WEST UNIVERSITY OF TIMIŞOARA
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
MASTER STUDY PROGRAM: Artificial Intelligence and Distributed Computing

MASTER THESIS

SUPERVISOR:
Lect. Dr. Sancira Monica

GRADUATE:
Cincu Andrada Maria Alexandra

TIMIŞOARA
2023

WEST UNIVERSITY OF TIMIŞOARA
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
MASTER STUDY PROGRAM: Artificial Intelligence and Distributed Computing

Deep Learning for Histopathology Image Analysis

SUPERVISOR:
Lect. Dr. Sancira Monica

STUDENT:
Cincu Andrada Maria Alexandra

TIMIŞOARA
2023

Abstract

A critical branch of medical diagnosis called histopathology depends heavily on the accurate interpretation of tissue microscopic images. This thesis offers a concentrated investigation into the use of deep learning methods for histopathology image analysis. It is explored the efficiency of convolutional neural networks (CNNs) in automating tasks like tumor detection, classification, and general pathology assessment, highlighting developments in deep learning.

This thesis looks at the particular challenges that come with histopathological imaging, such as the necessity for accurate localization of abnormalities, staining fluctuations, and complicated tissue structures. Presented is a thorough analysis of current deep learning models that have been modified for histopathology, focusing their architectures, training processes, and performance metrics. Particular focus is placed on the transferability and generalizability of these models across different datasets and pathology types.

Furthermore, the thesis investigates the integration of multimodal data to improve overall diagnostic accuracy, such as merging histopathological pictures with clinical information or molecular data. Current deep learning models in histopathology are critically analyzed for limitations and potential biases, providing insights into future research possibilities.

In conclusion, this thesis contributes to a better knowledge of the current status of deep learning for histopathological image interpretation, concentrating on both accomplishments and challenges in this rapidly expanding field. The findings are intended to enlighten researchers and practitioners about the existing environment and possible breakthroughs in using deep learning to make more accurate and efficient histopathological diagnoses.

Contents

Chapter 1

Introduction

The background of histology in medical diagnosis is established in this chapter, pointing out the limitations that traditional methods of analysis encounter. Given the rising complexity of histological data and the established efficiency of deep learning in medical imaging, the justification for using deep learning in histopathology is investigated. The context of the research is provided, emphasizing the need of incorporating deep learning techniques into histopathology analysis and its relevance to modern medical imaging practices.

1.1 Motivation

Histopathology, an important component of medical diagnosis, is based on the systematic study of tissue samples to detect and characterize disease. Traditional histopathological analysis, while successful, is labor-intensive, time-consuming, and sensitive to observer variability. Furthermore, the growing volume and complexity of histopathology data necessitates a more efficient and precise technique.
This thesis purpose is to offer a concentrated investigation into the use of deep learning methods for histopathology image analysis. It is explored the efficiency of convolutional neural networks (CNNs) in automating tasks like tumor detection, classification, and general pathology assessment, highlighting developments in deep learning.
The goal of this thesis is to investigate how deep learning might act as a catalyst for innovation in histopathology, opening the way for a future in which advanced technology complements and enhances pathology capabilities, resulting in more accurate and faster diagnoses.

1.2 Context of research

This thesis studies the usefulness of convolutional neural networks in automating histopathological image analysis tasks.
The study analyzes architectures, training methodologies, and performance across different datasets and disease types of existing deep learning models for histopathology. Furthermore, it investigates multimodal data integration, merging histopathological pictures with clinical or molecular information to improve diagnostic accuracy. The research intends to provide insights for improvement and recommendations for future efforts by critically analyzing the limitations and potential biases of present models.

Chapter 2

State of the art in the field

2.1 Addressed Problems

2.1.1 Introduction of the related field

Histopathology Image Analysis is a crucial aspect of modern medical diagnostics, playing a pivotal role in the identification and characterization of various diseases, especially cancers. Histopathological examinations involve the microscopic examination of tissue samples to detect abnormalities, assess disease progression, and guide treatment decisions. The analysis of histopathological images traditionally relies on the expertise of pathologists, making it a time-consuming and subjective process.
The addressed problem in the related work is the classification of breast cancer histopathology images. Specifically, the challenge involves distinguishing between different classes such as benign and malignant cases. The goal is to develop effective diagnostic systems to aid pathologists in improving the accuracy and efficiency of the diagnostic process. Traditional approaches, including handcrafted feature-based methods and machine learning classifiers, have been used, but they often rely on manual feature selection and may not fully capture complex and high-level features. In response to these limitations, recent studies have turned to deep learning models, particularly convolutional neural networks (CNNs), to automatically learn and extract intricate features from histopathological images. The work mentioned various studies that employed deep learning, especially CNNs, for breast cancer classification, with some using the BreakHis dataset. The authors of the current paper contribute to this area by introducing their dataset and exploring the classification of breast cancer histology images using deep learning models, specifically pre-trained VGG16 and VGG19 architectures.[2]
Accurate and efficient analysis of histopathology images is essential for timely and precise diagnosis. However, this process is often labor-intensive and subjective, relying on the expertise of pathologists to interpret complex visual patterns. As medical datasets grow larger and the demand for faster and more accurate diagnostics increases, there is a need for advanced computational tools to assist pathologists in image analysis[3].
In recent years, the advent of Deep Learning, particularly Convolutional Neural Networks (CNNs), has revolutionized histopathology image analysis. CNNs, a class of deep neural networks designed for image recognition tasks, have demonstrated remarkable capabilities in automated feature extraction and pattern recognition. Their application in medical image analysis, including histopathology, has shown great promise in enhancing the efficiency and accuracy of diagnosis[3].

Traditional histopathological analysis involves manual interpretation of complex images, a process prone to inter-observer variability and time constraints. Deep Learning techniques, and specifically CNNs, address these challenges by automatically learning hierarchical representations of features directly from the raw image data. CNNs excel at capturing intricate patterns and subtle details within histopathological images, making them invaluable tools for accurate and reproducible analysis. These advancements are particularly crucial in the context of breast cancer, the most prevalent cancer among women globally. Automated classification and subtype identification of breast cancer from histopathological images using CNNs offer the potential for faster, more consistent, and precise diagnoses[5].
The proposed solution [4] involves the development of an interpretable decision-support model, named ABN-DCN (Attention Branch Network with DarkCovidNet), for precise classification of benign and malignant tumors in breast tissue samples. The goal is to enhance the accuracy of classification while providing diagnostic interpretability through attention maps that highlight the regions of interest in the histopathological images.
It is also highlighted the challenge of accurately classifying specific soft-tissue sarcomas—clear cell sarcoma (CCS), alveolar rhabdomyosarcoma (aRMS), and embryonal rhabdomyosarcoma (eRMS)—based on histopathological images. The researchers intend to address this difficulty by developing a Convolutional Neural Network (CNN) model.This model is seen as a valuable tool to support pathologists in the histopathological classification of these rare pediatric sarcomas. The obstacles to overcome include a limited number of cases, potential misclassification issues, and the imperative for a reliable computational approach to facilitate the differential diagnosis of these malignancies[1].
This thesis aims to contribute to the ongoing efforts in integrating deep learning, specifically CNNs, into histopathology image analysis, with a focus on breast cancer diagnosis. By addressing the outlined goals through a rigorous methodology, the research seeks to advance the field, providing valuable insights for clinicians and researchers alike. Ultimately, the successful implementation of deep learning in histopathology stands to enhance the accuracy, speed, and reproducibility of medical diagnostics, thereby improving patient outcomes and healthcare efficiency.

2.1.2 Problem Statement

The research paper [3] addresses critical challenges in histopathology image analysis for endometrial cancer. It aims to accurately classify histological and molecular subtypes, including CNV-H, CNV-L, and MSI-high. Prediction of mutation status for key genes like TP53 and PTEN is explored, along with the correlation with specific subtypes.
Methods for visualizing deep learning model features at the whole-slide level are developed, including activation maps and two-dimensional tSNE plots. The study evaluates Panoptes' generalizability, clinical capability on diverse datasets, and performance against baseline models in predicting subtypes and mutation status.
Model optimization involves modifying the Panoptes architecture and exploring enhancements such as quantifying features. The paper emphasizes deploying AI-based models in clinical settings for efficient endometrial cancer subtype classification, highlighting the potential for rapid Hematoxylin and Eosin (H&E) slide analysis. Overall, the research contributes significantly to histopathology image analysis, showcasing

Panoptes' efficacy in predicting endometrial cancer characteristics.
The research paper [5] addresses several crucial challenges in the domain of histopathology image analysis, specifically focusing on breast cancer classification. One significant issue is the limited training data and imbalanced datasets commonly observed in histopathology datasets. The scarcity of samples and imbalances in class distribution can impede the training of deep learning models, highlighting the importance of representative datasets to ensure accurate classification across diverse tissue types. Another challenge addressed is the occurrence of overfitting and unbalanced class problems. Overfitting, where the model excessively learns from the training data, coupled with imbalanced classes, can hinder the model's generalization to new, unseen data. Addressing these issues is vital to ensure the robustness and generalizability of the model, contributing to reliable predictions on unfamiliar data. The paper employs data augmentation techniques as a solution, introducing random variations in intensity, rotation, flip, and translation to mitigate overfitting and imbalanced class challenges. Data augmentation aims to create a more diverse and unbiased training dataset, enhancing the model's ability to handle varied tissue appearances. The proposed hybrid CNN-LSTM model architecture introduces its own set of challenges. Combining convolutional neural networks (CNNs) with recurrent neural networks (LSTM) necessitates effective optimization of the model architecture and parameters. The hybrid model seeks to capture both spatial and temporal dependencies in the data, making optimization crucial for achieving high accuracy in the classification of different breast cancer subtypes. The primary objective of the research is to classify breast cancer histopathological images into benign and malignant subtypes, including various subcategories such as adenosis, fibroadenoma, and ductal carcinoma. Accurate classification of these subtypes is essential for informed treatment planning, prognosis, and overall patient management. Additionally, the paper addresses the challenge of handling different magnification factors within the dataset. With images captured at various resolutions (40×, 100×, 200×, and 400×), the model needs to adapt effectively to different magnification factors. This adaptability ensures the model's applicability to diverse imaging setups, enhancing its versatility for real-world histopathological analysis.
The research paper [2] addresses the need for an efficient and accurate classification system dedicated to breast cancer histopathology images. The challenges identified in the problem statement revolve around the intricacies and subjectivity inherent in manual histopathological analysis, a process known for its labor-intensive nature and susceptibility to variations in diagnostic decisions among pathologists.To address this, the research paper aims to develop an automatic diagnostic system to aid pathologists, minimizing differences in diagnostic decisions and enhancing accuracy. The creation of a specialized dataset, comprising whole slide images from breast cancer patients and curated with input from experienced pathologists, plays a pivotal role in achieving this goal. The primary task at hand is to classify histopathology images into non-carcinoma and carcinoma classes, with a specific focus on accurately identifying carcinoma images. This classification task aligns with the broader objective of improving diagnostic precision in breast cancer cases.Utilizing pre-trained VGG16 and VGG19 architectures as the foundation for deep learning models, the study seeks to overcome the limitations of manual analysis and elevate diagnostic accuracy through automation. In essence, the problem statement underscores the importance of leveraging advanced computational models to transform the breast cancer diagnostic process.

The research paper[4] focuses on addressing the challenge of accurately classifying breast cancer histopathological images, specifically in distinguishing between benign and malignant tumors. The complexity of analyzing intricate details within tissue samples underscores the need for a robust computational model to ensure precise diagnosis. In this context, the paper introduces the ABN-DCN model, a novel approach that integrates a convolutional neural network (CNN) with an attention branch. The primary goal is to not only enhance classification accuracy but also improve interpretability. The utilization of the BreaKHis dataset, derived from breast tissue slides and featuring images with varying magnification factors, provides a diverse and comprehensive set of data for model training and evaluation. The ABN-DCN model aims to achieve high accuracy in categorizing tumors while also generating attention maps. These attention maps serve the dual purpose of highlighting regions crucial for decision-making and enhancing the overall interpretability of the model's predictions.
The specific problem addressed in the research paper [1] is the histopathological classification of rare soft-tissue sarcomas, particularly clear cell sarcoma (CCS), alveolar rhabdomyosarcoma (aRMS), and embryonal rhabdomyosarcoma (eRMS). Soft-tissue sarcomas are challenging to classify accurately due to their rarity and the potential for misclassification by pathologists. The researchers aim to overcome these challenges by developing a Convolutional Neural Network (CNN) model. To train the CNN model, the researchers utilized a dataset consisting of histopathology slides from various sources, including cases of CCS, aRMS, and eRMS. The goal is to create a computational tool that can assist pathologists in the differential diagnosis of these subtypes. The research paper provides details about the training process, validation, and testing of the CNN model on both human and murine-origin histopathology images. The research paper also explores the CNN model in the presence of atypical features such as anaplasia and evaluates its performance on an external dataset of histopathology slides. Additionally, the challenges associated with variations in image quality, scanning protocols, and tissue preparation are highlighted, highlighting the need for standardized procedures in digital histopathology.

2.2 Methodologies

The research paper [5] a hybridConvolutional Neural Network-Long Short-Term Memory (CNN-LSTM) method for classifying histopathological breast cancer images, utilizing ResNet50 and InceptionV3 as the underlying CNN architectures. The input layer of each architecture takes images of size (299×299×3). In ResNet50, convolutional blocks with residual connections address the vanishing gradient problem, and max pooling layers facilitate downsampling. The architecture culminates in a fully connected layer with softmax activation for classification. InceptionV3 employs inception modules, incorporating parallel convolutional operations with different kernel sizes and auxiliary classifiers for mitigating the vanishing gradient problem. Average pooling is used for downsampling, and a fully connected layer with softmax activation concludes the architecture. The integration of LSTM in the hybrid model suggests an emphasis on capturing temporal dependencies or sequential patterns in histopathological images. This choice is particularly relevant in medical imaging, where understanding temporal aspects could be crucial for accurate diagnosis and classification. Data augmentation techniques, such as intensity variation, rotation, flipping, and translation,

are employed to diversify the training dataset, mitigate overfitting, and address class imbalance. These techniques contribute to the robustness of the model by exposing it to variations in orientation, position, and intensity. The study compares three optimizers, ultimately identifying Adam as the best optimizer for the task. Adam's efficiency and adaptive learning rate capabilities make it well-suited for optimizing the proposed CNN-LSTM model. The model performs subtype classification, categorizing breast cancer images into distinct subtypes, including adenosis, fibroadenoma, tubular adenoma, ductal carcinoma, lobular carcinoma, mucinous carcinoma, and papillary carcinoma. This class-specific analysis provides detailed insights into different cancer subtypes. The proposed method demonstrates good results compared to existing state-of-the-art models, achieving about 99% accuracy in binary classification (benign vs. malignant) and 92.50% accuracy for multi-class classification of benign and malignant cancer subtypes. The paper acknowledges certain limitations, such as the computational complexity of the model and the challenges associated with interpretability in deep learning-based diagnostics[5].

The Panoptes CNN architecture[3], tailored for histopathology image analysis, exhibits a sophisticated multi-resolution InceptionResnet-based design. Panoptes processes a set of three tiles, representing the same region on an H&E slide, rather than a single tile. Each set of tiles is amalgamated into a single matrix, treating it as a unified sample. Notably, the tiles within a set exhibit equivalent scanning resolutions of 2.5x, 5x, and 10x. The architecture comprises three InceptionResnet-based branches, each concurrently processing samples at a specific resolution. These branches operate independently until the third-to-last layer, where the outputs are concatenated. Following concatenation, a global average pooling layer and the final fully connected layer are applied. The multi-resolution design is a distinctive feature of Panoptes, enabling the simultaneous consideration of macro-tissue-level and cellular-level features. This approach captures features of various sizes on H&E slides, mirroring the reviewing strategy employed by human pathologists. Panoptes distinguishes itself by preserving spatial information through the use of multi-resolution tile sets as input, while maintaining a single output and loss function. This differs from models that combine decisions from separate models trained on tiles of distinct resolutions. An attempt is made to integrate clinical features, such as patients' age and body mass index (BMI), through the addition of a 1x3x1 feature pooling convolutional layer before global average pooling. Additionally, a fourth branch is dedicated to processing these clinical features. The exploration of four different Panoptes architectures, including variations with and without the clinical feature branch, is notable. Different types of InceptionResnet and Inception models were trained on single-resolution 10x3 tiles, and the selection of architecture was based on the area under the receiver operating characteristic curve (AUROC) of the test sets. In terms of performance, Panoptes outperformed baseline models (Inception and InceptionResnet) in various prediction tasks related to histopathology images. The multi-resolution architecture, in particular, demonstrated superior performance in analyzing endometrial cancer H&E slides for tasks such as histological subtyping, predicting molecular subtypes, and identifying gene mutations. For visualization and interpretation, activation maps were extracted to evaluate features learned by the models for each task. Furthermore, 2D tSNE plots were generated to illustrate the relationships between predicted groups based on the learned features. In conclusion, Panoptes showcases a multi-resolution approach, incorporates clinical features, and exhibits a good performance in histopathological image analysis tasks[3].

The paper's methodologies [2] involve the application of deep learning models, specifically employing VGG16 and VGG19 architectures for the classification of breast cancer histopathology images. To facilitate this, the authors created a private dataset consisting of whole slide images (WSI) from breast cancer patients, covering both non-carcinoma and carcinoma classes. Image patches were subsequently extracted from these WSI images to serve as the foundational elements for the classification task. In terms of deep learning architectures, the research paper utilized two well-established pre-trained models—VGG16 and VGG19. Four distinct models were trained, including fully-trained versions of VGG16 and VGG19, as well as fine-tuned versions where further training was performed on the specific dataset. The application of a 5-fold cross-validation ensured robust evaluation across different subsets of the dataset. An ensemble strategy was employed, combining the predicted probabilities from both fully-trained and fine-tuned models of VGG16 and VGG19. This ensemble approach involved calculating the average of predicted probabilities from individual models. Performance evaluation metrics, such as precision, recall, F1 score, and test accuracy, were computed for each model and fold, utilizing confusion matrices to assess classification results. Furthermore, the paper compared the effectiveness of the proposed ensemble approach with various state-of-the-art studies that focused on the classification of breast cancer histopathology images, particularly those utilizing the BreakHis dataset. The methodologies are thoroughly discussed, comparing results with recent studies and highlighting the competitive performance achieved by the proposed ensemble deep learning model. Overall, the methodologies include dataset preparation, the utilization of deep learning architectures, cross-validation techniques, ensemble strategies, and thorough performance evaluation to effectively address the classification problem in breast cancer histopathology images.

The methodologies applied in developing the ABN-DCN model [4] for breast cancer histopathological image classification involve several key components. The research paper employs the BreaKHis dataset from P & D Lab, Brazil, consisting of publicly available breast histopathological images from 82 patients. This dataset includes 7,909 images categorized into benign and malignant samples, spanning different magnification levels (40×, 100×, 200×, 400×) and various tumor classes. The Attention Branch Network (ABN) is designed to enhance CNN performance through an attention mechanism, employing a Class Activation Mapping (CAM) approach with a K×1×1 convolution layer, global average pooling (GAP), and a fully connected layer. Attention maps are generated, and the attention branch outputs probability values using GAP and the softmax function. Training involves separate training for different magnification factors (40×, 100×, 200×, 400×). Data augmentation techniques, including rotations and horizontal flips, are applied to augment the training set and prevent overfitting. Stochastic Gradient Descent serves as the optimizer with weight decay and momentum. Performance evaluation includes standard metrics such as accuracy, precision, recall, and F1 score. Evaluation is conducted for each magnification level, and results are compared with baseline models and other low-footprint CNN models. The ABN-DCN model is compared with other low-footprint CNN models, including baseline DCN, VGG16, EfficientNetB0, and MobileNet, with comparative performance metrics presented to emphasize the model's superiority. The research paper acknowledges limitations, including potential misclassifications, and suggests future work, such as fine-tuning the model and exploring its application on different datasets. In summary, the methodologies encompass a comprehensive approach, incorporating dataset description, model

architecture design, attention mechanism integration, training strategy, performance evaluation, interpretability analysis, and comparative assessments with other models.
The research paper [1] applied Convolutional Neural Network (CNN) models to histopathologically classify soft-tissue sarcomas, employing diverse methodologies. Dataset collection involved curating slides from various sources, encompassing clear cell sarcoma (CCS), alveolar rhabdomyosarcoma (aRMS), and embryonal rhabdomyosarcoma (eRMS). Preprocessing included color normalization, segmentation into 512 $\times$ 512-pixel tiles, and removal of extraneous blank tiles for standardized image characteristics. To address training set imbalances, oversampling techniques were applied, and the dataset was split into 70% training, 15% validation, and 15% testing. The CNN model construction utilized the DeepPATH software suite on the Inception v3 platform, training models for human- and murine-origin tissues. Evaluation metrics, such as ROC analysis, AUC scores, and MCC, assessed performance. External validation tested the trained model on a separate slide cohort, and robustness testing included diverse histopathology slides. Software tools encompassed Python modules, the DeepPATH suite, Pillow for image manipulation, and Inception v3. Statistical analyses used Python and GraphPad Prism. Acknowledging limitations, the study aimed to develop and assess CNN models for soft-tissue sarcoma diagnosis based on histopathological images.

2.3 Classification Approaches

In the research paper [2], VGG16 and VGG19 architectures were employed to classify breast cancer histopathology images, comparing fully-trained and fine-tuned models. The fine-tuned VGG16 model demonstrated improved sensitivity (95.68%), accuracy (91.67%), and average F1 score (91.63%) compared to its fully-trained counterpart. Similarly, the fine-tuned VGG19 model outperformed its fully-trained version in carcinoma sensitivity (95.68%), overall accuracy (91.67%), and average F1 score (91.63%). The comprehensive methodology involved training four models: fully-trained VGG16, fine-tuned VGG16, fully-trained VGG19, and fine-tuned VGG19. Additionally, an ensemble strategy combining fine-tuned VGG16 and VGG19 models aimed to boost classification performance. Evaluation metrics concentrated on non-carcinoma and carcinoma classification, with an emphasis on high sensitivity for carcinoma.The comparison between VGG16 and VGG19 highlighted specific metric values favoring the fine-tuned VGG16 model, emphasizing the efficacy of fine-tuning for superior performance in certain metrics within breast cancer histopathology classification.
The Panoptes [3] employs a deep learning approach to categorize histopathology images related to endometrial cancer, involving tasks such as predicting histological and molecular subtypes and the mutation status of specific genes. The primary metric for evaluating the model's performance is the AUROC, which measure its ability to distinguish between positive and negative instances across different threshold values. Additional metrics, including precision, recall, sensitivity, specificity, and F1 scores, offer a comprehensive evaluation of the model's classification accuracy.
The assessment is carried out at both per-patient and per-tile levels. Per-patient metrics are determined by calculating the mean of all tile metrics belonging to the same patient. Statistical significance is established through one-tail Wilcoxon tests, specifically examining prediction scores between positively and negatively labeled tiles.

This analysis illustrates the model's proficiency in distinguishing between different tiles within the test sets.
To compare the performance with baseline models, the paper utilizes bootstrap sampling and subsequent t tests, providing a robust evaluation of Panoptes against baseline architectures. The reported AUROC scores for specific tasks, such as histology, CNV-H from endometrioid, and CNV-H, serve as quantitative indicators of the model's ability to distinguish between positive and negative instances in these tasks.
Illustrative examples of per-patient AUROC scores from the paper [3] include:

- Histology: Panoptes2: 0.969 (95% CI: 0.905–1)
- CNV-H from endometrioid: Panoptes1: 0.958 (95% CI: 0.886–1)
- CNV-H: Panoptes4: 0.934 (95% CI: 0.851–1)

These scores provide a quantifiable measure of the model's performance in different classification tasks related to histological and molecular subtypes. The inclusion of confidence intervals (CIs) enhances the statistical robustness of the reported performance metrics, indicating the range of values within which the true performance is likely to fall. Overall, the combination of numerical metrics, statistical methods, and task-specific AUROC scores ensures a thorough and reliable evaluation of Panoptes' proficiency in identifying histopathological features in endometrial cancer images[3].
The research paper [5] adopts a classification approach for histopathology images using a hybrid CNN-LSTM model. The primary task is the classification of breast cancer images into distinct subtypes, differentiating between benign and malignant cases. The dataset used for this classification is the Breast Cancer Histopathological Image Classification (BreakHis) dataset, which comprises 7,909 microscopic images of breast tumor tissue collected from 82 patients. The images are classified into several subtypes, including adenosis, fibroadenoma, tubular adenoma, ductal carcinoma, lobular carcinoma, mucinous carcinoma, and papillary carcinoma.
To evaluate the performance of the classification model, the paper employs various metrics, including accuracy, precision, recall, top k-categorical accuracy, precision-recall curves, and area under the curve (AUC). For binary classification (benign vs. malignant), accuracy, precision, and recall are reported for different magnification factors (40×, 100×, 200×, and 400×). For multi-class classification of benign and malignant subtypes, accuracy, precision, recall, and top k-categorical accuracy are presented. Precision-recall curves and AUC are also used to assess the model's performance in distinguishing between different classes.
The study provides a detailed analysis of the classification results, highlighting the accuracy achieved for each magnification factor and subtype. The proposed hybrid CNN-LSTM model outperforms existing state-of-the-art models in terms of accuracy, demonstrating its effectiveness in histopathological image classification. The use of various metrics allows for a comprehensive evaluation of the model's performance, considering both binary and multi-class classification scenarios[5].
The classification approach in the proposed ABN-DCN model [4] involves a combination of convolutional neural network (CNN) architecture. he model utilizes a variant of DarkNet19, referred to as the DCN model, as the feature extractor. This model is a convolutional neural network designed to extract relevant features from

breast histopathological images. It comprises 17 convolutional layers with Batch Normalization and Leaky ReLU operations, adapted from its original purpose of detecting COVID-19 in chest X-ray images. The model is trained using a magnification-dependent BreaKHis dataset, which includes images at 40×, 100×, 200×, and 400× magnifications. Stochastic Gradient Descent is employed as the optimizer, with weight decay and momentum. The training is performed over 200 epochs, and the dataset is split into independent training and validation sets. The performance of the classification model is evaluated using standard metrics, including accuracy, precision, recall, and F1 score. These metrics provide a comprehensive assessment of the model's ability to correctly classify benign and malignant tumors in breast histopathological images.
The primary classification approach used in the research paper [1] is Convolutional Neural Network (CNN) modeling. CNNs are a type of deep learning model specifically designed for image analysis tasks. The researchers leveraged CNNs to classify histopathology images of pediatric soft-tissue sarcomas, focusing on clear cell sarcoma (CCS), alveolar rhabdomyosarcoma (aRMS), and embryonal rhabdomyosarcoma (eRMS). mages underwent preprocessing steps, including color normalization and segmentation into non-overlapping tiles of size 512 × 512 pixels, to standardize formats. Multiple CNN models were trained using 70% of the image tiles for each disease type. The model with the highest validation Area Under the Curve (AUC) scores was selected for downstream testing. The primary focus of the classification approaches was on the use of deep learning, specifically CNNs, to effectively classify and differentiate between different subtypes of soft-tissue sarcomas based on histopathology images.

2.4 Datasets

The research paper [2] utilized a dataset created by the researchers, consisting of 544 whole slide images (WSI) obtained from 80 patients with breast cancer. The images were collected from the pathology department of Colsanitas Colombia University in Bogotá, Colombia. The tumor tissue fragments were fixed in formalin, embedded in paraffin, and cut into 4 mm sections, which were stained with hematoxylin and eosin (H & E). The dataset includes 845 histopathology images, with 437 belonging to the carcinoma class (malignant tumors) and 408 to the non-carcinoma class (normal tissues and benign images of non-tumor glandular tissues). The images were captured at 200× magnification, resulting in dimensions of 1278 × 760 pixels. The researchers compared their dataset with the commonly used BreakHis dataset, which is frequently employed in studies related to breast cancer histopathology image classification. The BreakHis dataset typically includes images categorized into four classes: benign, atypical ductal hyperplasia (ADH), ductal carcinoma in situ (DCIS), and invasive ductal carcinoma (IDC).
The research paper [1] utilizes a dataset comprising histopathology slides of pediatric soft-tissue sarcoma, focusing on clear cell sarcoma (CCS), alveolar rhabdomyosarcoma (aRMS), and embryonal rhabdomyosarcoma (eRMS). The dataset, centrally-reviewed and totaling 424 sarcoma histopathology slides, serves as the primary source for model training. In addition to the main dataset, external datasets are incorporated for testing and validation. An external set of histopathology slides, independent of the main dataset, is used to assess the generalization capabilities of the Convolutional Neural Network (CNN) models. This external set includes cases of CCS, aRMS, and eRMS.

The research also involves a murine dataset derived from genetically engineered mouse models of aRMS and eRMS, alongside normal murine skeletal muscle tissue. The inclusion of these diverse datasets, representing both human and murine origins, contributes to a comprehensive evaluation of the CNN models in terms of their effectiveness in classifying and subclassifying histopathology images related to pediatric soft-tissue sarcoma.

2.4.1 Clinical Proteomic Tumor Analysis Consortium & The Cancer Genome Atlas

The research paper [3] relies on datasets sourced from two significant entities in cancer research, namely CPTAC (Clinical Proteomic Tumor Analysis Consortium) and TCGA (The Cancer Genome Atlas). The dataset encompasses a total of 456 patients from these cohorts, with each patient likely contributing clinical and histopathological information related to endometrial cancers. The collaborative nature of CPTAC and TCGA introduces heterogeneity, encompassing diverse patient demographics, tumor characteristics, and molecular profiles.
Regarding the slide processing workflow, a per-patient separation into training, validation, and test sets is implemented, ensuring that the model is trained on diverse data and evaluated on unseen patient cases. The slides undergo a tile extraction process, being cut into 299x299-pixel tiles while excluding background and contaminants. This tile-based approach allows the model to capture local features and patterns within the histopathological images.
The tile sets and resolutions strategy involves cutting slides into paired tile sets at different resolutions (2.5x, 5x, and 10x equivalent resolution of the same region). The research paper multi-resolution design enables it to consider both macro-tissue-level features and minute cellular-level features concurrently. Notably, the preservation of spatial information is emphasized as a distinctive feature of Panoptes[3] compared to models trained on tiles of three resolutions separately, achieved by using multi-resolution tile sets as input with a single output and loss function.
In terms of clinical feature integration, the paper introduces a fourth branch in Panoptes that processes clinical features, including patients' age and BMI. This suggests an effort to incorporate additional relevant information into the model for improved performance.
Highlighting the importance of representative datasets, endometrial cancers' diverse histopathological features are acknowledged, and the inclusion of patients from CPTAC and TCGA ensures a diverse representation of these features. The collaboration's significance lies in the molecular diversity it brings, with CPTAC's focus on proteomic analysis and TCGA's comprehensive genomic characterization likely providing a rich molecular landscape. The datasets' separation into training, validation, and test sets, combined with the multi-resolution tile sets, contributes to training robust models capable of generalizing well to unseen patient cases. In conclusion, the research paper carefully curates datasets from reputable sources, ensuring diversity in histopathological and molecular characteristics, with preprocessing steps tailored to enhance the model's ability to capture relevant features from histopathological images[3].

2.4.2 Breast Cancer Histopathological Image Classification

The research paper [5] utilizes the Breast Cancer Histopathological Image Classification (BreakHis) dataset, which is a publicly available dataset consisting of microscopic images of breast tumor tissue. The dataset was collected from 82 patients and includes images at different magnification factors (40×, 100×, 200×, and 400×). The images are in RGB format with a resolution of 700×460 pixels and an 8-bit depth in each channel, stored in PNG format.
Dataset Characteristics:

- Size: The BreakHis dataset contains a total of 7,909 microscopic images, with 2,480 labeled as benign and 5,429 labeled as malignant. The images cover various magnification factors, providing a diverse set of samples.
- Diversity: The dataset encompasses diverse histopathological images representing different subtypes of breast cancer. It includes distinct subtypes such as adenosis, fibroadenoma, tubular adenoma, phyllodes tumor, ductal carcinoma, lobular carcinoma, mucinous carcinoma, and papillary carcinoma.
- Preprocessing Steps: Data augmentation techniques are employed to address overfitting and unbalanced class issues during training. Augmentation includes random combinations of intensity variation, rotation, flip with horizontal and vertical direction, and translation. The goal is to create a more robust training dataset that captures variations in image appearance.

The BreakHis dataset undergoes thorough preprocessing steps in the paper to enhance the performance and robustness of the proposed model. Data augmentation techniques are employed during training, involving random intensity variation, rotation, flipping (both horizontally and vertically), and translation. Keras's ImageDataGenerator facilitates on-the-fly augmentation. Notably, the dataset's class imbalance, particularly the abundance of ductal carcinoma cases, is addressed through oversampling. This involves applying data augmentation to balance the number of images for each class. These preprocessing measures collectively aim to mitigate challenges related to overfitting, imbalanced class distributions, and ensure the model's ability to generalize to diverse and unseen histopathological images.
Representative datasets play a crucial role in training and testing machine learning models, especially in the medical domain. The BreakHis dataset [5], being a comprehensive collection of breast cancer histopathological images, allows the model to learn diverse features and patterns associated with different subtypes of breast cancer.
The dataset is used, also, in the context of the proposed ABN-DCN model[4] is the BreaKHis dataset, featuring histopathological images of breast tissue obtained from partial mastectomy or excisional biopsy specimens. The dataset's magnification-dependent nature encompasses images at different levels: 40×, 100×, 200×, and 400×. As part of the model's preparation, the images undergo preprocessing, likely involving tasks such as resizing to a consistent resolution of 256×256 pixels for compatibility with the model architecture. The dataset is strategically split into independent training and validation sets, adhering to a standard 70%–30% train-test split ratio common in machine learning practices.Organized based on varying magnification factors, the dataset facilitates diverse training and testing scenarios, allowing the model to examine tissue structures at different resolutions. The training strategy involves experimental

verification using a Tesla K-80 GPU, employing Stochastic Gradient Descent with a momentum of 0.9, weight decay, and a learning rate of 0.1. The training process spans 200 epochs, utilizing a mini-batch size of 32. Evaluation metrics such as accuracy, precision, recall, and F1 score are applied to assess the model's performance at each magnification level.

2.5 Main Results from the Research Papers

2.5.1 Experimental Setup

The experimental setup in the research paper [3] involves training, validation, and testing procedures for evaluating the performance of the proposed Panoptes model. The authors perform training and validation using convolutional neural networks (CNNs) on histopathology image datasets. The datasets are split into training, validation, and testing sets, and specific tiles from hematoxylin and eosin (H&E) stained slides are used for model training and evaluation.
The experimental setup in the paper [5] involves utilizing the BreakHis dataset for training, validation, and testing. The dataset, comprising 7,909 microscopic images of breast tumor tissue from 82 patients, is divided into different magnification factors (40×, 100×, 200×, and 400×). The training process involves employing a hybrid CNN-LSTM model with two main modules: a CNN and an independent RNN module. The CNN, with a (299×299×3) input shape, is pretrained using transfer learning from models like InceptionResNetV2 and ResNet50, leveraging ImageNet weights. The final convolutional layer of the CNN captures bottleneck features of size (batch size, 2048). Simultaneously, the RNN module consists of two LSTM layers, each with a size of (batch size, 2048). The outputs from both modules are merged through element-wise multiplication and fed into the classification layer with 8 nodes representing the different classes (benign and malignant subtypes). The model employs the SoftMax activation function.
For optimization, the paper [5] uses various hyperparameters, including batch size, activation function, optimizer (Adam), learning rate, and the number of epochs. Model checkpoints and callbacks are used during initial training and fine-tuning to prevent overfitting. The optimal parameters are determined through testing and validation to achieve robust and accurate predictions. The experiments are conducted on Google Collaboratory with specific hardware specifications, including 52GB of RAM, an Nvidia GeForce GPU, and development tools on the Ubuntu 64-bit operating system using Python 3.7.13. TensorFlow, Keras, CUDA, cuDNN libraries, and visualization tools like Matplotlib and Seaborn are employed for implementation.
The experimental setup [4] involves using the BreaKHis dataset, which includes images of breast histopathology at different magnification levels (40×, 100×, 200×, 400×). The dataset is divided into training and validation sets with a 70%–30% split. The ABN-DCN model is trained using a Tesla K-80 GPU on the Google Colaboratory platform, known for its reliability in image analysis. For each magnification factor, the model undergoes training over 200 epochs with a mini-batch size of 32. Stochastic gradient descent with a momentum of 0.9, weight decay, and a learning rate of 0.1 is employed as the optimizer. The DCN feature extractor, based on a variant of DarkNet19, is fine-tuned specifically for breast histopathological image analysis. The training process involves optimizing the DCN feature extractor based on the gradients from both the

attention and perception branches of the ABN-DCN model. Performance evaluation metrics such as accuracy, precision, recall, and F1 score are used to assess the model's classification capabilities.
The experiments [1], including training, testing, and validation, were conducted in a Linux CentOS 7 environment on a high-performance compute system. The system used Broadwell Xeon processors and two NVIDIA QUADRO P6000 graphics processing units (GPUs). The mention of GPUs indicates that these were employed for their computational capabilities, which are particularly useful in image processing tasks. The training dataset comprised 119 aRMS samples, 103 eRMS samples, and 15 CCS samples. The dataset was split into a 70% training subset, a 15% validation subset, and a 15% testing subset.
The experiments described in the research article[2] were conducted using Python 3.7.6, TensorFlow 2.1.0, and Keras 2.2.4. The computational setup involved a standard PC equipped with dual Nvidia GeForce GTX 2070 GPUs, 32.0 GB of RAM, and a 3.60 GHz Intel CoreTM i9-9900K processor with 16 logical threads and 16 MB of cache memory. The presence of dual GPUs indicates parallel processing capabilities, which is beneficial for deep learning tasks, allowing for faster training and inference times. The experimental setup in the research paper involved the collection of 544 whole slide images (WSI) from 80 breast cancer patients at the pathology department of Colsanitas Colombia University, Bogotá, Colombia. The tumor tissue fragments were fixed, embedded in paraffin, and subjected to 4 mm cuts stained with hematoxylin and eosin (H & E). Immunohistochemistry studies were conducted for various biomarkers. The WSI images were scanned at high resolution (400×) using a Roche iScan HT scanner.

2.5.2 Results and Findings

The paper [3] presents comprehensive results on the performance of the proposed Panoptes model compared to other CNN architectures, including Inception and InceptionResnet models. The evaluation metrics include the area under the receiver operating characteristic curve (AUROC), precision, recall, sensitivity, specificity, and F1 scores.
The Panoptes model [3] achieves good predictive performance across various histopathology image analysis tasks. Specifically, it outperforms other architectures in tasks related to predicting histological subtypes, molecular subtypes, and the mutation status of genes associated with endometrial carcinoma. The use of a multi-resolution design in Panoptes is emphasized as contributing to better predictive performance, allowing the model to consider both macro-tissue-level and minute cellular-level features simultaneously.

The results [3] are presented in terms of AUROC scores for different tasks, demonstrating the model's ability to discriminate between classes effectively. For instance:

- Histology Classification (Panoptes2): Per-patient AUROC: 0.969; Per-tile AUROC: 0.870
- CNV-H Prediction from Endometrioid (Panoptes1): Per-patient AUROC: 0.958; Per-tile AUROC: 0.864
- POLE Subtype Classification (Multi-model system): Per-patient AUROC: 0.890; Per-tile AUROC: 0.691

These scores indicate the model's robust performance in distinguishing different histopathological and molecular features. Additionally, the paper provides insights into the interpretability of the model's predictions by visualizing activation maps and using t-distributed stochastic neighbor embedding (tSNE) plots.
The proposed hybrid CNN-LSTM model [5] achieves notable results, surpassing existing state-of-the-art CNN models for both binary and multi-class classification of benign and malignant breast cancer subtypes. The performance is evaluated using various metrics such as accuracy, precision, recall, top-k categorical accuracy, precision-recall curves, and ROC curves. The binary classifier exhibits high accuracy for different magnification factors (99.03%, 99.75%, 99.64%, and 98.07% for 40×, 100×, 200×, and 400×, respectively). Similarly, the multi-class classifier demonstrates accuracy values of 96.5%, 92.6%, 88.04%, and 92.51% for the corresponding magnification factors. The precision-recall curves and ROC curves show the model's robustness and effectiveness in distinguishing between different classes. The proposed model outperforms existing methods in terms of accuracy and error loss, demonstrating its potential for automated diagnosis of breast cancer subtypes from histopathological images.
In the research paper [2], the authors trained four models based on pre-trained VGG16 and VGG19 architectures and evaluated their performance using a dataset comprising 544 whole slide images from 80 breast cancer patients. The models included fully-trained and fine-tuned versions of both VGG16 and VGG19, and 5-fold cross-validation was performed for each. The individual model performances varied, with the fine-tuned VGG19 model showing the highest average recall for carcinoma images at 95.68%, an overall accuracy of 91.67%, and an F1 score of 91.63%. The ensemble approach, combining fine-tuned VGG16 and VGG19 models, outperformed individual models, achieving a sensitivity of 97.73% for carcinoma, an overall accuracy of 95.29%, and an F1 score of 95.29%. Despite the comparatively smaller dataset, the ensemble model demonstrated competitive classification performance when compared to state-of-the-art studies using the BreakHis dataset. The research paper highlights the effectiveness of the proposed ensemble deep learning approach for the automatic classification of breast cancer histopathology images, particularly for carcinoma cases.
The ABN-DCN model [4] achieves impressive classification accuracies across various magnification levels: 98.4% for 40×, 98.6% for 100×, 98.7% for 200×, and 97.8% for 400×. Precision, recall, and F1 score metrics consistently indicate strong performance. In comparison with other CNN models (VGG16, MobileNet, EfficientNet) and the baseline DCN model, ABN-DCN outperforms, showing the superiority of its attention-based approach. The ABN-DCN model demonstrates robust performance in breast histopathological image classification, achieving high accuracy, precision, recall, and F1 score across different magnification levels (40×, 100×, 200×, and 400×). The model introduces diagnostic interpretability through attention maps, allowing visualization of regions influencing its decisions and aligning with pathologists' annotations. In comparisons with low-footprint CNN models (VGG16, MobileNet, EfficientNet) and the baseline DCN model, ABN-DCN consistently outperforms in classification metrics.
The primary focus of the research paper is to address the challenge of accurately differentiating between pediatric soft-tissue sarcomas, specifically clear cell sarcoma (CCS), alveolar rhabdomyosarcoma (aRMS), and embryonal rhabdomyosarcoma (eRMS) using deep learning techniques. The training set included 119 aRMS, 103 eRMS, and 15 CCS samples, and the models were validated using a 70% training, 15% validation, and 15% testing split. The research paper also involved a murine histopathology dataset,

which included 206 aRMS tissue slides, 54 eRMS tissue slides, and 58 normal murine skeletal muscle tissue slides. The CNN models trained on this dataset demonstrated high AUC scores, with values of 0.97 for aRMS, 0.98 for eRMS, and 1.0 for normal tissue slides on the withheld testing tile set. Despite these achievements, the study acknowledged certain challenges and limitations. These include the need for a more extensive and diverse dataset, potential biases arising from differences in slide scanning protocols, and the difficulty in handling histopathology images from various sources.

2.5.3 Architectures

The research paper [2] utilized VGG16 and VGG19 pre-trained convolutional neural network (CNN) architectures for the classification of breast cancer histopathology images. Two distinct configurations, namely fully-trained and fine-tuned, were employed for each of the VGG16 and VGG19 models. Additionally, an ensemble strategy was explored by combining the predictions from the fine-tuned VGG16 and VGG19 models. This ensemble approach aimed to enhance the classification performance, particularly emphasizing the accurate identification of carcinoma images within the dataset. The sensitivity for the carcinoma class reached 97.73%, with an overall accuracy of 95.29% and an F1 score of 95.29%. The research emphasized the effectiveness of the proposed ensemble deep learning approach for the classification of complex breast cancer histopathology images, showcasing its potential for practical diagnostic applications.
The proposed ABN-DCN [4] (Attention Branch Network - DarkNet19) model has a feature extractor, an attention branch, and a perception branch. The feature extractor is based on a variant of DarkNet19, referred to as the DCN model, which has 17 convolution layers with Batch Normalization and Leaky ReLU operations following each layer. The DCN model utilizes Max Pooling for downsizing inputs and adopts a 256×256 pixel resolution for input images. The attention branch is constructed based on the Class Activation Mapping (CAM) approach, which includes a K×3×3 convolution layer, global average pooling (GAP), and a 1×1×1 convolution layer. The attention branch generates attention maps through convolution operations and normalization using the sigmoid function. The perception branch processes the feature and attention maps to produce the final probability for each class. An attention mechanism is applied to the feature map, enhancing the interpretability of the model. The overall architecture leverages the collaboration between the attention branch and the feature extractor, allowing the ABN-DCN model to focus on relevant regions for decision-making in breast histopathology image classification.
Panoptes [3] employs a unique multi-resolution InceptionResNet-based design, featuring three branches processing samples at distinct resolutions (2.5x, 5x, and 10x). Each branch operates independently until the third-to-last layer, where their outputs are concatenated. Notably, the model processes sets of three tiles, representing the same region on an H&E slide, and amalgamates them into a single matrix. A fourth branch is dedicated to processing clinical features, such as patients' age and BMI.
The model is trained on datasets from two significant cancer research entities, CPTAC and TCGA [3], ensuring diversity in patient demographics, tumor characteristics, and molecular profiles. The tile-based approach and multi-resolution tile sets allow the model to capture local features and patterns effectively.
Performance metrics include AUROC, precision, recall, sensitivity, specificity, and F1 scores. Visualization tools, such as activation maps and 2D tSNE plots, pro-

vide insights into the interpretability of the model. The reported AUROC scores demonstrate the model's efficacy in tasks related to histological and molecular subtype classification[3].
The hybrid CNN-LSTM model[5] combines pre-trained CNNs (InceptionResNetV2 and ResNet50) for spatial feature extraction and LSTMs for capturing temporal dependencies. The CNN processes images with an input shape of (299×299×3), pre-trained using transfer learning from ImageNet weights. The LSTM module consists of two layers, and the final classification layer has 8 nodes representing different breast cancer subtypes.
Data augmentation techniques, including random variations in intensity, rotation, flip, and translation, address overfitting and imbalanced class challenges during training. The BreakHis dataset [5], containing 7,909 microscopic images of breast tumor tissue at various magnification factors, is used for training and testing.
The model is evaluated using metrics such as accuracy, precision, recall, top k-categorical accuracy, precision-recall curves, and ROC curves. The evaluation includes both binary (benign vs. malignant) and multi-class classification scenarios. The reported accuracy values demonstrate the model's effectiveness in breast cancer subtype classification [5].

2.6 Personal Results

In this section, I will discuss my individual experiments, where I use a dataset referenced in one of the cited research papers and implemented a methodology used in the research paper.

2.6.1 Experimental Setup

The experiment was conducted on a laptop featuring the NVIDIA GPU, specifically a GeForce GTX 1050 with 4GB of memory. The driver version is 537.13, and the CUDA version is 12.2. The GPU operates in WDDM (Windows Display Driver Model) mode.
The primary focus of the experiment was breast cancer classification using the BreakHis dataset, comprising 7,909 histopathological images from both benign and malignant cases. To ensure a balanced representation, an 80% training, 10% validation, and 10% test split were applied.
In terms of data preprocessing, images were resized to (224, 224, 3) pixels, and pixel values were normalized to the [0, 1] range. The dataset was divided into training (6,327 images), validation (791 images), and test sets (791 images).
The model architecture involved using the pre-trained VGG16 as the base model, with custom top layers including Flatten, Batch Normalization, and Dense layers. The total number of parameters for the model was 21,270,849, with 6,505,985 being trainable.
For training configuration, the Adam optimizer with a learning rate of 0.001 was used, employing Binary Crossentropy as the loss function. The training process spanned 19 epochs, with accuracy as the primary metric. EarlyStopping (patience=5) and ModelCheckpoint were used as callbacks.
Data augmentation techniques, such as random rotation, horizontal flip, width shift, height shift, and zoom, were applied to enhance the effective dataset size.
During training, a batch size of 32 images per batch was used for 19 epochs. The training process took approximately 8 hours on a GPU-powered environment. The training

progressed through iterative epochs, with the EarlyStopping mechanism monitoring the validation loss. If the validation loss did not showcase improvement over a predefined patience period of 5 epochs, the training process was halted. ModelCheckpoint was utilized to save the model's weights at the epoch where the validation loss reached its minimum, ensuring that the saved model represented the configuration where the model exhibited the best performance on the validation set.
The Adam optimizer dynamically adapted the learning rate during training, allowing the model to converge efficiently. The initial learning rate was set at 0.001, a commonly used value for training deep neural networks. The chosen batch size of 32 images per batch struck a balance between computational efficiency and model convergence, optimizing memory utilization.

2.6.2 Results

In the breast cancer classification experiment using the VGG16 model, the BreakHis dataset, consisting of 7,909 breast histopathological images, was utilized. The data was strategically split into 80% for training, 10% for validation, and 10% for testing to maintain a balanced representation of benign and malignant samples across the subsets.
For data preprocessing, all images were resized to (224, 224, 3) pixels, aligning them with VGG16's input size. Pixel values were normalized to the [0, 1] range to facilitate model convergence during training. The dataset was further divided into distinct training (6,327 images), validation (791 images), and test sets (791 images).
The model architecture employed VGG16 as the base model, complemented by custom top layers. These included a Flatten layer, Batch Normalization for stability, and Dense layers with specific neuron counts, ReLU activations, and dropout layers to mitigate overfitting. The entire model comprised 21,270,849 parameters, with 6,505,985 being trainable.
Training configuration involved the use of the Adam optimizer with a learning rate of 0.001 and Binary Crossentropy as the loss function. The primary metric for evaluation was accuracy. Callbacks such as EarlyStopping with patience=5 and ModelCheckpoint were implemented to enhance training efficiency.
The experimental results showcased notable improvements in both training and validation accuracy. Starting at 65.16%, training accuracy reached 87.59%, while validation accuracy progressed from 72.99% to 89.89%. The test set evaluation included a confusion matrix and key metrics, such as accuracy (89.38%), precision (88.70%), recall (97.09%), specificity (71.78%), and F1-score (92.71%). A detailed classification report highlighted metrics for both benign and malignant classes, emphasizing precision, recall, and F1-score values. The experimental results highlighted significant improvements in both training and validation accuracy.

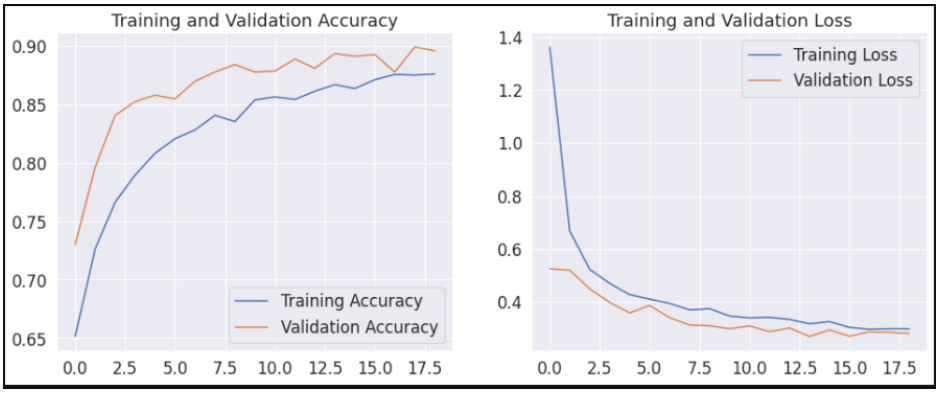

Figure 2.1: Training and Validation Accuracy & Training and Validation Loss

2.7 Coparison with the research paper

The experimental setup in the research paper[2] utilized a PC with dual Nvidia GeForce GTX 2070 GPUs, 32.0 GB RAM, and an Intel CoreTM i9-9900K processor for training and evaluating the breast cancer histopathology image classification models. In contrast, my experiment utilized a laptop with an Nvidia GeForce GTX 1050 GPU, 4GB memory, and an Intel CoreTM i7 processor.

In terms of datasets, the research paper used a private dataset with 544 whole slide images (WSI) from 80 breast cancer patients, while my experiment employed the BreakHis dataset, comprising 7,909 histopathological images from benign and malignant cases.

The preprocessing steps differed slightly, with the research paper resizing images without normalization due to color normalization challenges, while the other experiment resized images to (224, 224, 3) pixels and normalized pixel values to the [0, 1] range.

Both studies employed VGG16-based architectures with custom top layers for classification. The research paper used an ensemble of fine-tuned VGG16 and VGG19 models, where the other experiment utilized a single VGG16 model.

The research paper focused on achieving sensitivity of 97.73% for the carcinoma class and overall accuracy of 95.29%. In contrast, the other experiment reported a test set accuracy of 89.38%, precision of 88.70%, recall of 97.09%, specificity of 71.78%, and F1-score of 92.71%. The detailed classification report in the other experiment emphasized metrics for both benign and malignant classes.

While both studies demonstrated improvements in accuracy, precision, recall, and F1-score, the differences in datasets, preprocessing, and model architectures make direct comparisons challenging. Each study addressed specific challenges related to their datasets and experimental setups, contributing valuable insights to the field of breast cancer histopathology image classification.

2.8 What is new compared to existing SOTA

I've introduced the three research papers, discussed my experiments, and shared the implemented code available on GitHub. I'd like to highlight a significant shift from my previous work. In the last session, I focused on identifying tumors in chest X-rays. However, in my current project, I've narrowed my focus to tissue analysis. I employed Convolutional Neural Networks for classification, specifically utilizing the VGG16 model. The goal was to achieve results similar to the research I've selected. This shift represents a deliberate change from chest X-ray tumor detection to a more targeted exploration of histopathological features.

2.9 Github Link

Github link: Implementation

2.10 Conclusion

In conclusion, the examination of the research papers on deep learning applications in histopathology image analysis offers valuable insights into the evolving field of medical image processing, particularly in the context of breast cancer and pediatric soft-tissue sarcoma. Despite variations in datasets, preprocessing methods, and model architectures, certain key patterns and advancements emerge.

The consistent reliance on convolutional neural networks (CNNs) underscores their crucial role in effectively extracting essential features from histopathological images. The use of extensive datasets, exemplified by the BreakHis dataset, reinforces the effectiveness of these deep learning models in distinguishing patterns indicative of benign and malignant tissue samples.

Each research paper contributes uniquely to the field. Some introduce innovative network architectures, such as the hybrid CNN-LSTM model, which integrates spatial and temporal dependencies for breast cancer subtype classification. Others bring forth novel data preprocessing techniques, like meticulous augmentation strategies to address overfitting and imbalanced class challenges during training.

The reported metrics, including accuracy, precision, recall, specificity, and F1-score, serve as benchmarks for evaluating the models' performance. Notable improvements in training and validation accuracy, as demonstrated in the VGG16 model experiment, underscore the practical success of these deep learning approaches in achieving accurate classification outcomes.

Furthermore, the exploration of multi-modal data integration, as suggested in combining histopathology images with genetic or clinical information, holds promise for elevating diagnostic accuracy. This approach aligns with the broader trend in medical research toward comprehensive patient profiling for more personalized and effective treatment strategies. The persistent call for standardization in evaluation metrics and the creation of benchmark datasets underscores the need for robust frameworks to facilitate fair and meaningful comparisons between different deep learning algorithms. Such standardization is crucial for fostering collaboration, ensuring reproducibility, and advancing the field collectively to improve diagnostic capabilities.

In essence, these research papers deepen our understanding of feature representation in histopathological images and pave the way for future investigations. The convergence of cutting-edge technologies with medical image analysis continues to drive the revolution in pathology practices, aiming to enhance patient outcomes and advance the realm of diagnostic medicine.

Bibliography

[1] Arthur O. Frankel, Melvin Lathara, Celine Y. Shaw, Owen Wogmon, Jacob M. Jackson, Mattie M. Clark, Navah Eshraghi, Stephanie E. Keenen, Andrew D. Woods, Reshma Purohit, Yukitomo Ishi, Nirupama Moran, Mariko Eguchi, Farhat Ul Ain Ahmed, Sara Khan, Maria Ioannou, Konstantinos Perivoliotis, Pin Li, Huixia Zhou, Ahmad Alkhaledi, Elizabeth J. Davis, Danielle Galipeau, R.L. Randall, Agnieszka Wozniak, Patrick Schoffski, Che-Jui Lee, Paul H. Huang, Robin L. Jones, Brian P. Rubin, Morgan Darrow, Ganapati Srinivasa, Erin R. Rudzinski, Sonja Chen, Noah E. Berlow, and Charles Keller. Machine learning for rhabdomyosarcoma histopathology. *Modern Pathology*, 35(9):1193–1203, 2022.

[2] Zabit Hameed, Sofia Zahia, Begonya Garcia-Zapirain, José Javier Aguirre, and Ana María Vanegas. Breast cancer histopathology image classification using an ensemble of deep learning models. *Sensors*, 20(16), 2020.

[3] Runyu Hong, Wenke Liu, Deborah DeLair, Narges Razavian, and David Fenyö. Predicting endometrial cancer subtypes and molecular features from histopathology images using multi-resolution deep learning models. *Cell Reports Medicine*, 2(9):100400, 2021.

[4] Sruthi Krishna, S.S. Suganthi, Arnav Bhavsar, Jyotsna Yesodharan, and Shivsubramani Krishnamoorthy. An interpretable decision-support model for breast cancer diagnosis using histopathology images. *Journal of Pathology Informatics*, 14:100319, 2023.

[5] Mahati Munikoti Srikantamurthy, V. P. Subramanyam Rallabandi, Dawood Babu Dudekula, Sathishkumar Natarajan, and Junhyung Park. Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid cnn-lstm based transfer learning. *BMC Medical Imaging*, 23(1):19, 2023.