditional

and machine learning approaches [3, 16, 20, 26]. However, there are only a few studies that propose a complete pipeline for the classification of WSIs into LN stages using glomeruli features.

Most existing methods focus on a limited number of glomeruli classes, such as sclerosed vs. normal [6] or glomerular vs. non-glomerular [15], which are insufficient for comprehensive LN classification. Although these approaches have contributed to the field, they fall short in addressing the complexity of LN classification, which requires a more nuanced analysis of glomerular features.

Ginley et al. [16] attempted to classify the six stages of LN using hand-crafted glomerular features. While effective, this approach is limited by the reliance on manually designed features, which may not capture the full range of variability seen in glomeruli. The predefined feature set potentially overlooks other relevant features that could enhance classification accuracy. Other studies, such as those by [9], explored glomeruli classification at the kidney level, manually extracting glomeruli and employing bulk labeling to circumvent the need for pathologist annotations. This method, however, introduced label noise, as all glomeruli images were assigned the same label as the LN class, irrespective of their phenotype. To mitigate this issue, they utilized an approximate Bayesian DenseNet, claiming that deep neural networks (DNNs) are resilient to label noise. Despite achieving notable results, the scalability and interpretability of this method from a pathologist's perspective remain uncertain. The need to consider the phenotypical features of glomeruli during LN classification is emphasized in [47], highlighting a gap in existing methodologies that our proposed end-to-end automated method seeks to address.

Our work diverges from these studies by proposing a fully automated pipeline that not only detects glomeruli but also classifies WSIs into various LN stages based on the phenotypical features of the glomeruli. This end-to-end approach aims to reduce the reliance on manual intervention, improve classification accuracy, and offer a scalable solution suitable for clinical application.

Weakly Supervised Learning in WSI Classification: The classification of WSIs is a computationally intensive task due to the large size and high resolution of the images. Traditionally, this task has been approached through supervised learning, requiring detailed annotations from pathologists. However, obtaining localized annotations for every WSI is not only challenging but also time-consuming,

often taking several hours per slide [1, 14, 30, 32, 34, 38, 45]. As a result, weakly supervised learning (WSL) has gained traction as a viable alternative, enabling the training of models with minimal annotations while still achieving high performance.

WSL techniques have been particularly successful in computer vision applications, where they have reduced the dependency on dense annotations [51]. These methods have been adapted to WSI analysis, where the image-level labels are used to train models under the Multiple Instance Learning (MIL) framework

Classification of Brain Tumors Using Whole Slide Images: The classification of brain tumors, particularly low-grade gliomas such as Astrocytoma and Oligodendroglioma, has also been a focus of research in histopathology. Traditional methods involve pathologists reviewing gigapixel-scale WSIs to locate tumor regions, a process that is both time-consuming and subjective. Recent advances have leveraged deep learning techniques to automate this process, particularly using weakly supervised learning approaches that require minimal annotation effort.

Given the computational challenges of training deep learning models on gigapixel WSIs, a standard practice has emerged to divide these images into smaller patches. However, not all patches represent tumor regions, necessitating the extraction of discriminative patches—those that contain relevant tumor features. Several methods have been proposed to tackle this challenge. For example, WELDON [13] and WILDCAT [12] demonstrated state-of-the-art performance for object detection and localization in weakly supervised settings, using only image-level labels.

In the context of WSI analysis, recent studies [8, 10, 23, 24, 36, 46, 49] have shown exceptional performance using WSI-level labels for classification tasks under the Multiple Instance Learning (MIL) framework. In MIL, a WSI is treated as a bag of instances (patches), where the bag is labeled positive if at least one instance is positive. This approach allows for effective learning from weakly labeled data, which is particularly useful in histopathology, where obtaining detailed annotations is often impractical.

Hou et al. [23] and Campanella et al. [8] utilized MIL variants to select discriminative patches from WSIs, showing excellent results. However, these methods often require large datasets, which may

not always be feasible. For example, Campanella et al.'s [8] method involved a dataset equivalent to 88 ImageNet datasets, raising concerns about the practicality of such approaches for smaller labs or institutions. On the other hand, Kanavati et al. [25] used a top-K patch selection strategy, which reduces the dataset size but may overlook important tumor regions that span multiple patches.

Bagari et al. [5] introduced an anomaly detection method using isolation forests to separate tumor tissue patches from normal tissue patches. While effective, this approach is computationally expensive when extended to large datasets. Courtiol et al. [10] employed a MinMax layer to select the top and bottom R patches from their 1D convolution layer output, offering another strategy to improve patch selection.

The classification of low-grade glioma subtypes has also been explored in studies that combine radiology and pathology images. For example, prior work [4, 27] utilized both types of images for subtype classification but was limited by a small dataset of only 30 patients. In contrast, our study uses a larger dataset of 286 patients from the TCGA portal, covering a broader range of tumor variations and improving the model's generalization. This approach is not only applicable to brain tumors but can be extended to other cancer types, offering a generalized framework for histopathological image analysis.

Comparison and Future Directions:

While significant progress has been made in both glomeruli detection and brain tumor classification, several challenges remain. The reliance on manual feature extraction in many LN classification studies limits the scalability and accuracy of these methods. On the other hand, the MIL-based approaches for WSI classification, while powerful, often require large datasets and complex computational resources.

Future research could focus on integrating these approaches, leveraging the strengths of both manual and automated methods to create more robust and scalable solutions. For instance, combining handcrafted features with deep learning models could enhance the interpretability and accuracy of LN classification. Similarly, exploring alternative weakly supervised learning techniques that require fewer computational resources could make WSI classification more accessible to smaller labs and institutions.

In summary, while the field has made considerable strides, there is still room for innovation, particularly in developing methods that are both scalable and interpretable. The integration of diverse data types, such as radiology and pathology images, and the use of larger, more diverse datasets, will be crucial in advancing the state of the art in histopathological image analysis.

Chapter 3

Classification of histopathology images using ConvNets to detect Lupus Nephritis

Systemic lupus erythematosus (SLE) is an autoimmune disease in which the immune system of the patient starts attacking healthy tissues of the body. Lupus Nephritis (LN) refers to the inflammation of kidney tissues resulting in renal failure due to these attacks. The International Society of Nephrology/Renal Pathology Society (ISN/RPS) has released a classification system based on various patterns observed during renal injury in SLE [47]. Traditional methods require meticulous pathological assessment of the renal biopsy and are time-consuming. Recently, computational techniques have helped to alleviate this issue by using virtual microscopy or Whole Slide Imaging (WSI). With the use of deep learning and modern computer vision techniques, we propose a pipeline that is able to automate the process of 1) detection of various glomeruli patterns present in these whole slide images and 2) classification of each image using the extracted glomeruli features.

3.1 Overview

Systemic Lupus erythematosus (SLE) is a multisystem autoimmune disease, in which the person's own immune system attacks and damages the healthy tissues and organs. SLE can attack a wide range of places in the body. Lupus Nephritis (LN) refers to the case of SLE affecting the kidney. If left unchecked, LN can progress to renal failure. The classification of lupus nephritis is critical to the issue

Patient ID	LN class	No. of Images
P1-2984	IV	4
P2-3073	III	6
P3-3075	VI	5
P4-3409	IV	7
P5-3452	IV	2
P6-3729	III+V	11
P7-3463.20	V	4
P8-3526.20	V	14
P9-3594.20	IV	10
P10-3729.20	IV+V	14
P11-3780.20	IV	7
P12-3899.20	III+V	13
P13-4130.20	IV	3
P14-4152.20	not specified	8

Table 3.1 NIMS dataset

of patient care and for the comparison of outcome results and therapeutic trials between different clinics. The International Society of Nephrology/Renal Pathology Society (ISN/RPS) has released the classification of LN into 6 classes (Class I-VI) based on the presence of various glomeruli features seen in patients affected by LN [47]. The classification also states about the co-occurrence of LN Class IV with Class V and Class III with Class V based on the co-occurrence of glomeruli features in these classes. The features are summarized in Table 3.2. The major diagnosis of LN is done by a renal biopsy; the observed histological changes within glomeruli guide the treatment of patients and are correlated with patient outcomes. Several histological stains such as PAS and H&E reveal different information regarding the disease features. Digitizing such stained slides into images helps in developing computational algorithms that aid pathologists during diagnosis.

Typically the biopsy images are stored as Whole Slide Images(WSIs). A WSI is a digitized image of an entire microscopy slide and has a huge resolution (approx. $80,000 \times 20,000$). Due to the unavailability of the image scanner, the dataset used in the current method was taken by attaching a camera to the microscope and taking snapshots of the interested regions of the biopsy tissue sample. The current pipeline incorporates the Periodic Acid-Schiff (PAS) stained images as the input images. Contrary to the numerous works published for glomeruli detection and classification [3, 16, 20, 26], only a few works

Class	count
Normal	18
Sclerosed	25
Endocapillary Hypercellularity	19
Messangial Hypercellularity	10
ThickGBM	41
Wireloops	10
Hyaline Trombi	5
Cresent	14
Segmental Adhesion	2

Table 3.2 Glomerular classes

exist that propose a complete pipeline for the classification of WSIs into LN classes using glomeruli features. Additionally, they only consider a limited number of glomeruli classes: sclerosed v/s normal [6] or glomerular v/s non-glomerular [15], which is insufficient for LN classification. [9] focused on the glomeruli classification and Kidney-level classification. They extracted the glomeruli manually and used a bulk labeling scheme in order to avoid manual annotation by pathologists. This technique adds label noise in the dataset as glomeruli images were given the same label as the LN class regardless of their phenotype. They further used an approximate Bayesian DenseNet to resolve this issue claiming that DNNs are resistant to label noise. Although, they show significantly good results using this method, the scalability and genuinity from a pathologist's view seem unclear. As reported in [47], it is important to consider the phenotypical features of glomeruli present in the kidney while in the classification of LN. Ginley et al. [16] shows the classification of 6 stages of LN by using hand-crafted glomerular features. In our method, we propose an end-to-end automated method from extracting glomeruli from the images to classifying them into various stages of Lupus Nephritis. As shown in figure 3.1.

3.2 Data

Biopsy samples from 14 patients diagnosed with various stages of LN were collected from NIMS shown in Table 3.1. In our study, we are using only PAS stain images. Vahadane [44] color normalization was performed in order to normalize the color variation across the images. Due to less data, we are

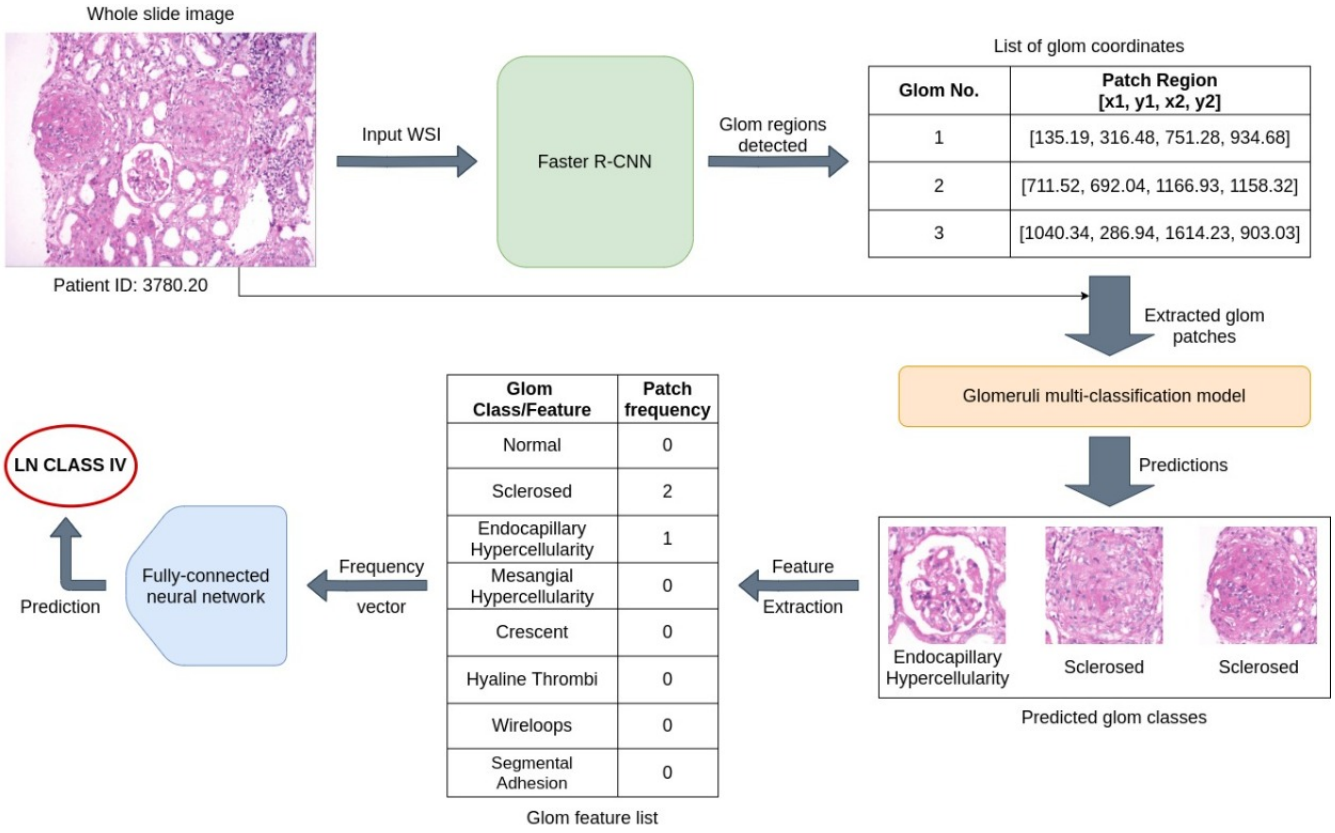

Figure 3.1 Pipeline of the proposed method

considering images of 10x, 20x, and 40x magnification scales. Aperio ImageScope1 software tool was used to annotate the glomeruli from the provided images. A bounding box is drawn around the glomeruli regions present in each image. The coordinates of these bounding boxes are stored in a .xml file and are used to extract the glomeruli regions. A total of 124 glomeruli were extracted from the 14 patients provided. The labels of these glomeruli were annotated and validated by Dr. Megha Uppin from NIMS, Hyderabad. Based on the classification of LN proposed by The International Society of Nephrology/Renal Pathology Society (ISN/RPS), the extracted glomeruli were divided into 9 classes based on the glomerular features seen in LN affected samples. The 9 classes selected and their respective number of glomeruli images extracted for each class are shown in Table 3.2. Figure 3.2 shows some examples showing varied features seen across the glomerular classes.

The detection pipeline for detecting different stages of Lupus Nephritis consists of 3 stages:

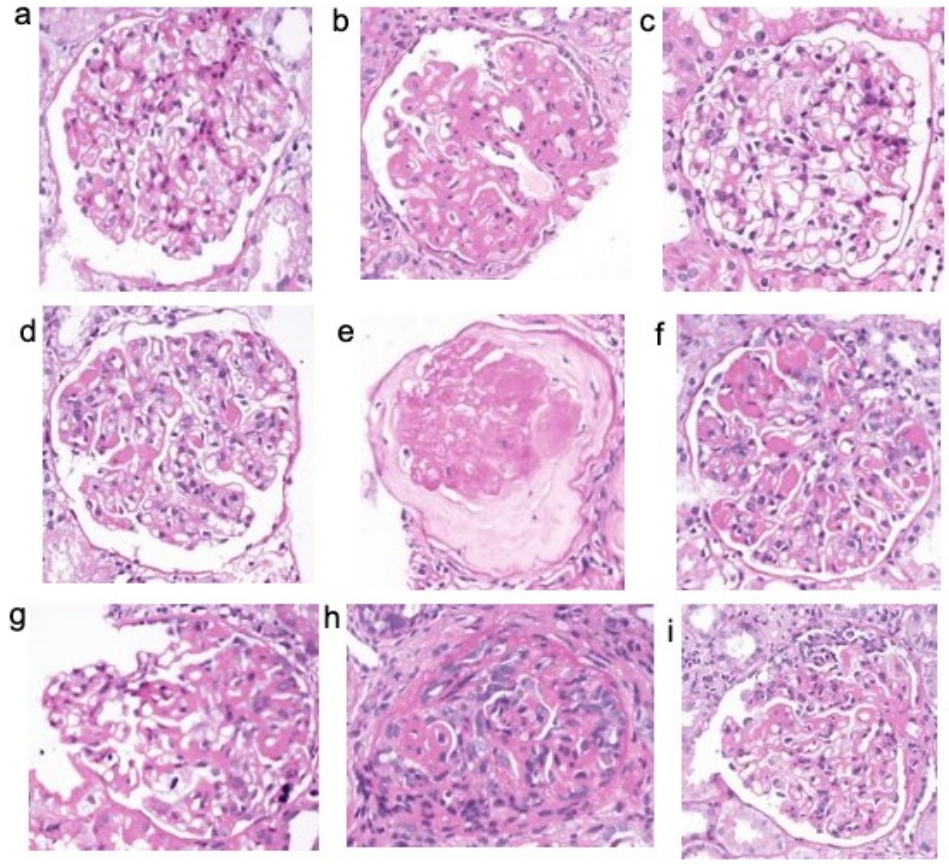

Figure 3.2 Examples showing the types of affected glomeruli. From left to right, a) Normal, b) ThickGBM, c) Mesangial Hypercellularity, d) Endocapillary Hypercellularity, e) Scerosed; f) Hyaline Thrombi, g) Wireloops, h) Crescent, i) Segmental Adhesion

3.2.1 Glomeruli detection

Although there are many approaches towards segmenting glomeruli [3, 6, 15–17, 26], they use very deep architectures for this task. Due to the limitations on the data availability from NIMS and the annotation effort required, we formulated the problem of extracting glomeruli as an object-detection task. Object detection is one of the common problems prevalent in the domain of computer vision. With the advantage of having large published datasets, deep learning has proven to provide state-of-the-art results for solving this problem. The general idea behind these deep learning methodologies is that they locate the smallest bounding box that comprises the complete object and also predict the class of the object. In the past years, there have been many methodologies for solving these problems based on the concept of region proposals [18, 19, 41] or one-shot detection [33, 39, 40] and various others [28, 31, 43]. We employed a region proposal-based technique and used a Faster R-CNN model for locating bounding boxes for the glomeruli present in the images provided by NIMS. We used a VGGNet pre-trained on ImageNet as the base network for generating the convolutional feature map. A Region Proposal Network (RPN) is used to find bounding boxes (also called region proposals) with the probability of containing an object. Further Region of Interest (RoI) pooling is applied to extract features for the proposals which is used by an R-CNN model for predicting the class of the object and its corresponding bounding box (defined by x_{center}, y_{center}, $width$, $height$). We fine-tuned the Faster R-CNN model on the slide images provided by NIMS for 30 epochs with a batch size of 5. The model gives the best estimates for the bounding boxes of the glomeruli present in the slide image. Based on the predicted coordinate values we crop the glomeruli from the slide images and only take the true positives for further classification.

3.2.2 Glomeruli Multi-Classification

With the extracted dataset of glomeruli images, we further classified them into the 9 classes shown in Table 3.1. Since a glomeruli image can have multiple labels, we used a multi-label multi-class classification approach. The classification model is shown in Figure. The model accepts the glomeruli image and passes it to 2 convolutional blocks (CONV $->$ BATCH_NORM $->$ RELU $->$ MAXPOOL).

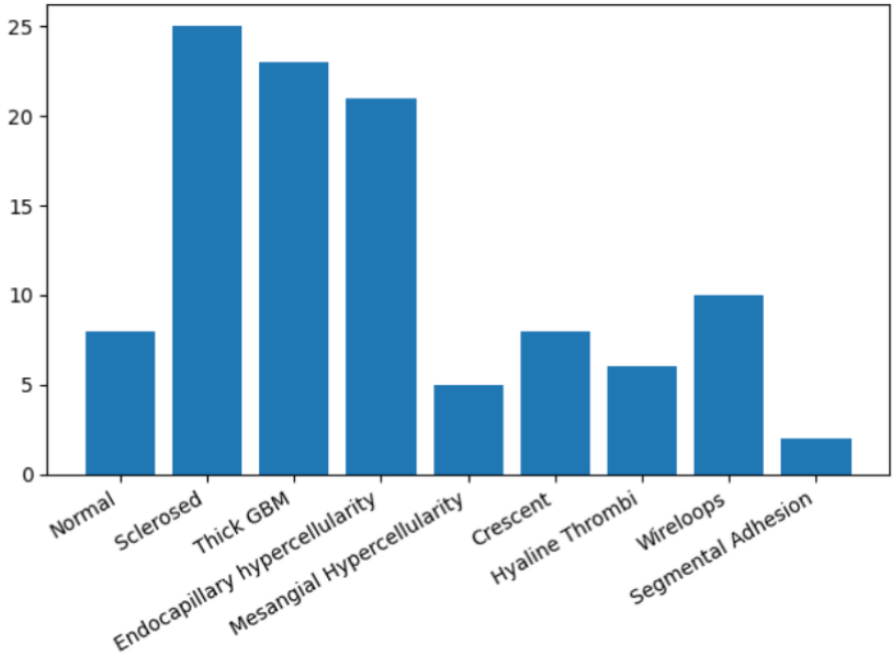

Figure 3.3 Glomeruli label imbalance

The resulting feature map is flattened to a feature vector and passed to a classifier with a sigmoid activation to output a vector of size 9×1 containing probabilities for each class. We found the best threshold value to be 0.45. If the probability value of a class is greater than the threshold, we assign the class to the image.

3.2.3 LN classification

Further, we aggregate all the glomeruli images for each patient based on their patient ID and generate a list of features. Once the glomeruli features list is obtained for each patient that depicts the types of glomeruli features present along with their counts, the final prediction of the LN class of the patient needs to be performed. We are currently exploring various aggregation methods to predict the slide level LN class from the glomeruli features list obtained for each patient. [23] proposes a Slide level level aggregation from patch-level labels. A similar approach can be used by learning a patient-level LN classifier based on the glomeruli feature list. A histogram/frequency vector can be generated for

	$\text{IoU}_{0.5:0.95}$	$\text{IoU}_{0.5}$	$\text{IoU}_{0.75}$
AP	0.73	0.976	0.843
AR	0.742	0.792	0.792

Table 3.3 Object detection scores on NIMS validation set.

each patient based on the obtained glomeruli features list and this can be input to a logistic regression model or an SVM to get the patient-level LN classification.

3.3 Results

3.3.1 Glomeruli detection

Out of the 108 images from NIMS, 100 images were kept for training the Faster R-CNN model and 8 for validation. The dataset was divided using patient ID such that slide images from the same patient were not found in both the training and validation sets. This was done to prevent the model from learning any kind of bias towards a patient. Table 3.3 shows the results on the validation set. We used the following metrics for evaluation:

- **Intersection over Union (IoU)**: It is defined as the area of the intersection divided by the area of the union of a predicted bounding box (B_p) and the ground-truth bounding box (B_gt).

$$IoU \ = \ \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})}$$

- **Average Precision (AP)**: It is defined as the area under the interpolated precision-recall curve, which can be calculated using the following formula:

$$AP \ = \ \sum_{i=1}^{n-1} (r_{i+1} \ - \ r_i) \ * \ p_{interp}(r_{i+1})$$

where $r_1, r_2, r_3, ..., r_n$ are the recall values (in ascending order) at which the precision is first interpolated.

- **Average Recall (AR)**: It is the recall averaged over all $IoU \in [0.5, 1.0]$ and can be computed as two times the area under the recall-IoU curve:

$$AR = 2 * \int_{0.5}^{1} recall(0)do$$

where o is IoU and $recall(o)$ is the corresponding recall.

3.3.2 Glomeruli multi-classification

We extracted 144 glomeruli images from the slide images. We trained on 100 glomeruli images and kept 44 glomeruli images for validation. For multi-label multi-class classification, the model was also pre-trained on DATASET_B_DIB by Bueno et. al. [7] for 20 epochs. This dataset contains 2340 glomeruli images of size 224×224 extracted from 31 WSIs of PAS-stained kidney biopsies generated in the European project AIDPATH. The dataset contains only 2 classes of glomeruli: normal (1170 images) and sclerosed (1170 images). The pretraining results on the test set (200 images) are shown in Table 3.4. We fine-tuned the model for another 20 epochs on our extracted set of glomeruli images from NIMS data. To deal with label imbalance as shown in Figure **??**, weighted oversampling was used. Table 3.5 shows the results of the multi-label multi-class classification on the validation set.

Acc	Precision	Recall	F1-score
97.9%	0.974	0.99	0.984

Table 3.4 Results for pre-training on Bueno et al. [7] dataset

Acc	Precision	Recall	F1-score
72.1%	0.87	0.7	0.775

Table 3.5 Results for multi-label multi-class classification

Classes	Accuracy
Sclerosed vs endHyper	0.94
MesHyper vs endHyper	0.84
Sclerosed vs normal	0.98
Sclerosed vs mesHyper	0.93
thickGBM vs endHyper	0.75
thickGBM vs MesHyper	0.85
thickGBM vs normal	0.84
thickGBM vs Sclerosed	0.74

Table 3.6 Results for binary classification

3.3.3 Binary Classification between glomeruli features

In addition to the mentioned pipeline, we have also performed a binary classification between the selected glomeruli features mentioned in Table 3.2 in order to verify the distinguishable features between the glomeruli classes. A glomerulus can belong to a single class or multiple classes from Table 3.2. We chose a combination of classes whose features cannot both occur in the same glomerulus, for example, a glomerulus cannot be normal and have any of the other features, and a combination of classes, where more than one feature might occur in the same glomerulus leading to multiple classes for a single glomerulus. The images were divided into train and validation sets. We perform data augmentations to increase the training data and keep the validation data as the original images. It is clear that there is a data imbalance between the classes as seen in Figure 3.3, which was addressed by oversampling the minority class and under-sampling the majority class in each batch during training by using a weighted random sampler. A simple CNN pre-trained on Mendeley Dataset [7] for the binary classification between normal and sclerosed glomeruli classes was used for the binary classification task. Table 3.6 summarizes the result of binary classification performed between classes with considerable individual data samples showing the combinations taken for binary classification and the accuracy achieved on the respective validation set.

3.4 Challenges

The major challenge faced is the availability of a published/verified annotated dataset for LN classification. The images provided for each patient are not in the WSIs, which gives the picture of the entire microscopic image of the tissue taken and can approximately have hundreds of glomeruli from each such image. This also prevents us from including state-of-the-art glomeruli detection methods in our pipeline which could increase the performance. Our data consisted of images taken covering a small region of tissue usually having around 1 glomerulus for a $40x$ magnification image, 2-3 glomeruli in a $20x$ magnification image, and 5-7 glomeruli in a $10x$ magnification image. Hence the total number of glomeruli extracted from each patient in our pipeline reduces significantly when compared to WSIs. This affects the performance of the classification of glomeruli into the 9 glomerular classes due to less and highly imbalanced data. Secondly, with the total number of patients being only 14, training a patient-level LN classifier becomes tough due to the smaller number of samples of each LN class covered. Finally, the glomerular feature list obtained from the extracted glomeruli is only from the provided set of images for each patient, which raises the question of whether it is clinically correct to predict the LN class of patients based on the limited set of images provided for each patient.

3.5 Summary

We propose a complete pipeline for the classification of WSIs into various LN stages based on glomeruli features. The steps include 1) detection of glomeruli from PAS stained images, 2) multi-classification of these glomeruli into the 9 classes selected based on the glomeruli features seen in LN affected kidney biopsies, and 3) predict the patient's stage of Lupus Nephritis (CLASS I to Class VI).

Chapter 4

Weakly supervised method for the classification of Astrocytoma and Oligodendroglioma using histopathology images

Analysis of digital histopathology slides or Whole Slide Images (WSI) is critical for several diagnoses. Pathologists review gigapixels of whole slide images from these slides to locate tumor regions. Low-grade glioma is a type of brain tumor that originates from glial cells. In our work, we perform the classification of low-grade gliomas into Astrocytoma and Oligodendroglioma. Due to the computational infeasibility of training deep learning models on gigapixel images, the standard practice is to divide them into smaller patches. However, not all the patches represent the tumor region. Discriminative patches are those that have the features of the tumor. Our work presents a weakly supervised method for extracting such discriminative patches representing the tumor region from each WSI. The Convolutional Neural Network (CNN) is trained on the discriminative patches for the final classification and aggregated to find the WSI label. We evaluated our method on a dataset of 5, 81, 616 patches extracted from 286 WSIs obtained from The Cancer Genome Atlas (TCGA) portal and achieved a slide-wise accuracy of 79.31%. Our results show a further increase of 10% when trained only on discriminative patches extracted from each WSI and gained a slide-wise accuracy of 89.65%.

4.1 Overview

Histopathology uses the microscopic examination of a biopsy or surgical specimen that is processed and fixed onto glass slides to study the disease symptoms [2]. It is a critical step in tumor detection and diagnosis. The tissue sections are counter-stained to visualize different cellular components under a microscope. Pathologists have used Hematoxylin-Eosin (H &E) staining [4] for over a hundred years. Hematoxylin stains cell nuclei in blue, while eosin stains cytoplasm and connective tissue in pink. Such histological sections are digitized into an image known as a Whole Slide Image (WSI). Recent developments in deep learning-based approaches have shown to be beneficial for automating various medical image tasks in histopathology [11, 22, 27, 35, 42, 48] and aid the pathologist in tumor detection tasks. WSI are gigapixel images with a high resolution of around $1,00,000 \times 1,00,000$ pixels. Training deep learning models on such large images is computationally infeasible. Fine-grained feature details seen in the tumor regions, essential to differentiate between the tumors, are lost by resizing images to smaller dimensions. Therefore, it is more reasonable to perform analysis on small patches with fine details cropped from WSIs as small blocks of the desired size, but on a much larger scale as thousands of patches are extracted from each WSI. Such features can be captured and learned from the high-resolution patches extracted from the WSI and predict the slide-level label based on the patch-level predictions. Therefore, most WSI classification methods classify or extract features from the patches and aggregate them to find the WSI label. Most histopathology datasets often comprise large images coarsely annotated according to the diagnosis, providing only the ground truth label for each WSI (slide-level label), but extracted patch labels remain unknown. The WSI label gets translated to the patches leading to a weakly supervised scenario. Developing deep learning models that give a good classification accuracy without pixel-level annotations based only on slide-level labels reduces the cost, and time spent in manual annotations and is very beneficial to histopathology image analysis. Secondly, the entire tissue region present in the WSI does not represent the labeled tumor. It is a tissue section covering regions of the tumor features and normal tissue. The patches extracted from WSI that have features and qualities of the WSI tumor label are considered discriminative patches. Since the WSI con-

tains the tumor and normal tissue regions, selecting discriminative patches from each WSI is essential to represent the tumor region. It might help in improving the performance of the model. Various studies [5, 8, 10, 23, 25, 37] to select the discriminative patches from each WSI have proven to be successful in improving the classification accuracy.

Contributions: A glioma is a type of brain tumor that originates from glial cells. Based on cell activity and aggressiveness, glial cancers are categorized into two types (i,e) High-Grade Glioma (HGG) and Low-Grade Glioma (LGG). In this paper, we propose a weakly supervised deep learning method to select discriminative patches and perform the classification of subtypes of LGG – Astro- cytoma, and Oligodendroglioma using histopathology images. This subtype classification is a challenging task because of the tumor cell diversity of gliomas. This work classifies the WSIs into two labels- Astrocytoma and Oligodendroglioma and the selected discriminative patches from each WSI are used to train the model and report the final classification result. We observe a 10% increase in the slide-wise accuracy from 79.31% to 89.65% when trained only on the discriminative patches selected from each WSI. We also show the high-probability regions predicted by the model which have been validated by pathologists. Although this method is applied to classify brain tumors into Astrocytoma and Oligodendroglioma, it is a generalized method and can be extended to any cancer type and future works can be extended to the multiclass classification of different types of glial cancers (Normal vs. Astrocytoma vs. Oligodendroglioma vs. Oligoastrocytoma vs. Glioblastoma). The pipeline of the proposed method is explained in figure **??**.

4.2 Approach

This work classifies the WSIs into two labels - Astrocytoma and Oligodendroglioma. Further, the discriminative patches that represent the tumor features are selected from each WSI, and analyze the effect on the classification performance when the model is trained only on such discriminative patches. The pipeline of the proposed method is explained in Figure 1.

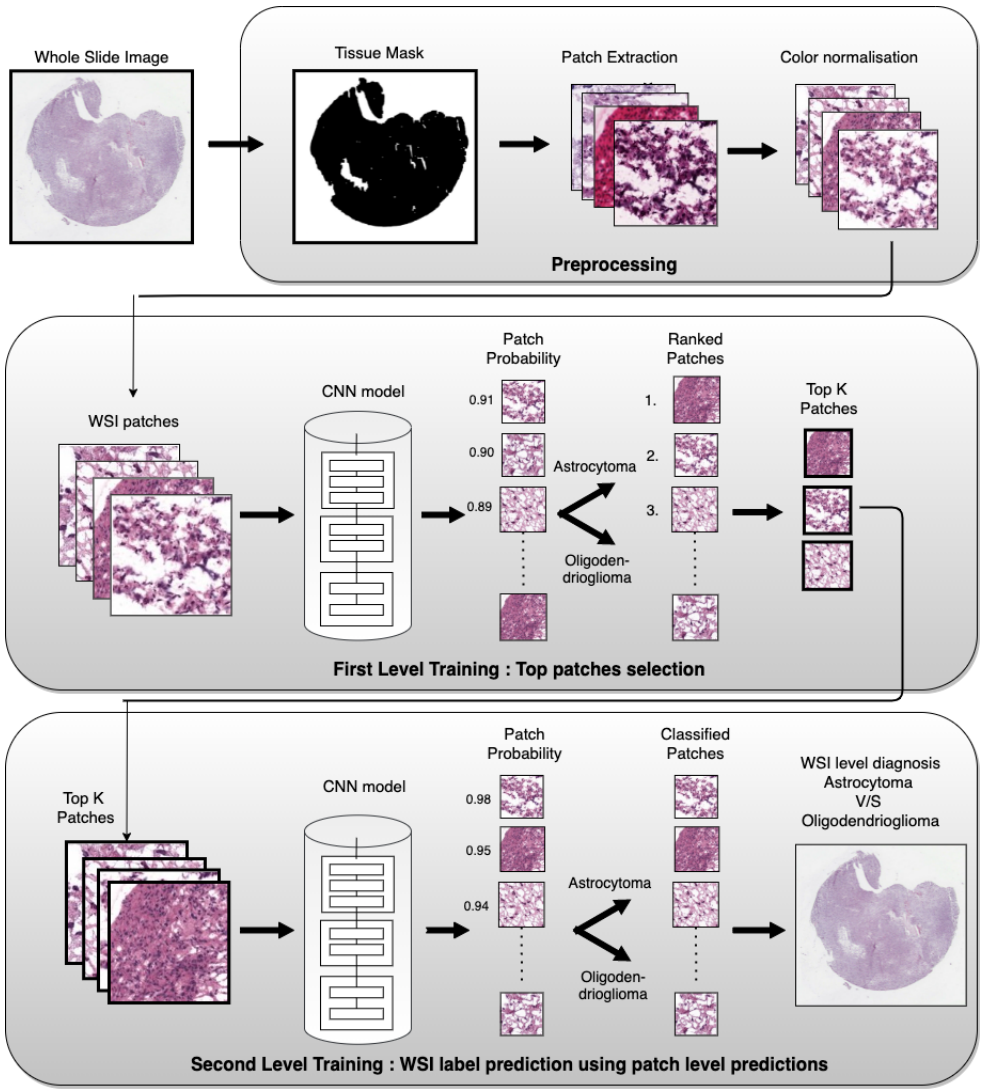

Figure 4.1 Pipeline of the proposed method. Workflow: a. Preprocessing steps are applied on each WSI to obtain patches. b. First-level training is performed on the patches, and the selection of top K percentile patches from each WSI is made based on their ranked sigmoid probabilities. c. Second-level training is performed on the selected discriminative patches and the obtained patch level predictions are used to calculate the WSI label.

4.2.1 WSI Preprocessing

We downloaded whole slide images from The Cancer Genome Atlas (TCGA) [21], an extensive data repository covering various cancers. The average dimensions of WSIs in our dataset are around $60,000 \times 30,000$. H&E staining has been done to preserve the cellular and tissue structure. We have considered 177 WSI from 130 patients available in Astrocytoma-Anaplastic and 109 WSI from 76 pa-

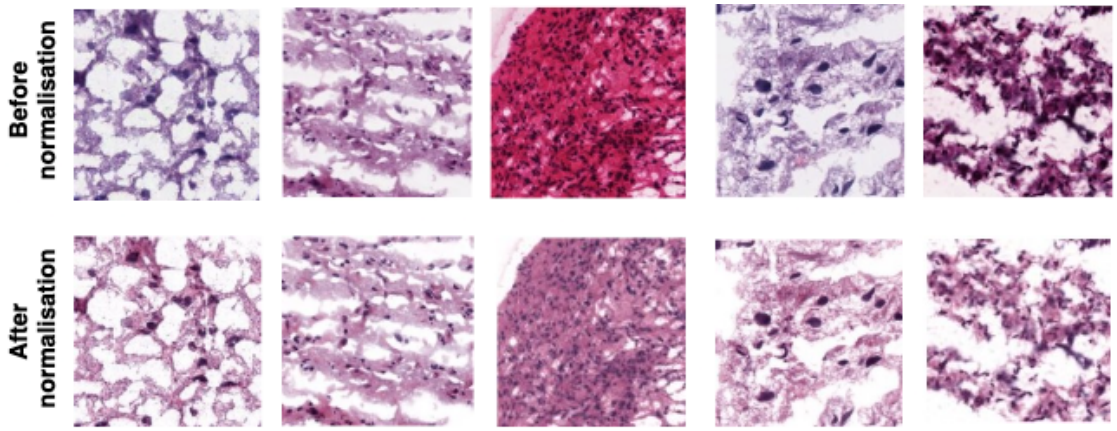

Figure 4.2 Patches shown in the top row are of different colors as they are obtained from different WSIs. Patches shown in the bottom row are after applying color normalization to achieve a uniform color across patches.

tients available in Oligodendroglioma-Anaplastic frozen specimens at 40× magnification level. Significant class imbalance exists in the dataset. We extracted the patches of size 448×448 without any overlap from each WSI at a 40× magnification level. Patches that contain noise or white backgrounds were discarded to save the computation cost and reduce the noise.

4.2.1.1 Patch filtering

Otsu binarization method [52] has been used to find the optimal threshold value to separate the background from the foreground. Binary dilation [29], is applied to enlarge bright regions and shrink the dark regions. We empirically found the optimum limit for both the threshold value returned by Otsu binarization and the ratio found from binary dilation. All the qualified patches were color normalized [44] to reduce the color variations among the patches across all slides and have been considered for training while all other patches are discarded. Pseudocode can be seen in algorithm 1.

Algorithm 1 Foreground patch selection and elimination of background patches

```
slide = Whole slide image
while patch in slide do
    gray = GRAYSCALEIMAGE(patch)
    (ret, threshold) = THRESHOLD(gray, Otsu + BinaryInversion)
    threshold = BINARYDILATION(threshold, iterations = 15)
    threshold_mean = MEAN(threshold)
    if ret < 200 and threshold_mean > 0.80 then
        Save the patch
    else
        Discard the patch
    end if
end while
```

4.2.2 WSI patch selection and classification

The objective is to predict whether a slide image is Astrocytoma or Oligodendroglioma. We formulate our problem as patch-level classification and predict the slide-level label based on the patch-level predictions. The proposed method can be divided into three parts.

4.2.2.1 First-level training

After performing the preprocessing steps, patches were extracted from each WSI. The WSI label is assigned to all patches extracted from it. Such patches across the available WSIs are passed as input to the ResNet-18 model pre-trained on ImageNet and are fine-tuned till convergence. The sigmoid function at the end of the last layer of the ResNet-18 model gives a probability of each patch belonging to the two classes. We refer to this network as First-level. The classification result obtained by first-level training can be considered as the baseline result obtained by using all the patches.

4.2.2.2 Selection of discriminative patches

Recent works on WSI classification that used the MIL-based approach for selecting discriminative patches have proven successful in achieving good results. However, non-discriminative patch elimination is done based on the probabilities predicted by the CNNs trained for 2-3 epochs in each iteration. When we trained the model for a similar number of epochs, the best patch-wise validation accuracy was approximately 57%, closer to random choice. To address this problem, we have trained the model till convergence and selected the discriminative patches based on the probabilities predicted by the model with comparatively much higher validation accuracy. Patches classified correctly with high confidence are hypothesized to represent the tumor region. Top K percentile patches classified correctly with high confidence are selected from each WSI based on the sigmoid scores of each patch obtained from the first-level training. These patches selected from each WSI represent the tumor regions better than considering all the patches as tumor patches.

4.2.2.3 Second-level training

In the second-level training, the fine-tuning of the last few layers of the ResNet-18 model initialized with ImageNet pre-trained weights is trained using only the top K-percentile patches extracted from each WSI. We refer to this network as second-level. The classification result obtained can be compared with the baseline result obtained from first-level training and analyze the selection effect of top K percentile patches.

4.3 Experiment settings

We downloaded whole slide images from The Cancer Genome Atlas (TCGA) [21], an extensive repository of data covering various cancers and data types collected from many hospitals, providing a huge collection of WSIs. These WSIs are available at different magnification levels $(5\times, 10\times, 20\times, 40\times)$ with resolutions ranging from $1,000 - 1,00,000$ pixels. H&E staining has been done to preserve the cellular and tissue structure. We have considered 177 WSI from 130 patients available in Astrocytoma-

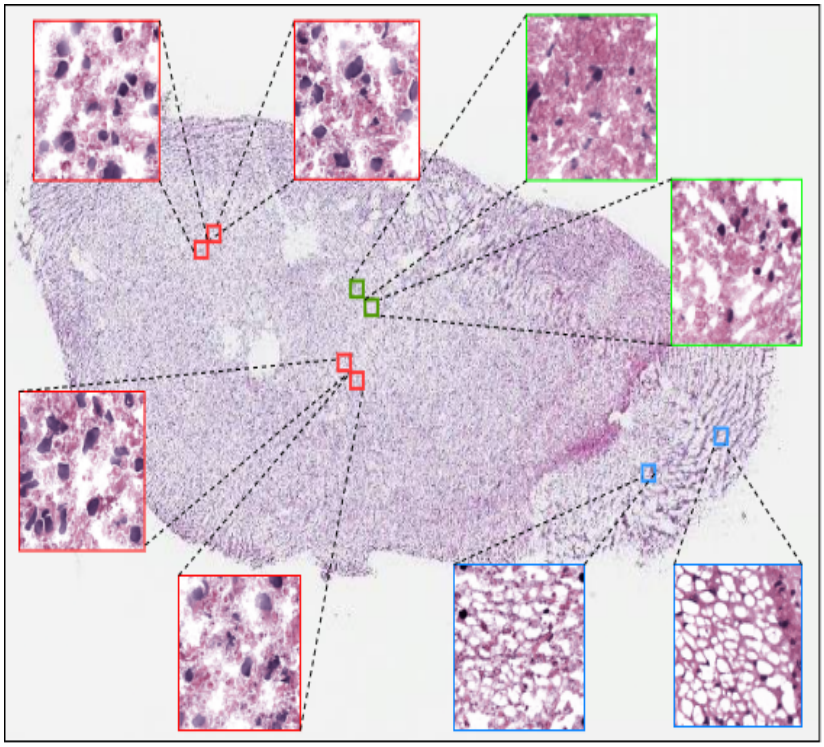

Figure 4.3 Astrocytoma WSI. Dimensions : 106624×28871 pixels. All the patches are shown at 40x magnification. The patches with red bounding boxes represent discriminative patches, green bounding boxes represent non-discriminative patches and blue bounding boxes represent artifacts.

Anaplastic and 109 WSI from 76 patients available in Oligodendroglioma-Anaplastic frozen specimens at $40\times$ magnification level. Significant class imbalance exists in the dataset. The average dimensions of WSIs in our dataset are around $60,000 \times 30,000$. We divided the dataset into 70-15-15 splits for training, validation, and testing.

4.3.1 Training

All the models are implemented in Pytorch. The input images are resized from 448×448 to 224×224. To introduce generalization, we applied data augmentation techniques like mirroring and random rotations to the training data. All the images were normalized using mean and standard deviation, calculated on the training set before being fed to the network. The last four layers of the ResNet-18 model have been fine-tuned using Adam Optimizer with 10^{-6} learning rate and binary cross-entropy as the loss function for 10-15 epochs during first-level training and 30-40 epochs during second-level

training. We used a weighted random sampler to deal with the class imbalance problem. The model was trained with batch size 128 in NVIDIA GTX 1080 Ti with 8 GB RAM. The hyperparameters are kept constant in both first-level and second-level training.

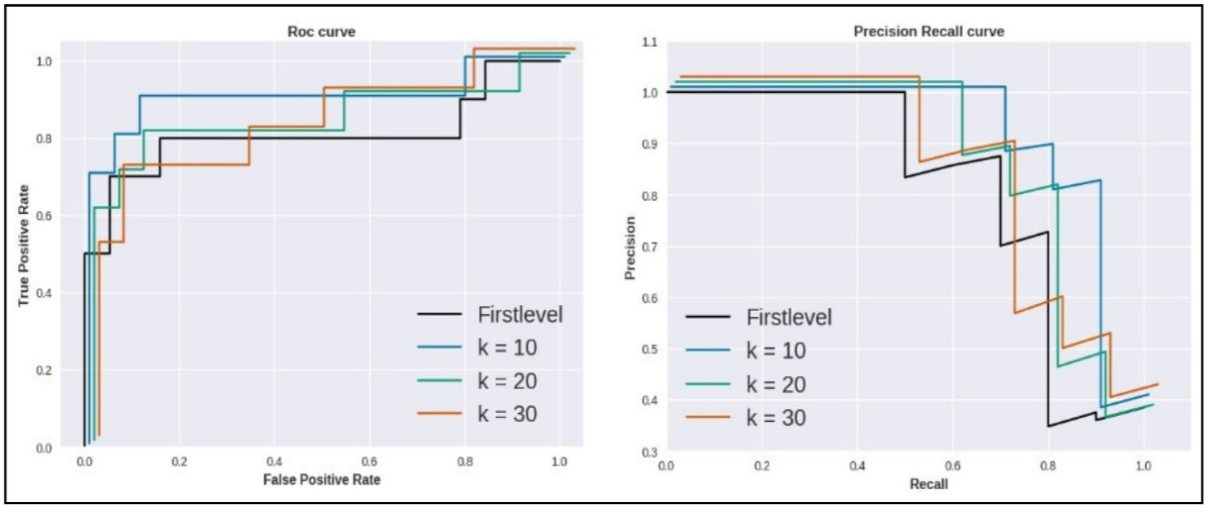

Figure 4.4 Receiver Operating Characteristic Curve (ROC) and Precision Recall Curve (PRC): Comparison of models trained in first level and second level ($K = 10, K = 20, K = 30$) has been shown using ROC on the left and PRC on the right. The curve with black, blue, green, and orange colors represents the models trained on all the patches extracted, top 10, 20, and 30 percentile patches respectively. The model trained on top 10 percentile patches ($K = 10$) performed the best.

.

The inference is made after the second-level training. During patch-level evaluation, all the patches extracted from a WSI in the test set are evaluated by the model. In the case of slide-level inference, the sigmoid probabilities of all patches in each WSI are averaged. If the mean value is greater than $\mathbf{0.5}$, then label $\mathbf{1}$ (Oligodendroglioma) is assigned, otherwise $\mathbf{0}$ (Astrocytoma). The result is influenced by the patches with extreme probabilities (either close to zero or close to one). The hypothesis is that patches with probabilities close to $\mathbf{0.5}$ which represent the non-discriminative patches, do not influence the result much.

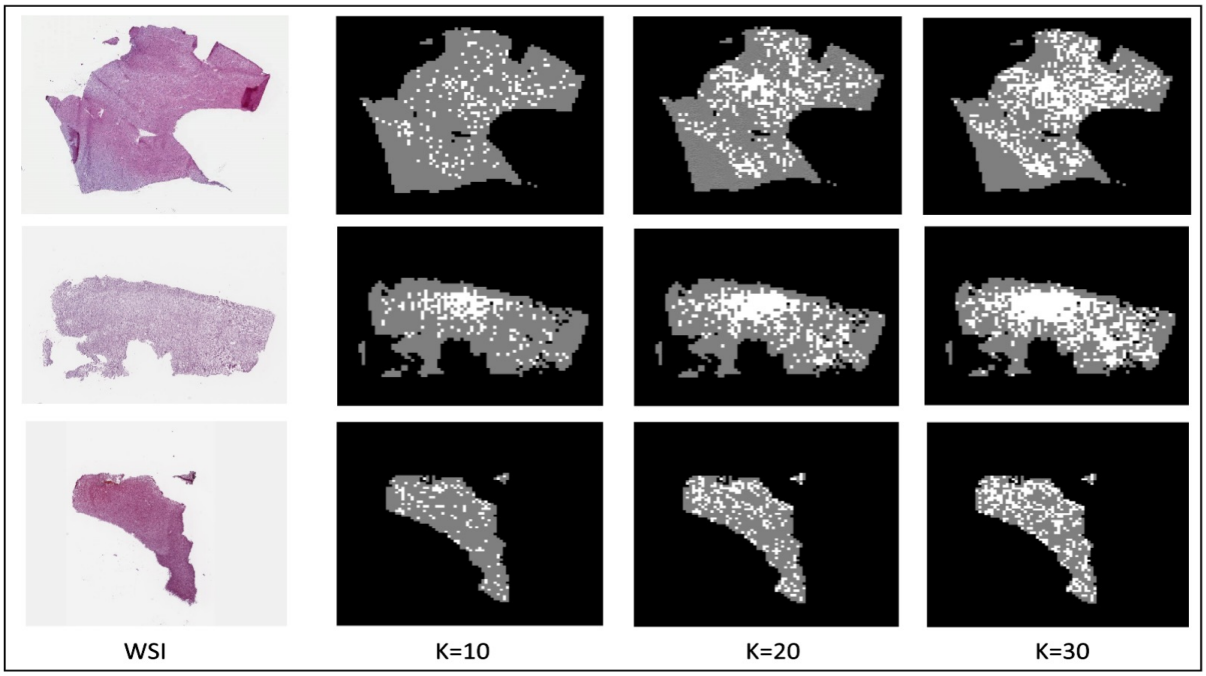

Figure 4.5 Each row represents the WSI and the high-probability areas from which patches for second-level training are selected. In Column 2,3,4, the white region represents the top 10, 20, and 30 percentile patches selected for second-level training while the grey region represents the areas with low probability patches and the black region represents the background area in the WSI.

4.4 Results

In the classification of Astrocytoma and Oligodendroglioma, the proposed classification framework obtained a slide-wise accuracy of 79.31% from first-level training of the last 4 layers in ResNet-18. Features computed by the first layers of a trained network seem to occur regardless of the exact cost function and image dataset, making it general. However, the features computed in the final layers of a trained network greatly depend on the chosen dataset and task, making it more specific. Thus first layer features are more general, whereas last layer features are more specific. We have tried varying the number of last layers selected to be trained on our target dataset. Table **??** shows the patch-wise, and slide-wise area under the receiver operating characteristic curve (AUC-ROC score) obtained by fine-tuning specified last n layers of the network on the target dataset, and the remaining first layers are frozen with pre-trained ImageNetweights. We observe that fine-tuning the ResNet-18 model on the last four layers of our target dataset yielded the best AUC ROC score. This model has been used for second-

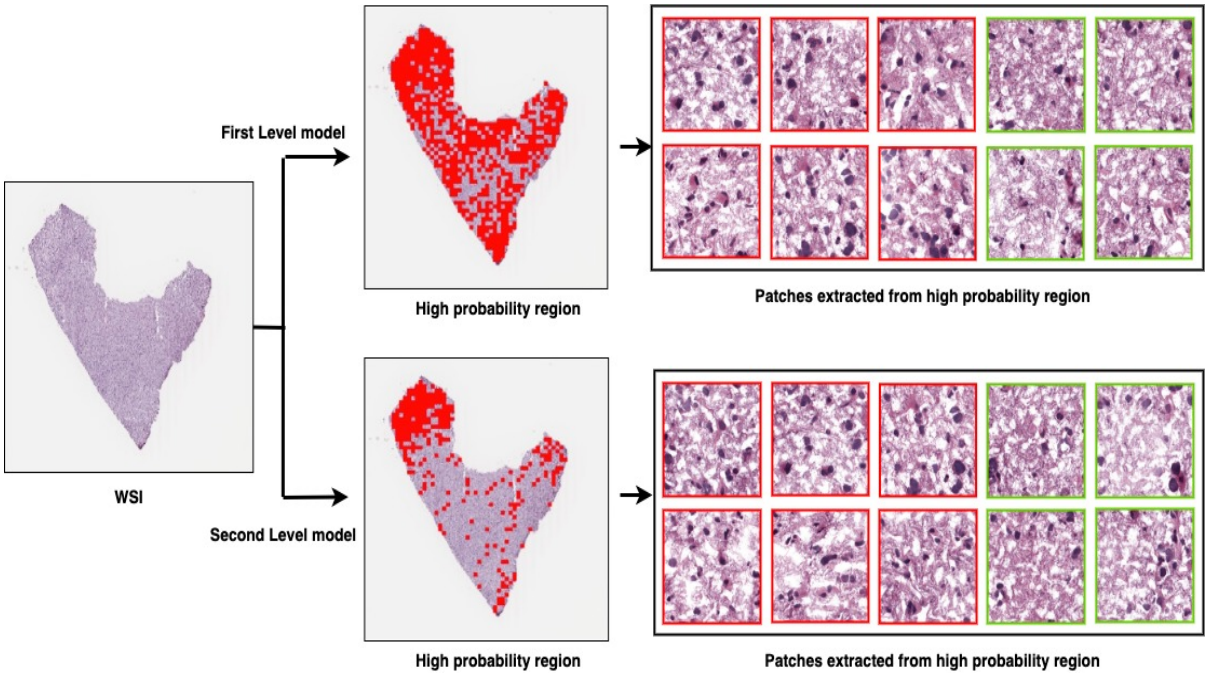

Figure 4.6 Comparison of high probability region predicted on a test set WSI by the first level model (baseline) and the best performing second-level model. The left image represents the WSI from the test set. Middle images represent the high-probability regions shown in red, which are predicted by the first and second levels models. The right images represent some sample patches taken from the respective high-probability regions. Patches shown in red bounding boxes represent the tumor patches and have high cellular foci. Patches shown in green bounding boxes represent regions with very low cellularity.

level training on top K percentile patches. The final result obtained for different K values is shown in Table 4.1. There is an increase of 10% in slide-wise accuracy when trained only on the top 10 percentile patches ($K = 10$) instead of considering all the patches for the baseline result obtained by first-level training. Table 3 shows the patch-wise, slide-wise accuracy, and AUC-ROC scores for the first-level and second-level training ($K = 10$). Precision recall curves and receiver operating characteristic curves of models fine-tuned on the last four layers in both first-level and second-level training on top 10, 20, and 30 percentile patches ($K = 10$, $K = 20$, $K = 30$ have been depicted in Fig 4.4. Despite the class imbalance, the model trained on the top 10 percentile patches (K = 10) has yielded the best average precision score of 91.07%. To avoid overlapping of the curves and a single color dominating over others, a very small offset value of 0.01, 0.02, 0.03 is added to $K = 10, K = 20, K = 30$ respectively along the X-axis and Y-axis to be able to distinguish the curves and visualize them clearly.

Table 4.1 Comparison of classification results of top 10, 20, and 30 percentile patches in second level training of finetuning last-4 layers of ResNet-18 Model
APS: Average Precision Score.

Experiment	Number of Patches	Accuracy	APS
K=10	40853	**89.65**	**91.07**
K=20	81525	82.75	85.19
K=30	121759	86.2	82.03

Table 4.2 Comparison of patch-wise and Slide-wise results in the first and second level with $K = 10$ obtained for fine-tuning the last 4 layers of ResNet-18 model.

	Experiment	Accuracy	AUC ROC
First Level	Patch Wise	65.34	67.09
	Slide Wise	79.31	79.47
Second Level	Patch Wise	69.84	67.43
	Slide Wise	**89.65**	**84.21**

Using the sigmoid probabilities predicted by the ResNet-18 model in the first-level training, a probability heat map is generated by highlighting the top K percentile patches for each WSI. These patches are then used for the second-level training. In Figure **??**, the white region shows the top K percentile patches selected for second-level training for $K = 10, 20, 30$ while the grey region shows the patches with low probability and the black region shows the background area in the WSI. Fig4.6 shows a comparison of high-probability regions predicted on a WSI from the test set by the first-level model and the best-performing second-level model (k = 10) trained on top 10 percentile patches. We can observe that the first-level model predicts most regions as high probability, whereas the second-level model, trained only on the discriminative patches, has given a much more concise high probability region. We have also sampled a few patches from these regions. We observe that the patches extracted from the high-probability region predicted by the second-level model have much fewer patches belonging to regions with very low cellularity, which were included in the first-level model's high-probability region. The qualitative results were validated by a pathologist.

Table 4.3 Patch-wise and Slide-wise AUC ROC scores of finetuning last n layers. The columns show the patch-wise and slide-wise AUC ROC scores obtained by First-level training and second-level training with $K = 10$, $K = 20$, and $K = 30$ respectively.

AUC-ROC	Experiment	First-Level	K = 10	K = 20	K = 30
Last 2 layers	Patch Wise	70.63	69.83	69.83	72.36
	Slide Wise	78.42	88.42	78.94	**83.68**
Last 3 layers	Patch Wise	69.84	74.42	72.82	73.35
	Slide Wise	77.89	86.84	**86.31**	**83.68**
Last 4 layers	Patch Wise	**71.26**	**75.97**	**75**	**73.47**
	Slide Wise	**81.05**	**90.52**	84.21	83.15
All layers	Patch Wise	67.09	67.43	71.13	70.53
	Slide Wise	79.47	84.21	85.78	82.10

4.4.1 Ablation study

Transfer learning is a method that initially trains the network on a general image dataset like ImageNet and transfers the learned feature to the network to be trained on the target dataset [50]. Features computed by the first layers of a trained network seem to occur regardless of the exact cost function and image dataset, making it general. However, the features computed in the final layers of a trained network greatly depend on the chosen dataset and task, making it more specific. Thus first layer features are more general, and last layer features are more specific. We have tried varying the number of last layers selected to be trained on our target dataset. Table 4.3 shows the patch-wise, slide-wise area under the receiver operating characteristic curve (AUC-ROC score) obtained by fine-tuning specified last n layers of the network on the target dataset, and the remaining first layers are frozen with pre-trained ImageNet weights. We observe that fine-tuning the ResNet-18 model on the last four layers of our target dataset yielded the best AUC ROC score.

4.5 Summary

This work presented the classification of brain tumors into Astrocytoma and Oligodendroglioma from histopathology images. We solved this problem in two stages. The first stage identifies the discriminative patches with the cancerous region, and the second stage trains them. In the second stage, the model's performance boosts by 10% in slide-wise accuracy when trained on the high probability patches keeping the other parameters constant. We conclude with the above results that training deep learning models on discriminative patches extracted from WSI based on slide-level labels gives accurate results. We believe that weakly supervised learning methods have great potential to assist pathologists in histology image diagnosis soon. A further improvement in the accuracy can lead to clinical adoption. Future works can be extended to the multiclass classification of different types of glial cancers (Normal vs. Astrocytoma vs. Oligodendroglioma vs. Oligoastrocytoma vs. Glioblastoma). Although this method is applied to classify brain tumors into Astrocytoma and Oligodendroglioma, it is a generalized method and can be extended to any cancer type.

Chapter 5

Conclusion and Future work

5.1 Conclusion

This thesis has explored the application of advanced machine learning techniques to the classification of histopathology images in two distinct medical domains: Lupus Nephritis (LN) and glioma subtypes (Astrocytoma and Oligodendroglioma). By developing specialized methodologies for each problem, we have demonstrated significant improvements in classification accuracy, providing valuable tools for enhancing diagnostic processes.

Classification of Lupus Nephritis In the context of Lupus Nephritis, we proposed a comprehensive pipeline designed to classify whole slide images (WSIs) into various LN stages based on glomerular features. The process encompasses the following steps:

Detection of Glomeruli: Utilizing PAS-stained images, we developed methods to accurately detect glomeruli, which are critical structures in kidney biopsies. Multi-Classification of Glomeruli: We classified the detected glomeruli into nine distinct classes. These classes were selected based on the specific glomerular features observed in LN-affected kidneys, providing a nuanced understanding of the disease's progression. Stage Prediction: Using the classified glomeruli, we predicted the patient's stage of Lupus Nephritis, ranging from Class I to Class VI, as defined by the International Society of Nephrology/Renal Pathology Society (ISN/RPS). This automated approach addresses several limitations of existing methods, such as label noise and limited class consideration. By incorporating phenotyp-

ical features and providing a detailed classification scheme, our pipeline enhances the accuracy and reliability of LN diagnosis, ultimately contributing to improved patient care.

Our work has several important implications. To our knowledge, this is the first study to leverage convolutional neural networks (CNNs) for detecting Lupus Nephritis using visual glomerular features. While previous studies like Ginley et al. used hand-crafted glomerular features for classifying LN stages, our approach allows deep learning models to implicitly learn all relevant features, thereby broadening the feature set. We emphasize that this work is not intended to replace neuropathologists but to assist them. However, there are potential negative implications. Due to the limited availability of LN data, we used a set of nine commonly seen glomerular classes, whereas there are more than 25 glomerular classes identified in some patients. Including images with such additional glomerular classes could alter predictions and increase false positives.

Classification of Glioma Subtypes For the classification of glioma subtypes, specifically Astrocytoma and Oligodendroglioma, we introduced a weakly supervised deep learning method. This approach addresses the challenges posed by the high-resolution and weakly annotated nature of WSIs. Our methodology involves two main stages:

Identification of Discriminative Patches: We identified high-probability patches within WSIs that contain the cancerous regions. This step is crucial for focusing the analysis on the most relevant areas of the tissue, thereby enhancing the model's learning process. Training on Discriminative Patches: By training the model on these selected patches, we achieved a significant improvement in classification accuracy. Specifically, the slide-wise accuracy increased by 10% compared to training on the entire WSI. Our results demonstrate that training deep-learning models on discriminative patches based on slide-level labels can yield accurate and reliable classification outcomes. This method shows great potential in assisting pathologists with histology image diagnosis, providing a scalable and effective solution for tumor classification. Further improvements in accuracy could facilitate clinical adoption, thereby enhancing diagnostic workflows and patient outcomes.

This method, although applied specifically to classify brain tumors into Astrocytoma and Oligodendroglioma, is generalized and can be extended to other types of cancer. Future work can focus on the multiclass classification of various glial cancers, including Normal, Astrocytoma, Oligodendroglioma, Oligoastrocytoma, and Glioblastoma. Expanding the method's applicability can further validate its utility across different cancer types.

5.2 Future Directions

The findings of this thesis underscore the potential of machine learning to revolutionize histopathology image analysis. However, several avenues for future research and development remain:

Enhancement of Weakly Supervised Learning: Further refinement of weakly supervised learning techniques could improve model accuracy and robustness, making them more suitable for clinical application. Extension to Other Cancer Types: The methodologies developed for glioma subtype classification could be extended to other types of cancer, enabling broader applicability and impact. Future research can explore the multiclass classification of different glial cancers and other tumor types. Integration with Clinical Workflows: Developing user-friendly tools that integrate seamlessly with existing clinical workflows will be crucial for the practical adoption of these techniques. By continuing to advance these areas, the potential for machine learning to significantly enhance histopathological diagnostics will be fully realized, leading to better patient outcomes and more efficient healthcare delivery.

In conclusion, this thesis has made significant strides in applying deep learning to the classification of histopathology images, addressing critical challenges in both Lupus Nephritis and glioma subtype diagnosis. The proposed methods demonstrate the feasibility and effectiveness of automated image analysis, paving the way for future advancements in the field.

Bibliography

[1] Abhinav Agarwalla, Muhammad Shaban, and Nasir Rajpoot. Representation-aggregation networks for segmentation of multi-gigapixel histology images. 07 2017.

[2] Derek C Allen and R Iain Cameron. *Histopathology specimens: clinical, pathological and laboratory aspects*. Springer, 2017.

[3] Nicola Altini, Giacomo Donato Cascarano, Antonio Brunetti, Francescomaria Marino, Maria Teresa Rocchetti, Silvia Matino, Umberto Venere, Michele Rossini, Francesco Pesce, Loreto Gesualdo, and Vitoantonio Bevilacqua. Semantic segmentation framework for glomeruli detection and classification in kidney histological sections. *Electronics*, 9(3), 2020.

[4] Madhuri R Ankle and Priya S Joshi. A study to evaluate the efficacy of xylene-free hematoxylin and eosin staining procedure as compared to the conventional hematoxylin and eosin staining: An experimental study. *Journal of oral and maxillofacial pathology: JOMFP*, 15(2):161, 2011.

[5] Aditya Bagari, Ashish Kumar, Avinash Kori, Mahendra Khened, and Ganapathy Krishnamurthi. *A Combined Radio-Histological Approach for Classification of Low-Grade Gliomas: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I*, pages 416–427. 01 2019.

[6] Gloria Bueno, M. Milagro Fernandez-Carrobles, Lucia Gonzalez-Lopez, and Oscar Deniz. Glomerulosclerosis identification in whole slide images using semantic segmentation. *Computer Methods and Programs in Biomedicine*, 184:105273, 2020.

[7] Gloria Bueno, Lucia Gonzalez-Lopez, Marcial García-Rojo, and Arvydas Laurinavicius. Data for glomeruli characterization in histopathological images, mendeley data, v3.

[8] Gabriele Campanella, Matthew Hanna, Luke Geneslaw, Allen Miraflor, Vitor Silva, Klaus Busam, Edi Brogi, Victor Reuter, David Klimstra, and Thomas Fuchs. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine*, 25:1, 08 2019.

[9] P. A. Cicalese, A. Mobiny, Z. Shahmoradi, X. Yi, C. Mohan, and H. Van Nguyen. Kidney level lupus nephritis classification using uncertainty guided bayesian convolutional neural networks. *IEEE Journal of Biomedical and Health Informatics*, 25(2):315–324, 2021.

[10] Pierre Courtiol, Eric W Tramel, Marc Sanselme, and Gilles Wainrib. Classification and disease localization in histopathology using only global labels: A weakly-supervised approach. *arXiv preprint arXiv:1802.02212*, 2018.

[11] Angel Cruz-Roa, Hannah Gilmore, Ajay Basavanhally, Michael Feldman, Shridar Ganesan, Natalie Shih, John Tomaszewski, Fabio González, and Anant Madabhushi. Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent. *Scientific Reports*, 7:46450, 04 2017.

[12] Thibaut Durand, Taylor Mordan, Nicolas Thome, and Matthieu Cord. Wildcat: Weakly supervised learning of deep convnets for image classification, pointwise localization and segmentation. pages 5957–5966, 07 2017.

[13] Thibaut Durand, Nicolas Thome, and Matthieu Cord. Weldon: Weakly supervised learning of deep convolutional neural networks. pages 4743–4752, 06 2016.

[14] Babak Ehteshami Bejnordi, Guido Zuidhof, Maschenka Balkenhol, Meyke Hermsen, Peter Bult, Bram Ginneken, Nico Karssemeijer, Geert Litjens, and Jeroen van der Laak. Context-aware

stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images. *Journal of Medical Imaging*, 4, 05 2017.

[15] Jaime Gallego, Anibal Pedraza, Samuel Lopez, Georg Steiner, Lucia Gonzalez, Arvydas Laurinavicius, and Gloria Bueno. Glomerulus classification and detection based on convolutional neural networks. *Journal of Imaging*, 4(1), 2018.

[16] Brandon Ginley, Kuang-Yu Jen, Avi Rosenberg, Giovanni Maria Rossi, Sanjay Jain, and Pinaki Sarder. Fully automated classification of glomerular lesions in lupus nephritis. In John E. Tomaszewski and Aaron D. Ward, editors, *Medical Imaging 2020: Digital Pathology*, volume 11320, pages 231 – 237. International Society for Optics and Photonics, SPIE, 2020.

[17] Brandon Ginley, Brendon Lutnick, Kuang-Yu Jen, Agnes B. Fogo, Sanjay Jain, Avi Rosenberg, Vighnesh Walavalkar, Gregory Wilding, John E. Tomaszewski, Rabi Yacoub, Giovanni Maria Rossi, and Pinaki Sarder. Computational segmentation and classification of diabetic glomerulosclerosis. *Journal of the American Society of Nephrology*, 30(10):1953–1967, 2019.

[18] R. Girshick. Fast r-cnn. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1440–1448, 2015.

[19] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Region-based convolutional networks for accurate object detection and segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1):142–158, 2016.

[20] Darshana Govind, Brandon Ginley, Brendon Lutnick, John E. Tomaszewski, and Pinaki Sarder. Glomerular detection and segmentation from multimodal microscopy images using a Butterworth band-pass filter. In John E. Tomaszewski and Metin N. Gurcan, editors, *Medical Imaging 2018: Digital Pathology*, volume 10581, pages 297 – 303. International Society for Optics and Photonics, SPIE, 2018.

[21] David A Gutman, Jake Cobb, Dhananjaya Somanna, Yuna Park, Fusheng Wang, Tahsin Kurc, Joel H Saltz, Daniel J Brat, Lee AD Cooper, and Jun Kong. Cancer digital slide archive: an informatics resource to support integrated in silico analysis of tcga pathology data. *Journal of the American Medical Informatics Association*, 20(6):1091–1098, 2013.

[22] Zhongyi Han, Benzheng Wei, Yuanjie Zheng, Yilong Yin, Kejian Li, and Shuo Li. Breast cancer multi-classification from histopathological images with structured deep learning model. *Scientific Reports*, 7, 06 2017.

[23] Le Hou, Dimitris Samaras, Tahsin Kurc, Yi Gao, James Davis, and Joel Saltz. Patch-based convolutional neural network for whole slide tissue image classification. volume 2016, pages 2424–2433, 06 2016.

[24] Zhipeng Jia, Huang Xingyi, Eric Chang, and Yan Xu. Constrained deep weak supervision for histopathology image segmentation. *IEEE Transactions on Medical Imaging*, PP, 01 2017.

[25] Fahdi Kanavati, Gouji Toyokawa, Seiya Momosaki, Michael Rambeau, Yuka Kozuma, Fumihiro Shoji, Koji Yamazaki, Sadanori Takeo, Osamu Iizuka, and Masayuki Tsuneki. Weakly-supervised learning for lung carcinoma classification using deep learning. *Scientific Reports*, 10, 06 2020.

[26] Shruti Kannan, Laura A. Morgan, Benjamin Liang, McKenzie G. Cheung, Christopher Q. Lin, Dan Mun, Ralph G. Nader, Mostafa E. Belghasem, Joel M. Henderson, Jean M. Francis, Vipul C. Chitalia, and Vijaya B. Kolachalama. Segmentation of glomeruli within trichrome images using deep learning. *Kidney International Reports*, 4(7):955–962, 2019.

[27] Pegah Khosravi, Ehsan Kazemi, Marcin Imielinski, Olivier Elemento, and Iman Hajirasouliha. Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images. *EBioMedicine*, 27, 12 2017.

[28] Hei Law and Jia Deng. Cornernet: Detecting objects as paired keypoints. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[29] JI Liang, Jim Piper, and Jing-Yan Tang. Erosion and dilation of binary images by arbitrary structuring elements using interval coding. *Pattern Recognition Letters*, 9(3):201–209, 1989.

[30] Huangjing Lin, Hao Chen, Qi Dou, Liansheng Wang, Jing Qin, and Pheng-Ann Heng. Scannet: A fast and dense scanning framework for metastatic breast cancer detection from whole-slide images. 03 2018.

[31] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[32] Geert Litjens, Clara Sánchez, Nadya Timofeeva, Meyke Hermsen, Iris Nagtegaal, Iringo Kovacs, Christina Hulsbergen-van de Kaa, Peter Bult, Bram Ginneken, and Jeroen van der Laak. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific Reports*, 6:26286, 05 2016.

[33] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. Ssd: Single shot multibox detector. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 21–37, Cham, 2016. Springer International Publishing.

[34] Yun Liu, Krishna Gadepalli, Mohammad Norouzi, George E. Dahl, T. Kohlberger, Aleksey Boyko, Subhashini Venugopalan, Aleksei Timofeev, Phil Q. Nelson, G. Corrado, J. Hipp, Lily H. Peng, and Martin C. Stumpe. Detecting cancer metastases on gigapixel pathology images. *ArXiv*, abs/1703.02442, 2017.

[35] Yun Liu, Krishna Gadepalli, Mohammad Norouzi, George E Dahl, Timo Kohlberger, Aleksey Boyko, Subhashini Venugopalan, Aleksei Timofeev, Philip Q Nelson, Greg S Corrado, et al. Detecting cancer metastases on gigapixel pathology images. *arXiv preprint arXiv:1703.02442*, 2017.

[36] M. Lu, Drew F. K. Williamson, Tiffany Y Chen, Richard J. Chen, Matteo Barbieri, and Faisal Mahmood. Data efficient and weakly supervised computational pathology on whole slide images. *Nature biomedical engineering*, 2021.

[37] Alexandre Momeni, Marc Thibault, and Olivier Gevaert. Dropout-enabled ensemble learning for multi-scale biomedical data. In *International MICCAI Brainlesion Workshop*, pages 407–415. Springer, 2018.

[38] Talha Qaiser, Korsuk Sirinukunwattana, Kazuaki Nakane, Yee Tsang, David Epstein, and Nasir Rajpoot. Persistent homology for fast tumor segmentation in whole slide histology images. *Procedia Computer Science*, 90:119–124, 12 2016.

[39] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[40] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement, 2018.

[41] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

[42] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv 1409.1556*, 09 2014.

[43] Mingxing Tan, Ruoming Pang, and Quoc V. Le. Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[44] Abhishek Vahadane, Tingying Peng, Amit Sethi, Shadi Albarqouni, Lichao Wang, Maximilian Baust, Katja Steiger, Anna Melissa Schlitter, Irene Esposito, and Nassir Navab. Structure-

preserving color normalization and sparse stain separation for histological images. *IEEE Transactions on Medical Imaging*, 35(8):1962–1971, 2016.

[45] Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad, and Andrew Beck. Deep learning for identifying metastatic breast cancer. 06 2016.

[46] Xi Wang, Hao Chen, Caixia Gan, Huangjing Lin, Qi Dou, Efstratios Tsougenis, Qitao Huang, Muyan Cai, and Pheng-Ann Heng. Weakly supervised deep learning for whole slide lung cancer image analysis. *IEEE Transactions on Cybernetics*, PP:1–13, 09 2019.

[47] Jan J. Weening, Vivette D. D'agati, Melvin M. Schwartz, Surya V. Seshan, Charles E. Alpers, Gerald B. Appel, James E. Balow, J. A. N. A. Bruijn, Terence Cook, Franco Ferrario, Agnes B. Fogo, Ellen M. Ginzler, L. E. E. Hebert, Gary Hill, Prue Hill, J. Charles Jennette, Norella C. Kong, Philippe Lesavre, Michael Lockshin, Lai-Meng Looi, Hirofumi Makino, Luiz A. Moura, Michio Nagata, ON Behalf of the International Society of Nephrology, and Renal Pathology Society Working Group on the Classification Oflupus Nephritis. The classification of glomerulonephritis in systemic lupus erythematosus revisited. *Kidney International*, 65(2):521–530, Feb 2004.

[48] Jason Wei, Laura Tafe, Yevgeniy Linnik, Louis Vaickus, Naofumi Tomita, and Saeed Hassanpour. Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks. *Scientific Reports*, 9, 03 2019.

[49] Yan Xu, Jun-Yan Zhu, E. Chang, M. Lai, and Zhuowen Tu. Weakly supervised histopathology cancer image segmentation and classification. *Medical image analysis*, 18 3:591–604, 2014.

[50] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? *arXiv preprint arXiv:1411.1792*, 2014.

[51] Zhi-Hua Zhou. A brief introduction to weakly supervised learning. *National Science Review*, 5, 08 2017.

[52] Ningbo Zhu, Gang Wang, Gaobo Yang, and Weiming Dai. A fast 2d otsu thresholding algorithm based on improved histogram. In *2009 Chinese conference on pattern recognition*, pages 1–5. IEEE, 2009.