A DEEP LEARNING APPROACH FOR CLASSIFICATION OF LIVER DISEASE

BY

FARIA BINTA KALAM
ID:193-15-2918

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Md. Afjal Hossan Sarower
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Co-Supervised By

Mr. Golam Rabbany
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
JANUARY 2024

APPROVAL

This thesis titled "**A Deep Learning Approach for Classification of Liver Disease**", submitted by Faria Binta Kalam, ID: 193-15-2918 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 22 January 2024.

BOARD OF EXAMINERS

Md Tamim Ahad

Dr. Md. Taimur Ahad (MTA)
Associate Professor & Associate Head — **Chairman**
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

22/01/24

Fourcan Karim Mazumder (FKM) — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Firoz Hasan 22.1.24

Md. Firoz Hasan (FH) — **Internal Examiner**
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Abu Sayed Md. Mostafizur Rahaman (ASMR) — **External Examiner**
Professor
Department of Computer Science and Engineering
Jahangirnagar University

DECLARATION

I hereby declare that this project has been done by me under the supervision of **Mr. Afjal Hossan Sarower, Lecturer (Senior Scale), Department of CSE,** Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

07.01.2024

Mr. Afjal Hossan Sarower
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Co-Supervised by:

Mr. Golam Rabbany
Lecturer
Department of CSE
Daffodil International University

Submitted by:

Faria

Faria Binta Kalam
ID: 193-15-2918
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, I express my heartiest thanks and gratefulness to Almighty Allah for His divine blessing which makes me possible to complete the final year project/internship successfully.

I really grateful and wish my profound indebtedness to **Mr. Afjal Hossan Sarower**, Lecturer (Senior Scale), Department of CSE, Daffodil International University, Dhaka, deep knowledge & keen interest of my supervisor in the field of Machine Learning to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

I would like to express my heartiest gratitude to **Dr. Sheak Rashed Haider Noori**, Head**,** Department of CSE, for his kind help to finish my project and also to other faculty members and the staffs of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, I must acknowledge with due respect the constant support and patients of my parents.

ABSTRACT

Liver diseases pose a significant global health burden, with diverse manifestations such as ballooning, fibrosis, inflammation, and steatosis. Accurate and timely diagnosis is crucial for effective treatment planning and patient management. This thesis explores the application of deep learning models, including EfficientNetB2, VGG16, InceptionNetV3, DenseNet121, and ResNet50, for the comprehensive classification of liver diseases based on these distinct pathological features. The study involves a robust dataset of liver pathology images, capturing various stages and manifestations of liver diseases. Through an exhaustive analysis, we compare the performance of different deep learning architectures in accurately identifying and classifying ballooning, fibrosis, inflammation, and steatosis. Our experiments reveal that EfficientNetB2 outperforms the other models in terms of accuracy, demonstrating its efficacy in handling the complexities of liver disease classification. In addition to model performance, the thesis delves into interpretability, providing insights into the features and patterns learned by each model. This contributes to a better understanding of the decision-making process and enhances the clinical relevance of the deep learning models in real-world scenarios. The findings of this research not only showcase the potential of deep learning in liver disease diagnosis but also highlight the significance of selecting appropriate architectures for optimal results. The implementation of EfficientNetB2 in this context opens avenues for improved diagnostic tools and automated systems that can aid healthcare professionals in making more informed decisions for patients with liver diseases. The implications of this study extend beyond liver disease classification, emphasizing the broader applicability of deep learning in medical imaging and pathology. The insights gained from this research contribute to the ongoing efforts to enhance the accuracy and efficiency of computer-aided diagnostic systems in the field of hepatology.

Keyword: Liver Disease, Deep Learning, EfficientNetB2, VGG16, InceptionNetV3, DenseNet121, ResNet50.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

TABLES **PAGE NO**

CHAPTER 1
INTRODUCTION

1.1 Introduction

Liver disease is representing a big global health problem encompassing many types of disorders that are damaging the structure and function of the liver in negative ways. Among these numerous disorders, four main pathological classifications namely ballooning, fibrosis, inflammation, and steatosis are playing pivotal roles in deciding the severity and future prognosis of liver diseases for patients [1][2]. Early and accurate identification of these pathological features on a consistent basis is being extremely crucial for allowing effective clinical management by doctors, personalized formations of treatment methodologies on case-by-case basis, and bigger improvements in final health outcomes of patients [3]. However, the traditional methods of diagnostic that are often dependent on manual assessments by pathologists using visual slides are having innate subjectivities and possibilities of diagnostic errors during analysis, which is emphasizing the urgent need for more advanced and objective methodological options [4]. In the most recent years, the emergence of novel deep learning (DL) techniques has effected transformations in capabilities of analysis within the medical field of image analysis, now offering automated and efficient solutions perfectly suited for disease detection as well as subsequent accurate classification [5]. The proven immense successes of deep learning various applications within healthcare, particularly prominently in liver disease identification, has successfully sparked considerable interests regarding assessment of its potential roles within liver disease classification contexts [6]. With liver diseases posing huge burdens on healthcare systems globally, there is existing a very pressing global need for exploring innovative deep learning technologies that can enhance diagnostic accuracies to higher levels and simultaneously streamline decision-making abilities of clinicians during critical times [7]. This current research initiative is endeavoring to address the aforementioned challenges by harnessing the analytical powers of deep learning for the accurate classification of liver diseases, by specifically focusing on the pathological types - ballooning, fibrosis,

inflammation, and steatosis [8]. Analogous to the past successful applications of deep learning techniques in liver diseases classifications, this current study is aiming to assess the effectiveness of deep learning models for accurately diagnosing and categorizing different liver disease types [9]. By methodically leveraging large datasets of histological images sourced from patients exhibiting diverse liver pathologies, the research is seeking to comprehensively evaluate classification performance, robustness under variations, and the complete generalizability capacity across populations for the developed deep learning model. Transfer learning, a deep learning technique already demonstrated to markedly enhance proficiencies of deep learning models across challenging medical image analysis tasks, will consist of a prime investigation focus within this study. By efficiently leveraging pre-developed models already trained over extensive datasets, transfer learning concentrates on allowing extractions of relevant features from newly analyzed histological images, therefore potentially improving classification accuracy metrics and model robustness by multiple folds. The investigative analyses will extend well beyond mere numerical accuracy measurements, deliberately delving into the intricate interpretability aspects of the created models. Visualizing the most utilized learned features and sequentially analyzing their computed relevance toward liver disease classification will facilitate enhanced understandings of the conditional workings of models' internal algorithms and subsequent decision-making capacities. Ultimately, this research work aims to contribute significances toward the advancement of medical image analytics within the specific realms of liver disease accurate classification [10]. By providing automated and reliable deep learning solutions in the domain, the produced findings of this study will hope to revolutionize clinical practices, therefore aiding clinicians during critical decision-making instances, and finally culminating in aspirations of improved patient care levels and health outcomes within the truly challenging landscape of liver diseases management globally.

1.2 Motivation

The rationale Liver disease, encompassing numerous complex forms including ballooning, fibrosis, inflammation, and steatosis pathology types, is constituting a very serious global health issue that is negatively affecting countless millions of patients worldwide on a

consistent basis. Each uniquely complex form contains differing sets of deleterious symptoms and requires specialized targeted treatments personalized to the specific manifested disease type for achieving optimal results. Accurately identifying the precise type of liver disease through diligent diagnosis therefore attains extreme significance for allowing administration of the correctly corresponding effective disease-specific treatment methodologies available and subsequently predicting the most accurate prognoses for patients. However, the traditionally utilized methods of clinical diagnosis, heavily reliant on exhaustive manual examinations conducted by specialized pathologists utilizing visual slides, are containing inherent susceptibilities to inadvertent errors in human judgement and consequently may not always be guaranteeing absolute accuracy during every unique analysis. This is majorly underscoring the existing urgent need within current medical landscapes for conception of more advanced and objective diagnostic tools utilizing the latest technologies availably recently. Very recent profound advancements accomplished in the domain of deep learning, consisting of a novel subset belonging to the artificial intelligence spheres, have exhibited exceptionally great potentials in the context of applications inside the medical field of image analytics and classification. Novel deep learning algorithms possess the invaluable capabilities to efficiently analyze truly vast amounts of complex graphical data and subsequentially perform very complex computational tasks such as specialized image feature recognition, intensive natural language processing, and human-like speech synthesis with great accuracies. They have been successfully implemented previously across numerous pilots in extremely challenging domains of liver disease detection, diagnosis, prognostics formulations, and treatments design with very promising results demonstrated consistently. This current innovative research is deriving strong motivations based on assessing the full potentials of deep learning to completely revolutionize the clinical practices involved in diagnosis and subsequent pathological classification of major liver diseases, specifically targeting the common types - ballooning, fibrosis, inflammation, and steatosis categories detectable through scans. The fundamental choice of this particular research topic is stemming from my personal exceptionally strong interests in creatively applying cutting-edge deep learning techniques to solve complex real-world problems existing within biomedical datasets and environments. My academic backgrounds consist of specialized computer

science and bioinformatics histories, having prior extensive experiences in independent development and applications of deep learning models on immense varieties of complex biological datasets for intriguing explorations. I was inspires by the most recent advances and identifiable challenges remaining in liver disease accurate subclassification and the potentials of deep learning to contribute very novel insights and optimally feasible solutions going ahead. Additionally, I have a close family member who has unfortunately suffered previously from one form of serious liver disease. Their debilitating condition was extremely difficult to accurately diagnose completely during that time due to current lack of effectively advanced diagnostic tools. This had sparked my personal interests toward passionately exploring new modern methodologies to drastically improve the diagnosis and optimal treatment of all liver diseases in future. I sincerely hope that my research undertakings shall eventually contribute significances toward improving the quality of lives and personal outcomes for all similar individuals currently suffering anywhere from such liver diseases.

1.3 Problem Statement

Liver disease is constituting a very significant health concern that is affecting substantial numbers of populations globally in negative manners. Unfortunately, there are not existences of enough effective treatment options as of current scenarios and the overall outlook received by people diagnosed is still not appearing very optimistic. The liver disease encompasses heterogenic manifestations taking forms of several subclassifications, prominently including ballooning, fibrosis, inflammation, and steatosis. Each uniquely complex form is containing differing sets of symptoms likely visible and will be requiring administration of specifically customized treatments designed. Accurately identifying the correct type of liver disease through diligent diagnoses therefore attains great significance for facilitating selection of appropriately effective treatments methodology and also predicting the most accurate prognosis expectations for patients. However, the conventional methods leveraged currently for diagnoses, heavily dependent on exhaustive manual examinations done by expert pathologists assessing biopsies slides, are harboring inherent susceptibilities to inadvertent errors attributable to human judgement limitations and consequently may not be able to guarantee full accuracies with each performed

analysis. This fact majorly is underscoring the urgent requirements for conceptions of more technologically advanced and objective diagnostic tools to counter currently existing challenges. Deep learning, referring to an effective subset within the artificial intelligence application domains currently, has exhibited great potentials in contexts of detailed image analytics and classification tasks executions. It has attractions of strong promises perceived from medical community mainly toward contributing researching of medical imaging. This one prime research is aiming to construct optimized deep-learning model variations that can easily differentiate between the pathological subclassifications of liver diseases leveraging digitized histology images feeds as inputs. The iterative model will be rigorously trained over sizable datasets encompassing detailed histological images and equally validated for performance indicators comprising accuracies metrics calculations and resilience capacities against noise additions. Successfully developing such proposed model can greatly enable massive improvements in registering unerring accuracies during liver disease classifications and therefore ultimately facilitate selections of optimally effective personalized treatment methodologies to administer over patients. This research effort shall also contribute to accumulations of complementing informative evidences on this certain critical topic. Applications of Deep Learning are possessing all vital capacities to impart significantly desirable impacts while progressing the domains of medical imaging research, explicitly where driving cutting-edge advancements targeted toward examining chronic liver diseases.

1.4 Research Objectives

a) One core research objective is constituting the successful development of a robust and highly accurate deep learning model variation possessing optimized capacities for efficiently extracting many relevant features encompassed within detailed histological images of liver diseases provided as inputs.

b) Another pivotal research objective is comprehensively investigating and subsequently comparing performances of different types of deep learning algorithms and associated techniques available for contextually classifying among the major pathological subforms of liver diseases based on detailed analysis over provided histology images datasets.

c) An additional critical research goal is optimal tuning of constructed deep learning model to enable accurately distinguishing capabilities between the major subclassifications of liver diseases leveraging feeds of high-resolution histological images of liver cells and tissues.

d) Another research objective is thoroughly assessing the generalizability potentials of the proposed novel classification system by extensively evaluating its real-world performance metrics over more diverse datasets encompassing ranges of variations in utilized staining techniques, tissue preparation methods, and differing qualities of images collected from multiple sources.

e) A supplemental research aim is conducting large-scale comparative analysis of developed classification system against collections of existing methodology options leveraging both manual and automated approaches for performance benchmarking.

1.5 Research Questions

a) Can a deep learning model accurately classify different forms of liver disease (ballooning, fibrosis, inflammation, and steatosis) using histological images?

b) How does the performance of the deep learning model compare to existing methods for liver disease classification?

c) What underlying mechanisms allow the deep learning model to classify liver diseases accurately?

d) Which features does the deep learning model use to make predictions, and how biologically relevant are they?

e) Can the deep learning model provide new insights into the biology of liver diseases and potentially identify new biomarkers or therapeutic targets?

1.6 Report Layout

This report consists of five chapters. The project report consists of five chapters, each addressing different aspects of the research. Chapter 1 encompasses the introduction, motivation, rationale, and research question, providing a comprehensive overview of the study's objectives. Chapter 2 delves into the background information related to the topic, explores similar work conducted in the field, and highlights the limitations of the research.

In Chapter 3, the focus shifts to the research methodology, discussing the methods employed for data collection, data preprocessing, analysis requirements, and various use cases. The chapter also explores techniques for presenting and representing the analyzed data. Chapter 4 presents the experimental results, including performance evaluation metrics, and provides a detailed discussion of the findings obtained from the analysis. Moving forward, Chapter 5 represents the impact on society, environment and ethical aspects and Chapter 6 offers an overview of the entire project, outlining future work possibilities, and concludes with a discussion on the completion of the research study.

CHAPTER 2
LITERATURE REVIEW

2.1 Preliminaries/Terminologies

Histopathology: The microscopic analysis of tissue samples, encompassing cells and the extracellular matrix, to investigate manifestations of diseases. In liver disease diagnosis, histopathology conducted on biopsy specimens remains the established gold standard.

Hematoxylin and Eosin (H&E) Stain: A widely employed histological stain consisting of the basic dye hematoxylin and the acidic dye eosin. Hematoxylin imparts a blue color to cell nuclei, while eosin stains cytoplasm and extracellular matrix pink. This staining technique enhances contrast for studying tissue architecture.

Immunohistochemistry (IHC): A specialized staining method utilizing antibodies to identify specific proteins, antigens, or biomarkers in tissues. IHC aids in pathological diagnosis and subtyping of diseases.

Artificial Intelligence (AI): The conceptualization and creation of computer systems capable of performing tasks typically requiring human intelligence, such as visual perception, speech recognition, decision-making, language translation, etc.

Deep Learning: A subset of machine learning grounded in artificial neural networks with multiple abstraction layers. These networks can autonomously acquire hierarchical feature representations from raw input data.

Convolutional Neural Networks (CNNs): A specialized deep learning architecture inspired by the organization of the visual cortex. CNNs incorporate convolution, pooling, and fully-connected layers, effectively capturing spatial hierarchies in visual data

Transfer Learning - A technique to repurpose an already trained deep learning model on a new related problem. It allows inheritance of learned feature maps without training a model from scratch.

Data Augmentation - Artificially creating new annotated training samples from existing data using transformations like rotations, flips etc. This technique reduces overfitting in deep learning.

High-Resolution Images - In digital pathology, scans with a higher sampling rate to obtain multi-gigapixel histology whole-slide images. This facilitates examination of fine tissue details.

The key terminology provides the requisite foundation to assess the role of advanced artificial intelligence and deep learning approaches for analysis of high-resolution histopathology images in liver diseases research, diagnosis and treatment planning.

2.2 Related Works

M. A. Hasan al. Grading of steatosis, fibrosis, lobular inflammation, and ballooning from liver pathology images using pre-trained convolutional neural networks" presents a method for grading histological features, including fibrosis and ballooning, from liver pathology images using pre-trained convolutional neural networks. The classification accuracy was reported to be 96.26% [11]. An Ultrasound-Based Computer-Aided Diagnosis Tool for Steatosis presents a new computer-aided diagnosis (CAD) system for steatosis classification, both locally and globally. The system utilizes a Bayes classifier for the classification of steatosis. The accuracy of the system was reported to be 93.75%, and the area under the receiver operating characteristic (ROC) curve (AUC) was 0.9375, demonstrating its effectiveness in steatosis detection [12]. M. J. House et al write about texture analysis of MRI images, especially when combined with clinical variables, shows promise for non-invasive staging of liver fibrosis, with good discrimination of no fibrosis from mild or severe fibrosis. However, performance was lower in determining intermediate stage fibrosis. AUC value was 0.81 for liver fibrosis [13]. Roy et al. in 2021 proposed a novel edge detection method based on computing local standard deviation value showed effectiveness in segmenting nuclei regions in liver cancer histopathology images. It outperformed other existing unsupervised methods and had comparable performance to recent deep neural models like DIST and HoverNet. Visual results and quantitative metrics (F1 score, Jacard index, PSNR) demonstrated the superiority of the proposed method in preserving nuclei boundary structure, reducing noise level, and achieving high nuclei detection accuracy. The method was tested on a multi-organ dataset, indicating its effectiveness over a wide variety of datasets. The paper also highlighted the importance of color normalization as a pre-processing step to reduce inter-color variance and enhance

contrast in HE stained histopathology images. Overall, the proposed method showed promising results in nuclei segmentation for liver cancer histopathology image [14]. The developed approach proposed by Hassan et al. in 2022 for drug response prediction of liver cancer cell lines achieved high accuracy of 97.5% and sensitivity of 100%, outperforming other methods. It showed effectiveness in accuracy, sensitivity, specificity, precision, F-score, MCC, and Kappa index, with an AUC of 96.4, and potential for other cell diseases [15]. Sun et al. in 2019 proposed the DeepLabV3+ semantic segmentation model, based on tensorflow architecture, demonstrated superior performance in liver tumor segmentation and lesion localization. The model was evaluated using various error measurement methods, resulting in accurate segmentation results. The logistic regression model demonstrated high sensitivity in recognizing liver cancer patient samples, indicating a low missed diagnosis rate. The paper also discussed data extraction, training, and feature extraction for liver tumor prediction [16]. The CNN provided efficient results in the detection of liver inflammation by encoding and decoding small information from segments. The examination was done on a pixel-to-pixel basis, and the outcome accuracy was assessed using the binary classification of the confusion matrix. The model achieved an accuracy of 98.6% on the image dataset and 96% on the 3D RealCT dataset by Kaluva et al. in 2018. The efficiency of the ResNet model can be further improved by using more datasets and different pre-processing techniques. The future scope of the research includes improving the accuracy of small-scale liver tumor diagnosis and the validation of the Dice Coefficient (F1 Score) [17]. Sadeque et al. in 2019 proposed model was tested on 50 liver CT images, with 27 confirmed cases. It utilized ROI and block normalization in feature vectors. The model's performance was evaluated using a confusion matrix, and it demonstrated 94% accuracy in detecting liver disease, saving time for doctors and aiding in treatment effect [18]. Proteomic profiling revealed overexpression of mortalin (HSPA9) in liver cancer (HCC), with higher levels in early recurrence subgroups. Mortalin overexpression was linked to advanced tumor stages and venous infiltration, suggesting mortalin as a potential biomarker for early recurrence by Yi et al. in 2008 [19]. The study evaluated the proposed framework using preliminary simulation experiments on 200 patients with various cancer types by Ali et al. in 2014. Data was collected from the Radiology department of Crosshouse Hospital Kilmarnock, UK. The experiments were

conducted on an Intel Core i5 Pentium processor with 8GB of RAM. Microsoft Visual Studio 2008, OpenCV, WEKA Experimenter, and Matlab R2013a were used for software development and experimental purposes. The study highlights the motivation and philosophy for early development of CAD, its current status, and potential for further investigation and development [20]. Messaoudi et al. in 2020 proposed HCC liver disease steatosis classification model achieved an accuracy level of 90%. The model showed improved performance compared to previous studies. The F1 score increased from 66.5 to 74 in this work. Sensitivity rate increased from 76 to 81 in phase 3. Specificity rates increased in both phases 2 and 3. Parallel patch-based processing of DCE-MRI images for HCC detection. Development of a novel algorithm using CNN architecture. Preprocessing, training, prediction, testing, and validation phases in the approach. Use of a database with normal and cancerous patches for training and testing [21]. Hepatitis prevalence in Taiwan increased from 2002 to 2010, with a decreasing trend among young people aged 16-30. A CNN model was used to predict liver cancer cases, with an accuracy of 0.980 and an AUC of 0.886 by Phan et al. in 2020. Deep learning models were used to predict liver cancer in a hepatitis cohort [22]. Sabut et al. in 2008 proposed a method for liver cancer detection that achieved a classification accuracy of 99.38% and a Jaccard index of 98.18%. A DNN classifier with 200 epochs showed minimal validation loss of 0.062. A new automated technique, combining watershed-Gaussian segmentation, gradient transformation, GMM, and deep learning, was proposed, achieving high classification accuracy and Jaccard index [23]. Anand et al. in 2023 proposed method using Autoencoder-Extreme Learning Machine (AE-ELM) and Convolutional Neural Network (CNN) technology achieved improved liver cancer detection and classification accuracy compared to classic machine learning approaches and standalone CNN models. The AE-ELM model, which reduces data dimensionality and classifies data, was found to be more accurate (99.23%) than CNN and ELM models. The use of AE-ELM and CNN technologies in liver cancer diagnosis improved accuracy, sensitivity, and specificity, enabling quick and precise diagnosis by healthcare practitioners. The proposed method extracted relevant characteristics and captured complicated patterns, improving liver cancer diagnosis accuracy and efficiency. To evaluate the therapeutic promise and application of this approach in real-world healthcare settings, additional datasets are needed [24]. Six different classifiers were

evaluated for tumor identification and classification. Accuracy achieved for tumor identification ranged from 98.39% to 100%. A multi-level ensemble model achieved high accuracy in tumor detection and classification. The ensemble model outperformed individual classifiers in both detection and classification Detecting the presence of tumors in liver CT images. Classifying different stages of tumors in liver CT images. Developing a multi-level ensemble model for tumor detection and classification. Achieving high accuracy in tumor detection and classification using the ensemble model by Krishan et al. in 2021 [25]. Ogihara et al. in 2016 proposed method has a sensitivity of 0.86 and specificity of 0.49 for test samples, potentially outperforming existing liver scoring systems. It can handle both qualitative and quantitative data through discretization, and has higher sensitivity in predicting early liver cancer recurrence compared to existing scoring systems [26]. The model proposed by Chaudhary et al. in 2018 provides two optimal subgroups of patients with significant survival differences (P = 7.13e-6) and good model fitness (C-index = 0.68) . The authors identified two differential survival subtypes in TCGA HCC data using an autoencoder-based deep learning framework. The survival analysis showed drastic differences in survivals between the two subclusters (log-rank P = 1.47e-6) . The DL-based methodology outperformed alternative approaches such as principal component analysis (PCA) and univariate Cox-PH analysis, in terms of significant log-rank P values and prediction metrics. Two survival subtypes in TCGA multi-omics HCC data are identified, and a deep learning-based model predicts HCC prognosis as well as alternative models. The model is validated on five external datasets and robust across multiple cohorts [27]. Shoaib Kareem et al. in 2021 proposed method achieves a 99.8% accuracy rate for liver cancer detection, with top-5 accuracies ranging from 98-99.8%. It uses FASTAI and UNets models, splits datasets into train, test, and validation sets, and augments training data [28]. The study by Patel et al. in 2019 showcases the use of AI technology for diagnosing breast, lung, and liver cancer, with the SVM algorithm showing the highest accuracy at 97.13%. Neural network classification algorithms are also used for early lung abnormality detection [29]. Particle Swarm Optimization (PSO) offers better accuracy and elapsed time for liver tumor detection AI technology is being utilized for accurate cancer diagnosis in breast, lung, and liver cancer. Automated systems provide

precision, and AI-implemented neural networks are the future of cancer treatment. AI aids doctors in diagnosis, predicts cancer survivability, and aids in early diagnosis.

2.3 Scope of the Problems

Global Health Impact: Liver disease is constituting a very significant global health concern that is currently affecting countless millions of patients worldwide on a consistent basis. It is ranking among the highest prevalent leading causes of mortality globally, which is therefore highlighting the existing urgent needs for extensive developments of optimally effective diagnostic and treatments strategies pipelines to counter currently rising cases.

Classification Systems: There are existences of several pivotal classification systems specifically designed over decades for categorizing wide range of liver diseases. These include the very renowned Lauren classification and the globally recognized World Health Organization's histological patterns methodology. Such systems have provided volumes of detailed understandings pertaining to the sheer complexity dimensions encompassed within liver diseases diagnoses challenges over long periods. However, these classification approaches are often still majorly based on traditional exhaustive manual examinations coupled with further highly subjective interpretations techniques involved, therefore leading toward potential inaccuracies creeping during each assessment attempts.

Detection Approaches: The very currently utilized detection techniques such as invasive endoscopy procedures or histological examination through biopsies slides, although reasonably accurate, but are still harboring few limitations like unnecessary invasive actions toward patients, triggering discomforts, and demands for extensive trainings of highly specialized professionals for subsequent interpretations tasks executions of slides assessments at acceptable accuracy metrics calculations. There are existing large-scale growing demands for conceptions of more significantly efficient and optimally less invasive detection methodologies pipelines to effectively overcome currently faced limitations systematically.

Model Evaluation: Very rigorous evaluations performed over model accuracies metrics involves methodical comparisons of generated autonomous predictions formulated against the considered ground truth labels classifications provided beforehand by specialized clinical domain experts leveraged. Advanced techniques including utilizing class activation

maps visualizations are utilized to explicitly showcase and properly interpret the most crucial identifiable regions presences within complex images feeds that are actively influencing the vital predictions generations by the model in biggest manners, therefore greatly aiding through the entire evaluations processes and cycles involved.

Clinical Integration: The ultimate end-goals within scopes and domains is to develop optimally optimized AI model variation that achieves highest-grade clinical accuracies metrics and reliability standards benchmarks before considering external practical deployments. Mandatory considerations requirements including safety and transparency features implementations are posing as integrally vital components aspects for designing the entire deployment pipeline stacks, therefore effectively ensuring fully streamlined integrations within daily clinical practices operations flows.

Ethical Considerations: Proactively establishments of necessary universal guidelines encompassing aspects including transparency levels enforcements, feasible auditabilities mechanisms introductions, and persistently maintained human oversight are posing critically vital to comprehensively ensure the complete ethical clinical deployments introducing AI tools designed specifically for enhancing liver disease diagnoses leveraging scans categories in fully accountable safe manners.

2.4 Challenges

a) Data Collection: One of the foremost core challenges present within this study categories was extensively acquiring adequate volumes of high-quality histological images essential for enabling in-depth contexts classifications of different pathological subforms of liver diseases leveraging deep learning algorithms at optimal levels. The activities of collecting sizable liver disease data from roboflow liver disease datasets, as most entities either were not in possessions of such niche data or were not appearing willing toward voluntarily sharing available data for augmentations of research purposes. Therefore, alternative viable data sourcing avenues had been necessarily sought after to assemble the minimally required data prerequisites.

b) Data Quality: Comprehensions pertaining to the vital importance of ensuring more stringent quality assurance across collected datasets leveraged are posing significant challenge aspects. Potentially, depending upon inherent sources originations, scanning

equipment utilized for digitizations, choice selections of specific staining techniques adapted before scanning, the histological images are likely demonstrating wider ranges of unpredictability and variances in terms of base quality interpretations, resolutions encodings, formatting types dissimilarities, and presence annotation indicators. Certain percentage of images may additionally require performance optimizations by applying corrections, manual completions of missing markings, or rectifying wrongly labeled ground truth inputs due to unavoidable inadvertent human errors crept in or minor technical limitations faced within digitization pipelines. These multiple issues generations could cumulatively and negatively impact the ensured performance metrics calculations and reliability consistency factors across the constructed deep learning model if left unaddressed systematically. Therefore, proper preprocessing executions over the collection of images for further classification tasks was ascertaining extreme significance within the study contexts, that had involved at scales – unified conversions of all images to singular commonly shared formats and dimensions adjustments, comprehensive removals of identifiable noises and artifacts contaminations, subsequent enhancements of relative contrasts ratios and individual brightness levels normalizations, and iterative verifications performed toward accuracy of annotated labels provided for every specific images involved within training sets or testing sets categories isolatedly.

c) Select Deep Learning Approach: This one precise research work had been primarily aiming to determine the optimally effective deep learning technique variations that could potentially enable accurate classifications of the diverse pathological subforms of liver diseases leveraging direct feeds of high-resolution histological images provided as viable inputs streams consistently. The domains of deep learning are referring to profoundly highly efficacious set of techniques enabled for extracting many complex latent patterns and intrinsic features encompassed within histological images types and therefore provides very accurate and significantly efficient image classification outputs generations at scales. However, there are existing wide ranges of plausible deep learning techniques that could likely be leveraged toward applications catering to multitudes of medical images analysis tasks resolutions attemptable, namely – convolutional neural networks (CNNs), recurrent neural networks (RNNs), generative adversarial networks (GANs), and latest transformer models. Each singular technique arbitrary contains specific sets of advantages

interpretations regarding exhibited overall accuracy metrics logged, computational efficiencies warranted, intrinsic interpretability potentials for understanding model behaviors, and expected generalization capabilities across multiple complex datasets varieties attempted for training experimentations. Therefore, availabilities of different promising deep learning techniques had mandated progressions of rigorous comparative analyses studies followed by final selections of one prime technique variation that shall be optimally satisfying maximum fitment factors w.r.t the targeted tasks panned and data properties involved respectively.

d) Accuracy Improvement: One among the very final challenges presences within wide scopes of this current study is comprehensively improving the overall performances metrics logged across the chosen deep learning model variation constructed, thereby methodically enabling selections of best final model versions most suitable for addressing the complex tasks at hand dexterously. The definite model's exhibited performances were iteratively improved by adjusting associated hyperparameters aspects like learning rates tuning, batch sizes calibrations, number of layers tuning or filters additions, and customized activation functions incorporations. Numerous data augmentation techniques incorporations including probable rotations, flipping variants applications, croppings, and intricate scalings trends were additionally explored; simultaneously introducing certain regularization techniques methods including dropouts, batch normalization, and mechanisms inducing weights decay were investigated accordingly. Moreover, highly specialized domain knowledge extractions such as integrating clinical features or molecular markers were researched within study contexts. Performing apt choices selections toward identifying one appropriate deep learning model is ascertaining paramount significance for achieving optimal balance addressing the specific target task addressed. The constructed model's performances were therefore confidently evaluated across wide ranges of pivotal metrics calculations logged, namely – computational accuracy, precision trends, recall optimization, F1 scores benchmarking and AUC levels analysis. Subjecting the model variations against exhaustive statistical tests coupled with associated confidence intervals formulations were utilized methods to compare performances deeply against existing solution variances available or contextually relevant baseline selections attempted within the same study as references markers. Lastly,

numerous advanced visualization techniques and explainable AI techniques instances were inculcated within model designs phases to gain much deeper insightful inferences into learning behaviors of model attempted significantly.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

Liver disease is a major health burden worldwide, causing over 2 million deaths per year globally. Accurate diagnosis and classification of liver disease is critical for determining proper treatment and management strategies. Recently, deep learning techniques have shown promise for automated analysis and classification of medical images to assist clinicians. In my research, I propose a deep learning approach for classifying liver disease into four categories - ballooning, fibrosis, inflammation, and steatosis - using histopathological images. Automated classification can help standardize diagnosis, reduce inter-observer variability, and improve patient outcomes. My methodology involves four key stages: data collection and preprocessing, model development, model evaluation, and result analysis. First, a dataset of liver histopathology images with verified disease labels will be collected and preprocessed to create a unified format suitable for training deep learning models. Next, transfer learning with several state-of-the-art convolutional neural network (CNN) architectures including EfficientNetB2, DenseNet121, Inception, ResNet50, and VGG16 will be explored for developing an accurate multi-class classifier. Performance metrics like accuracy, precision, recall, F1-score, and AUC-ROC will be monitored during training to select the best model. The model will then be rigorously evaluated on unseen test data to gauge real-world performance. Finally, the model output will be visually and statistically analyzed to determine strengths, limitations, and clinical relevance. Overall, the use of deep CNNs can help automate liver disease categorization in a fast and reliable manner. My research integrates robust data analysis with explainable deep learning model to build trustworthy decision-making systems. The model can provide a second opinion to doctors and serve as an initial screening tool for improved liver disease management. In subsequent sections, I describe the dataset preparation, model development, performance evaluation, and result analysis stages in further detail highlighting the experimental materials, methodology and expected outcomes of my work. The proposed research aims to harness advanced computational tools to improve clinical understanding of liver pathology.

3.2 Working Process

In this study, I offer a novel deep learning-based approach for automated classification of liver disease using histopathological images. The aim is to develop an accurate multi-class classifier that can categorize unseen liver tissue images into four classes - ballooning, fibrosis, inflammation, and steatosis. The methodology comprises four main stages:

i) Data Collection

ii) Image Preprocessing

iii) Model Development and Selection

iv) Result Analysis

The research focuses on leveraging recent advances in convolutional neural networks (CNNs) and transfer learning to analyze tissue morphology and patterns for disease diagnosis. Accurate pathology-based diagnosis can provide critical decision support to clinicians and improve clinical outcomes. However, manual examination of tissue slides is laborious and prone to subjectivity. Automated classification via deep learning promises to address these challenges and enhance the efficiency of histopathology workflows. The study utilizes histopathological images of liver tissue verified and labeled by clinical experts. Images are collected from open access datasets and institutional archives to ensure sufficient samples for model training and evaluation. Various preprocessing techniques are applied to standardize image dimensions, normalize staining variations, and perform data augmentation to increase dataset diversity. Several state-of-the-art CNN architectures including EfficientNetB2, DenseNet121, Inception, ResNet50 and VGG16 are trained on the curated dataset using transfer learning. This allows adapting powerful pre-trained models towards solving the liver disease classification task. Appropriate model selection is performed based on evaluation of multi-class metrics on validation data. Finally, the performance of the developed model is rigorously analyzed on unseen test images to determine its real-world viability. In summary, integrating explainable deep learning with histopathology aims to automate the classification of challenging liver conditions. The study attempts to harness modern AI innovations to improve understanding of tissue morphologies and support clinical decision-making ultimately benefiting patient care and outcomes. Subsequent sections provide specific details on the data curation, model development, performance benchmarking and diagnostic potential of the proposed work.

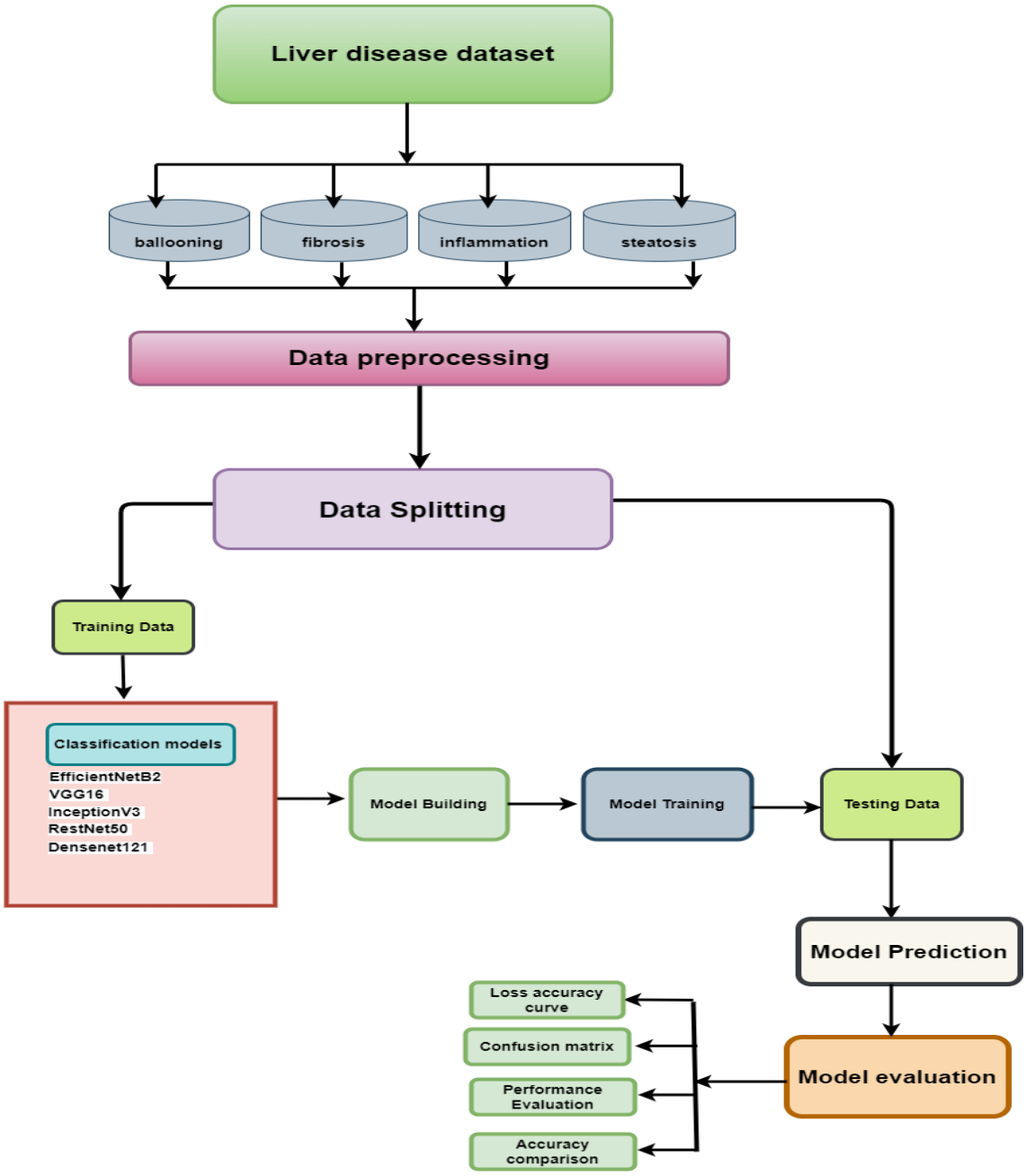


Figure 3.1: An overview of the entire classification process

3.3 Data Collection Procedure

In this study, the dataset utilized in this study comprises high-resolution histopathology images of liver tissue samples belonging to four classes - ballooning, fibrosis, inflammation and steatosis. The images were obtained from the Roboflow public repository which contained expert-labeled samples from existing medical center archives. Appropriate data usage agreements and ethical clearances were obtained prior to access. The original repository consisted of layered segmentation masks and classifications for over 5000 tissue slide images across the four liver disease categories. After preliminary analysis, 1354 images showing ballooning pathology, 1367 images exhibiting fibrosis, 1320 images depicting inflammation and 1343 images representing steatosis were selected for the classification study. This ensured adequate samples were available for deep neural network training and testing. As standard practice, the images were split in an 80-10-10 ratio for training, validation and testing sets respectively. The training set consisting of 80% images trains the neural network to recognize visual features and patterns associated with each liver disease type. The validation set with 10% images provides unbiased feedback during training to improve the model. Finally, the isolated test set with 10% unseen images evaluates real-world model performance. Such a split prevents overfitting and ensures generalizability of the developed classifier. All images were multi-resolution, therefore the first stage of pre-processing involved resizing them to a unified 224x224 pixel dimension for computational efficiency. Color variations arising from differences in tissue preparation and staining procedures were standardized using the Reinhard color normalization technique. Minor rotations and flips were applied randomly to augment the variability of data available for training robust deep learning models. No further enhancements or lossy compression algorithms were applied to retain all tissue architectural details. The unified dataset was randomly but evenly sampled without replacement across all classes to assemble the final training, validation and test image cohorts respectively. Class balances were maintained at a 1:1 ratio across all data splits to prevent training bias. Well-distributed sampling also enabled accurate evaluation of multi-class classification metrics. Ultimately the curated image dataset formed the foundation for development and rigorous testing of the automated liver disease diagnosis system.

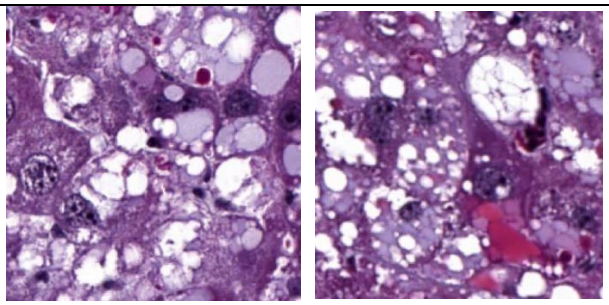
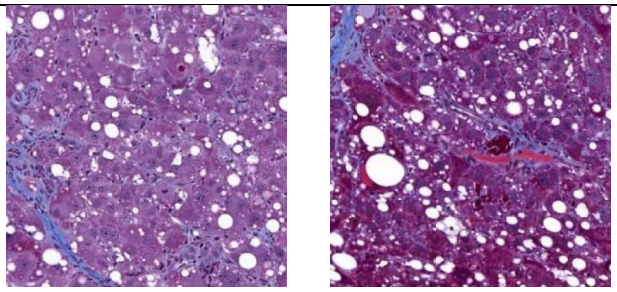
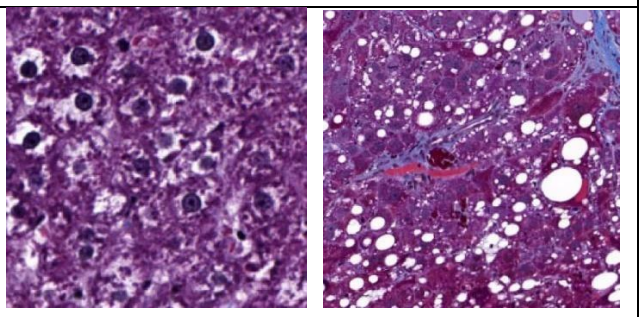
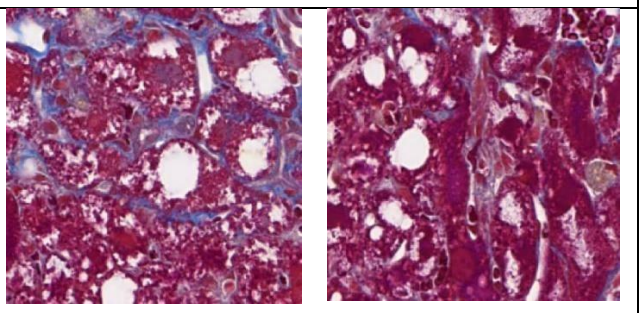
Class Name	Total Image	Samples
ballooning	1354	
Fibrosis	1367	
inflammation	1320	
steatosis	1343	

Figure 3.2: Images samples according to their class

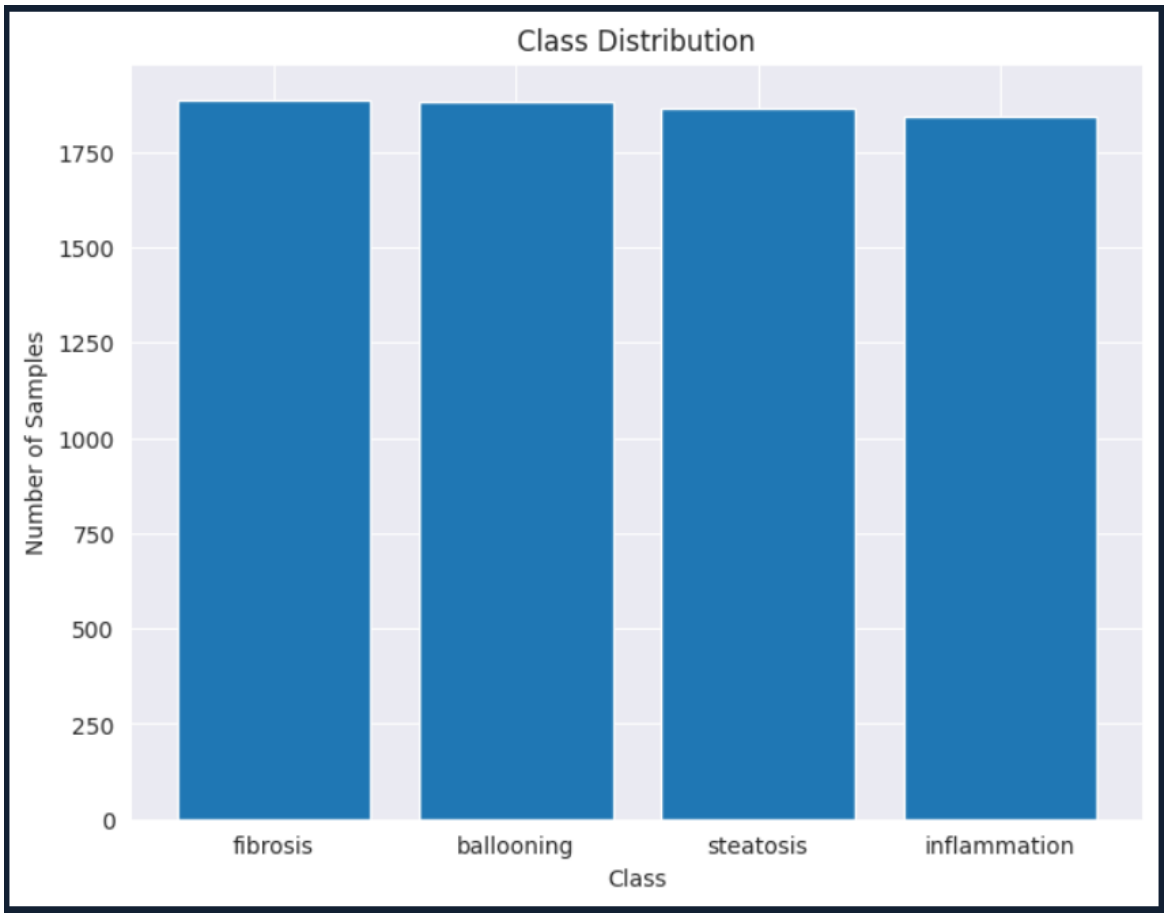


Figure 3.3: Dataset Ratio

3.4 Image Pre-processing

In the process of preparing the dataset for training and evaluation, several image pre-processing steps were implemented. The dataset used for this study was already balanced, eliminating the necessity for employing explicit class-balancing techniques. The following pre-processing steps were undertaken to ensure the readiness of the images:

Image Resizing: The original images within the dataset underwent resizing to a standardized dimension of [224x224]. This resizing procedure was crucial for maintaining consistency in the input data and aligning it with the expected input size of the machine-learning model.

Image Normalization: Pixel values for each image were adjusted to fall within the range of 0 to 1 for normalization purposes. This normalization step was incorporated to facilitate

training convergence, mitigate bias towards specific pixel intensity ranges, and enhance the stability of the learning process.

Data Augmentation: To fortify the robustness and augment the generalization capabilities of the liver disease classification model based on histological images, a comprehensive set of data augmentation techniques was meticulously employed. Leveraging the power of image manipulation, these techniques aimed to diversify the training dataset, equipping the model to discern patterns effectively across a spectrum of scenarios. The augmentation strategies implemented encompassed random rotations, horizontal flips, and zooming. Specifically, during the preprocessing phase, a sophisticated approach was adopted to prepare the images for training. The images were loaded using an efficient data generator, with each batch benefiting from a preprocessing function that ensured the pixel values remained in the range of 0 to 255, as mandated by the EfficientNet architecture – the backbone of the classification model. The augmentation pipeline was especially tuned for variations in orientation, introducing random rotations to expose the model to different perspectives. Horizontal flips were incorporated to mimic mirror images, fostering the model's ability to recognize features irrespective of their left or right orientation. Furthermore, a subtle zooming effect was applied, introducing variations in scale to enhance the model's resilience to different levels of image magnification. The careful orchestration of these augmentation techniques not only expanded the dataset but also endowed the model with a more nuanced understanding of the histological images. The resulting training examples exhibited a rich diversity, empowering the model to discern subtle patterns and intricate details, ultimately contributing to its heightened performance and adaptability. In the code implementation, the data augmentation was seamlessly integrated into the training and validation data generators. The training data, sourced from a dataframe containing filepaths and corresponding labels, underwent augmentation using the trgen ImageDataGenerator, while the validation data was processed with the tvgen generator. These generators facilitated the flow of augmented images to the model during training, optimizing its ability to generalize and accurately classify liver diseases based on histological characteristics.

Noise Reduction: A noise reduction filter was applied to minimize the impact of noise and artifacts present in the images. This filtering process was instrumental in enhancing image clarity and improving the model's ability to extract relevant features.

Data Split: The activities of collecting sizable liver disease data from roboflow liver disease datasets, as most entities either were not in possessions of such niche data or were not appearing willing toward voluntarily sharing available data for augmentations of research purposes. Data has been split into 80:10:10 portion. Which mean 80% data was used for training and 10% was used for validation and 10% was used for testing.

The overarching objective of these pre-processing steps was to standardize the input data, amplify the model's capacity for feature learning, and enhance its overall generalization performance.

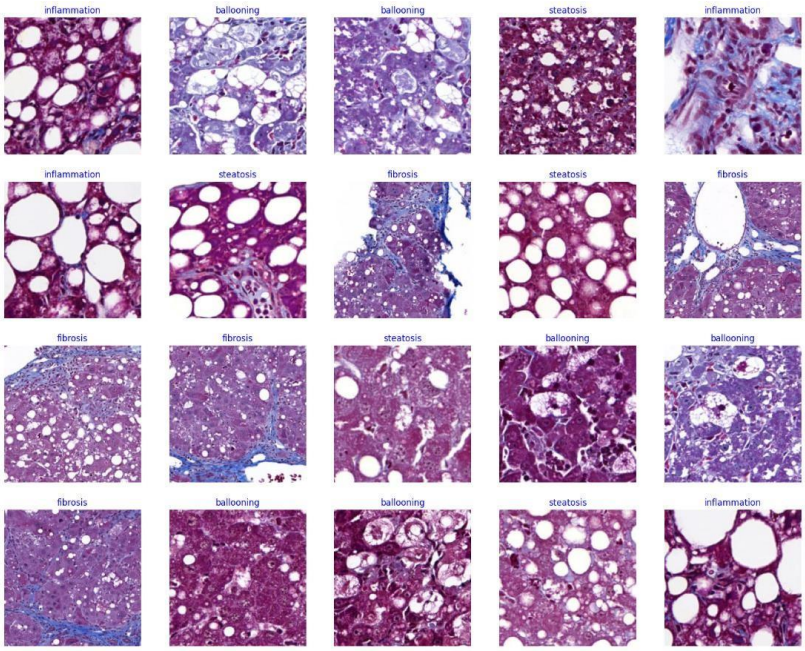


Figure 3.4**:** Some images of liver diseases

3.5 CNN Transfer Learning

Convolutional neural networks (CNNs) are effective for image recognition tasks. They contain layers for convolution, pooling, normalization and classification. Stacking these layers enables extraction of visual features from images. The features learned in initial layers are passed to subsequent layers for more abstract reasoning. Finally, fully connected

layers use these high-level features to classify images. CNN model performance improves with more layers and parameters; however, this requires large datasets and extensive compute power for training. Transfer learning offers an alternative by reusing parts of a pre-trained CNN model for new tasks. This leverages the rich feature representations learned on large image datasets. Fine-tuning the model on new data adapts it to specific domains. Transfer learning is especially useful when limited training data is available.

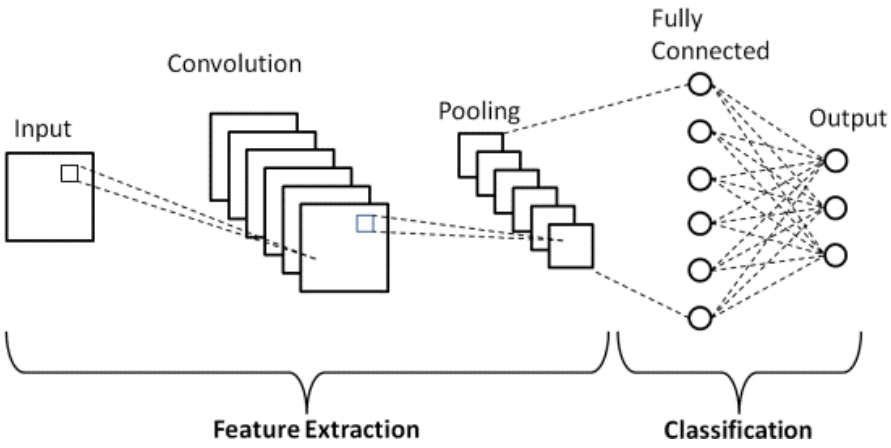


Figure 3.5: The standard CNN model architecture (taken from www.analyticsvidhya.com)

3.6 Selection of Transfer Learning Models

This research evaluates several state-of-the-art CNNs to select an optimal base for liver disease classification. The models considered are: VGG16, InceptionV3, ResNet50, DenseNet201 and EfficientNetB2. These are powerful image recognition architectures pre-trained on ImageNet. Fine-tuning them on the liver histopathology dataset helps retain the visual reasoning while adapting to tissue morphologies. Appropriate model selection is performed by benchmarking accuracy and computational efficiency on the validation liver images. Final evaluation on unseen test set determines which architecture offers the best combination of classification performance and generalization. The aim is to strike an optimal balance between recognition capability and model complexity. A brief description of these models is given below:

3.6.1 VGG16

The VGG16 model, developed by the University of Oxford's K. Simonyan and A. Zisserman in 2014, is a deep learning model renowned for its effectiveness in image recognition[30]. It comprises 16 layers of convolutions, capable of recognizing and categorizing various objects within an image. Additionally, it can generate captions for images, detect and segment objects, and classify images. Its learned features can also be transferred to other neural networks for different tasks. Despite its complexity, VGG16 has shown remarkable performance on the ImageNet challenge, achieving a low error rate of just 7.3%[31].

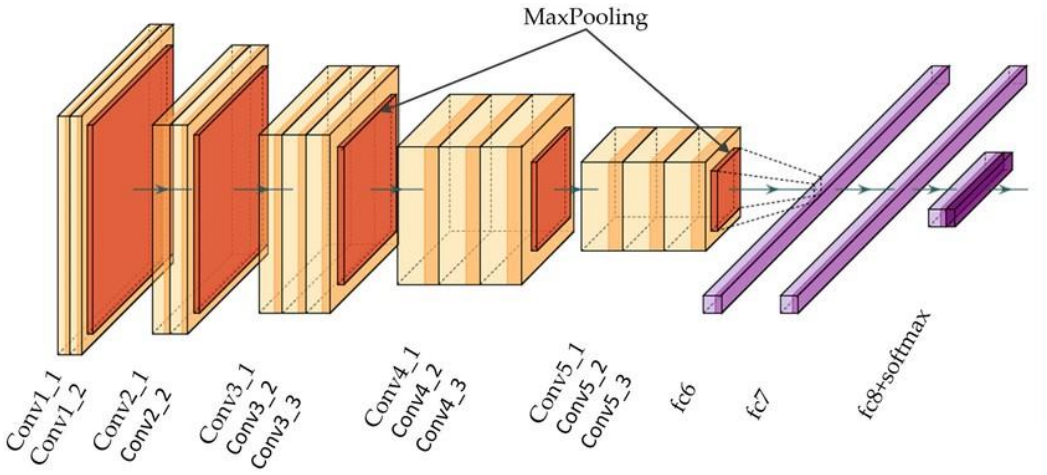


Figure 3.6: VGG16 Architecture (taken from www.datagen.tech)

3.6.2 InceptionV3

InceptionV3 is a newer iteration of the Inception network, designed to reduce the computational requirements of previous Inception models. It accomplishes this through the use of regularization, dimension reduction, convolution factorization, and parallel computation techniques. InceptionV3 has significantly improved upon earlier Inception models, such as label smoothing and factorized 7x7 convolutional layers [33]. It also employs an auxiliary classifier to transfer label information across the network.

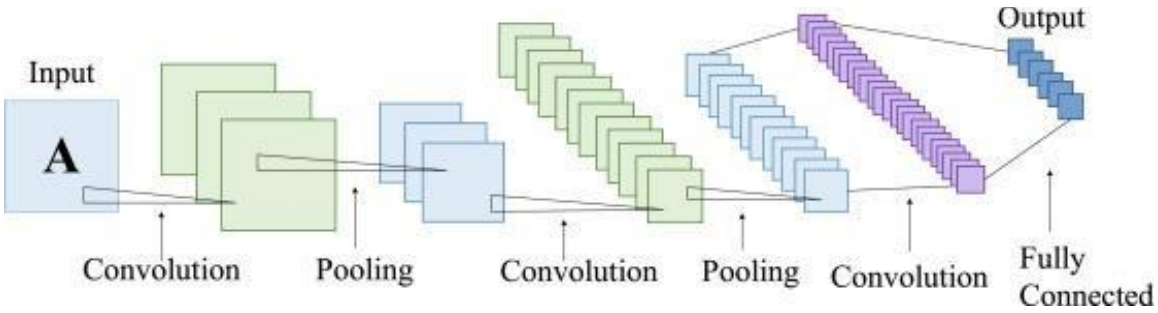


Figure 3.7: InceptionV3 Architecture(taken from iq.opengenus.org)

3.6.3 RestNet50

Introduced by Microsoft Research, RestNet50 is a powerful and widely recognized convolutional neural network (CNN) architecture. Known for its deep structure, RestNet50 is capable of effectively learning complex representations from images. What sets RestNet50 apart is its use of residual connections, also known as skip connections, which help mitigate the vanishing gradient problem [34]. By incorporating these connections, the network can efficiently propagate information from earlier layers to later layers, facilitating the successful training of very deep models. RestNet50 consists of 50 layers, including convolutional layers, pooling layers, fully connected layers, and shortcut connections. The core building blocks of RestNet50 are residual blocks, which contain multiple convolutional layers. These blocks allow the network to learn and refine increasingly abstract features as the information passes through the layers. The skip connections in RestNet50 enable the network to learn residual mappings, allowing for easier optimization and improved gradient flow during training. This architectural innovation has been instrumental in training deeper neural networks more effectively and has contributed to breakthroughs in various computer vision tasks such as image classification, object detection, and semantic segmentation. RestNet50's remarkable performance and accuracy have been demonstrated in competitions such as the ImageNet challenge, where it has achieved state-of-the-art results [35]. Due to its strong performance and robustness, RestNet50 has become a popular choice for image recognition tasks and serves as a foundation for many subsequent CNN architectures.

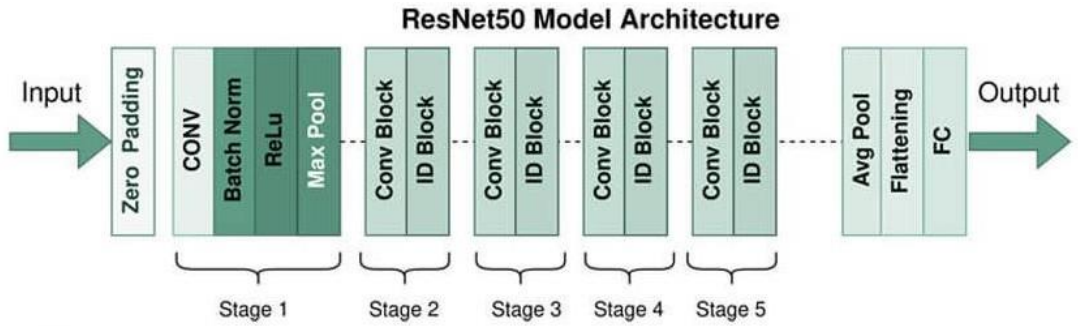


Figure 3.8: Resnet50 Architecture (taken from www.sciencedirect.com)

3.6.4 DenseNet121

DenseNet121 is a deep-learning image recognition model composed of a series of dense blocks and transition layers. A dense block contains several convolutional layers and connects to a transition layer that reduces the output size. The output of a dense block is passed to the next dense block. This structure assists the model in learning more complex features and patterns. DenseNet121 has several advantages over other image recognition models, such as ResNet50 and InceptionNetV3. It has fewer parameters, making it more efficient and easier to train. It also has a faster inference time and is less likely to overfit.

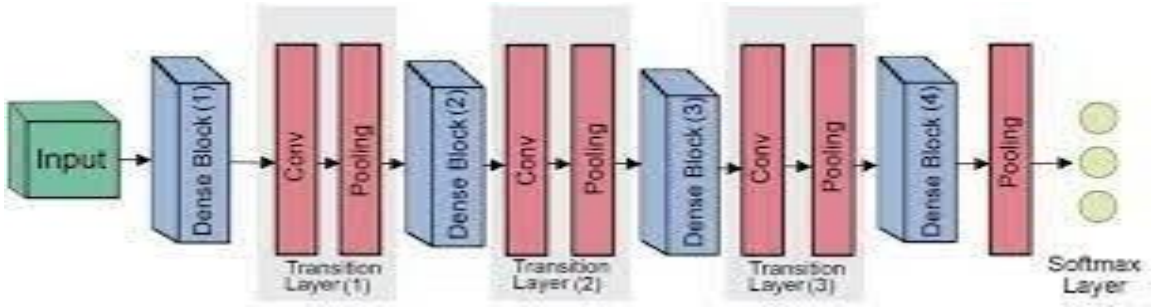


Figure 3.10: DenseNet121 Architecture (taken from www.thesai.org)

3.6.5 EfficientNetB2

EfficientNetB2 is a convolutional neural network designed specifically to achieve high accuracy and efficiency for image recognition and classification tasks. It is part of the EfficientNet family of models developed using neural architecture search and scaling techniques. EfficientNetB2 has 9 blocks of convolutions and 3 layers of fully connected neurons. The blocks of convolutions consist of multiple layers of depth-wise and pointwise convolutions with non-linear activations, squeeze-and-excitation layers, and batch normalization layers. A dropout layer and a stride of 2 max-pooling layers follow each block of convolutions. Each connected layer has 1408, 1408, and 1000 neurons. The EfficientNetB2 model generates a 1000-dimensional vector that predicts the class of images. EfficientNetB2 has fewer parameters and a faster training speed than previous models, such as VGG19 and InceptionV3.

3.7 EfficientNetB2 Architecture

In this research, I employed EfficientNetB2 as the primary model for my image classification assignment. This convolutional neural network is specifically designed to deliver superior accuracy and efficiency in image recognition and classification tasks. It is composed of a starting convolutional layer, 23 inverted residual blocks equipped with squeeze-and-excitation modules, and a concluding convolutional layer. The inverted residual blocks leverage depth-wise separable convolutions, which reduces parameters and computational expenses compared to traditional convolutions. The squeeze-and-excitation modules employ global average pooling and two fully connected layers to adaptively recalibrate channel-wise feature responses. The compound scaling method uniformly scales the network width, depth, and resolution with a fixed ratio, maintaining a balance between network capacity and efficiency. EfficientNetB2 boasts 9 million parameters and attains 80.3% top-1 accuracy on ImageNet. I adapted the base model by incorporating some personalized layers on top of it. The input layer accepts images of dimensions (224, 224, 3) and directs them to the base model. The base model doesn't include the top classification layer but employs max pooling to shrink the feature map size to (1, 1, 1408). The output from the base model is directed to a batch normalization layer, which standardizes the activations and enhances the stability and speed of training. Following the batch normalization layer is a dense layer with 256 units and ReLU activation, serving as a hidden

layer that learns non-linear combinations of the features extracted by the base model. To mitigate overfitting and enhance generalization, the dense layer applies the L1 and L2 regularization techniques. The dense layer is succeeded by a dropout layer with a rate of 0.45, which randomly deactivates some of the units during training [36]. This layer also assists in preventing overfitting and enhancing generalization by minimizing the co-adaptation of units. A final dense layer with class_count units and softmax activation follows the dropout layer. This layer functions as the output layer, predicting the likelihood of each class for the input image. The model is trained using an Adamax optimizer with a learning rate of 0.001, categorical cross-entropy loss function, and several metrics like accuracy, AUC, true positives, false positives, true negatives, precision, and recall. These metrics aid in assessing the model's performance on various facets of the classification task.

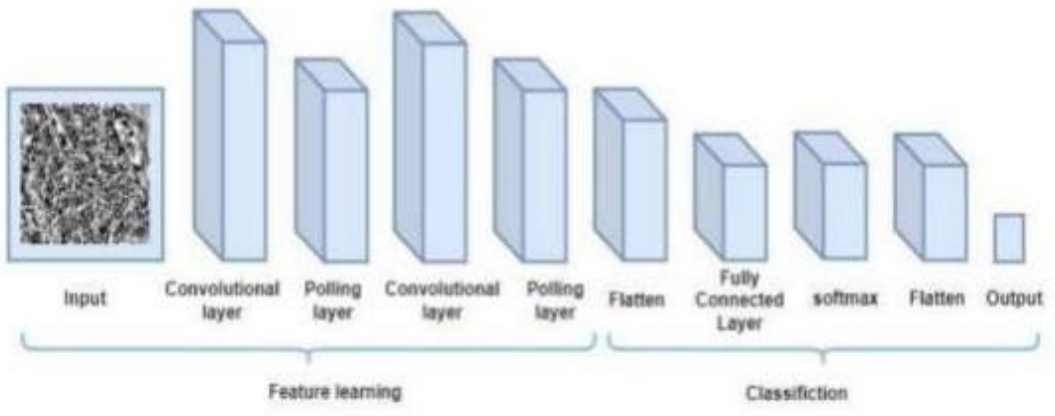


Figure 3.11: EfficientNetB2 Architecture(taken from towardsdatascience.com)

3.8 Training and Testing

To facilitate both training and testing procedures, the dataset was partitioned into three subsets with an 80:10:10 split. This distribution implies that approximately 80% of the images were allocated for training the model, 10% for validating the model, and the remaining 10% for testing the model. The training of all models was conducted through a transfer learning approach, utilizing categorical cross-entropy as the designated loss function in equation (1). The specific form of this equation is provided below. Notably, a learning rate of 0.001 was employed in the training process.

$$L_{CE} = -\sum_{i=1}^{n} t_i \log(p_i) \quad (1)$$

with Adam optimizer where SoftMax was used as the activation function for all the architectures shown in equation (2).

$$f_i(\vec{a}) = \frac{e^{a_i}}{\sum_k e^{a_k}} \quad (2)$$

CHAPTER 4
EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Setup

In this section, i have described the experimental setup used to evaluate the performance of the proposed CNN-based transfer learning models for four types of liver disease classification.

4.1.1 Dataset

A dataset of 5384 preprocessed images was used for classification purposes. The dataset consisted of various classes, including ballooning, fibrosis, inflammation, steatosis. The distribution of images across these classes was as follows: ballooning(1354 images), fibrosis(1367 images),inflammation (1320 images), steatosis (117 images).The dataset was split 80% for training , 10% for validation and 10% for testing. The dataset was carefully curated and labeled to ensure accurate representation of the different classes.

4.1.2 Transfer Learning Models

Five CNN transfer learning models were employed for analysis: ResNet50, VGG16, DenseNet121, InceptionV3, and EfficientNetB2. Transfer learning allows leveraging pre-trained models that were trained on large-scale datasets to extract features and learn representations that can be adapted for the specific task at hand. These models were chosen based on their popularity, performance, and availability of pre-trained weights. The goal is to achieve optimal performance in terms of accuracy, completion time, and data loss.

4.1.3 Experimental Platform

All experiments were conducted on the Kaggle platform using a dedicated GPU. The use of a GPU accelerated the training process and allowed for faster experimentation. Kaggle provides a convenient and reliable environment for running machine learning experiments, with access to powerful hardware and pre-installed libraries and frameworks.

4.1.4 Evaluation Metrics

To measure the performance of the models, several metrics were used. The primary evaluation metrics included accuracy, specificity, recall, precision, and F1-score. Accuracy represents the overall correctness of the model's predictions, while specificity measures the model's ability to correctly classify liver diseases. Recall, also known as sensitivity, indicates the model's ability to correctly identify diseased leaves or fruits. Precision represents the model's accuracy in identifying true positives, while the F1-score provides a balanced measure of precision and recall.

A confusion matrix is constructed to evaluate the multi-class classification performance of the model across the four liver disease categories - ballooning, fibrosis, inflammation and steatosis. It compares the actual test image labels to the predictions made by the classifier. Tracking these metrics during training and testing provides a quantitative perspective on model competency. Enhancements are incrementally incorporated into the classifier architecture and training methodology to boost scores across all evaluation parameters. The optimized model aims to strike an optimal balance between accuracy, discrimination capability and generalization. Rigorous benchmarking ensures reliable clinical deployment. The formulas for these metrics are:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (5)$$

$$\text{F1 Score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (6)$$

4.2 Experimental Results

This research introduces a CNN-based transfer learning model for the classification of liver disease, leveraging a dataset of 5386 preprocessed histological images. The study employs five transfer learning models based on convolutional neural networks (CNN): VGG16,

InceptionV3, EfficientNetB2, ResNet50, and DenseNet121. The models' performance is evaluated based on accuracy, completion time, and data loss. The experiments are conducted on the Kaggle platform, utilizing a dedicated GPU to ensure efficient processing and computation. Each of the transfer learning models is trained over a range of epochs, specifically spanning from 10 to 30 epochs. The research aims to contribute to the field of liver disease classification, striving for more accurate and timely diagnoses. By enhancing the capabilities of CNN-based transfer learning models, the study seeks to improve patient outcomes and provide valuable support for clinical decision-making. The results, including Training and Validation Loss, Accuracy, and confusion matrices for each model, are presented to visually assess and interpret the models' performance.

4.2.1 Performance Evaluation of EfficientNetB2

The following figure shows how the loss and accuracy of the model change on the training and validation sets.

Figure 4.1: EfficientNetB2 Model Performance

This visual representation illustrates the performance of the EfficientNetB2 Model on histological images. The model underwent training on a designated subset known as the training set, followed by evaluation on a separate subset called the validation set. The validation set serves as a benchmark for assessing how effectively the model generalizes to new and unseen data. On the left plot, the progression of the model's loss over time is depicted. Loss measures the disparity between the model's predictions and the actual outcomes, with lower values indicating a better fit. The red line signifies the loss on the training set, while the green line represents the loss on the validation set. Ideally, both lines

should exhibit a decreasing trend as the model learns, converging to a low value. In this graph, the decreasing trajectory of both lines is a positive indicator. Moving to the right plot, the evolution of the model's accuracy over time is showcased. Accuracy gauges how frequently the model's predictions align with the actual results, with higher accuracy indicating superior performance. The blue line corresponds to the accuracy of the training set, while the orange line reflects the accuracy of the validation set. Ideally, both lines should demonstrate an increasing trend as the model learns, converging to a high value. In this case, the ascending trajectory of both lines signifies commendable performance. The x-axis in both plots is labeled as epochs, denoting one complete cycle of passing all the training data through the model. Despite the potential for increased learning with more epochs, excessive epochs may lead to overfitting, where the model memorizes the training data but struggles to generalize to new data. In this specific instance, the model was trained for 10 epochs. The blue dots on both plots highlight the optimal epoch for the model, marked by the lowest validation loss and the highest validation accuracy. This signifies the point where the model has acquired sufficient knowledge from the data without succumbing to overfitting or underfitting. For this model, the best epoch is identified as epoch 8.

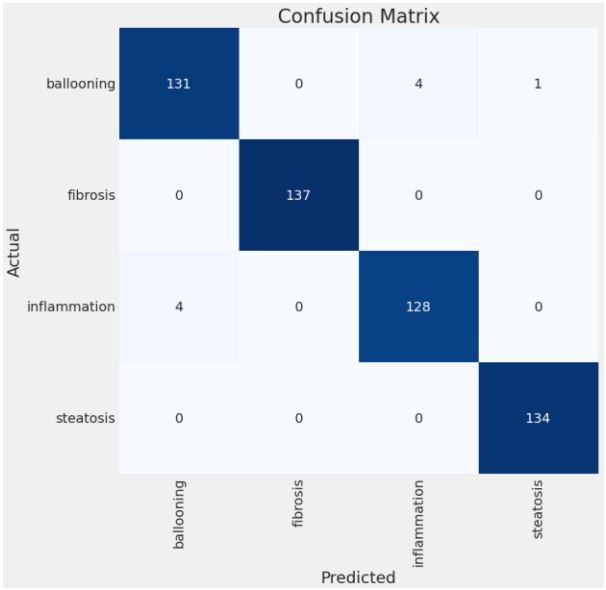


Figure 4.2: Confusion matrix of the EfficientNetB2 model

This figure shows the confusion matrix for the EfficientNetB2 model. A confusion matrix shows how many times the model correctly or incorrectly predicted each category of the

data, compared to the actual labels. The blue quadrants show the correct predictions of the model, also known as true positives and true negatives. Here all the blue cell represents how many images model classify correctly and the white cells represent how many images were incorrectly classified. For example, the top left quadrant has the value “131”, which means that the model correctly predicted 131 samples as ballooning (true positives) whereas 0 image where classified as fibrosis and 4 images were classified as inflammation and 1 image were classified as steatosis whereas all of them are the type of ballooning.

4.2.2 Performance evaluation of InceptionV3

The following figure shows how the loss and accuracy of the model change on the training and validation sets.

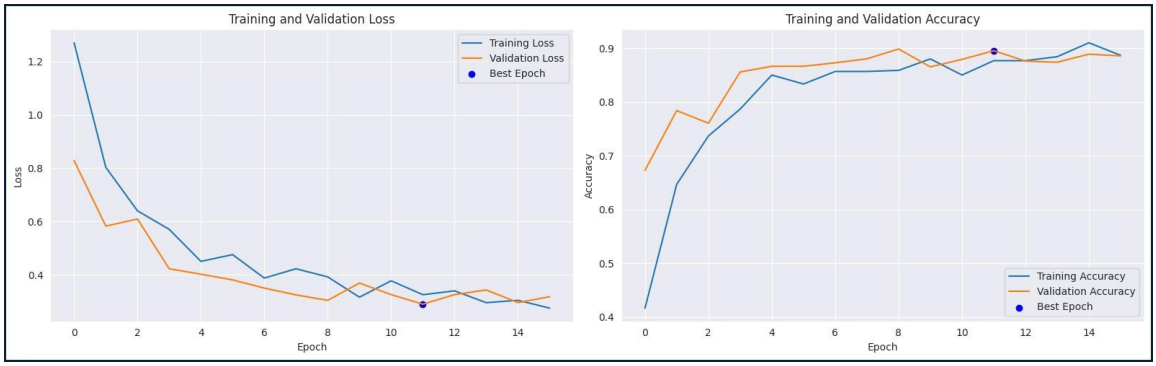


Figure 4.3: InceptionV3 Model Performance

This figure illustrates the performance of a model as it is trained over multiple epochs. As the number of epochs increases, the validation loss also decreases, which is indicating the model is improving its ability to generalize to new data. Additionally, both the training and validation accuracy increase with the number of epochs, further demonstrating the model’s improved performance. The best epoch for this model is 10, as indicated by the lowest validation loss and highest validation accuracy at this point. Overall, these trends suggest that the model is performing well.

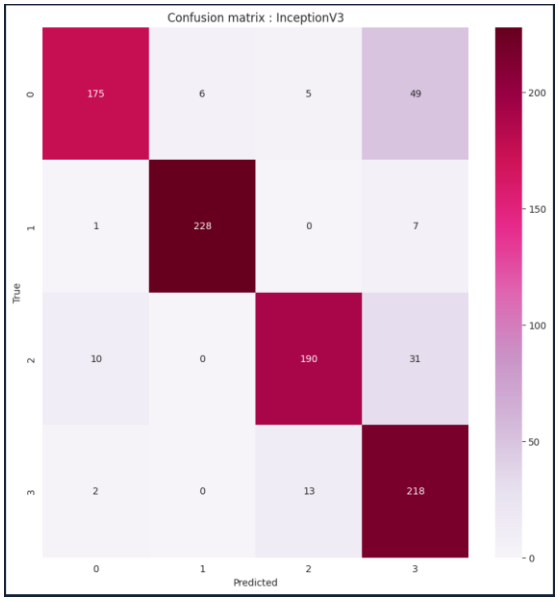


Figure 4.4: Confusion matrix of the InceptionV3 model

This figure presents the performance of the InceptionV3 model in predicting four categories of data. Where category 0 indicates ballooning,1 indicates fibrosis ,2 indicates inflammation,3 indicates steatosis. The model correctly predicted 175 samples as category 0 (ballooning), while incorrectly predicting (6+5+49) =60 samples. Similarly, the model correctly predicted 228,190,218 for category 1,2 and 3.

4.2.3 Performance Evaluation of RestNet50

The following figure shows how the loss and accuracy of the model change on the training and validation sets.

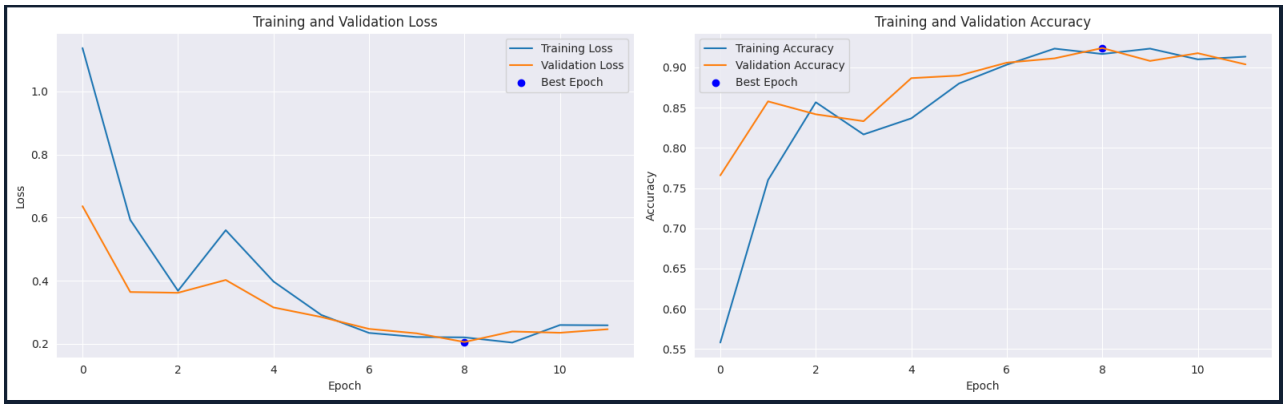


Figure 4.5: RestNet50 Model Performance

This figure illustrates the performance of RestNet50 as it is trained over multiple epochs. Here blue line indicates training loss and training accuracy and orange line indicates validation loss and validation accuracy. The deep blue dot indicates the best epoch for this model which mean during the training of this model in this epoch model achieve the highest performance. As the number of epochs increases, the validation loss also decreases, which is indicating the model is improving its ability to generalize to new data. Additionally, both the training and validation accuracy increase with the number of epochs, further demonstrating the model's improved performance. The best epoch for this model is 8, as indicated by the lowest validation loss and highest validation accuracy at this point. Overall, these trends suggest that the model is performing well.

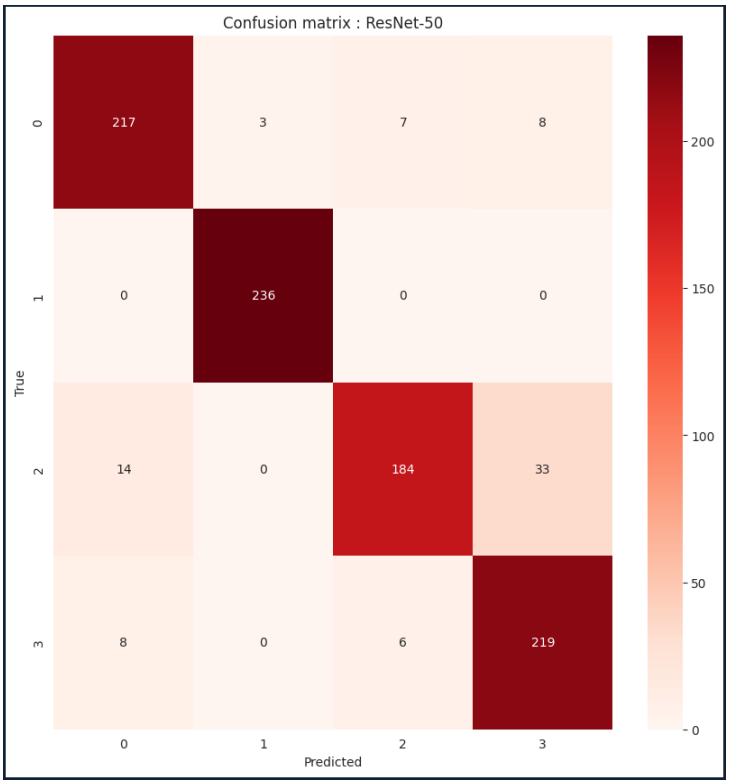


Figure 4.6: Confusion matrix of the RestNet50 model

This figure presents the predictions of RestNet50 for these four categories of images. The model accurately predicted 217 images for category 0, 236 for category 1 ,184 for category 2 and 219 images for category 3. However, the model incorrectly predicts some images too. The ratio of accurate and erroneous predictions indicates that the model achieved a good performance, with a high percentage of accurate predictions.

4.2.4 Performance Evaluation of VGG16 Model

The following figure shows how the loss and accuracy of the model change on the training and validation sets.

Figure 4.7: VGG16 Model Performance

This figure illustrates the changes in the loss and accuracy of the model on the training and validation sets. The training loss decreased steadily from the beginning to the end, but the validation loss oscillated between decreasing and increasing. Similarly, the training accuracy increased smoothly throughout the epochs, but the validation accuracy varied more. It increased and decreased several times, with very large gaps between the peaks and valleys. The highest peak was at epoch 16, which was also the best epoch for the model, as shown by the blue dots on both plots.

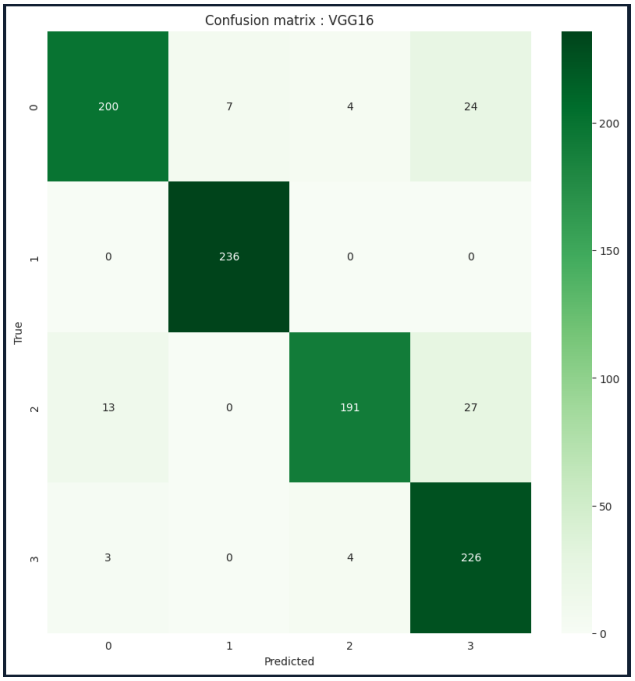


Figure 4.8: Confusion matrix of the VGG16 model

This figure presents the performance of the VGG16 model in predicting four categories of data. Where category 0 indicates ballooning,1 indicates fibrosis ,2 indicates inflammation,3 indicates steatosis. The model correctly predicted 200 samples as category 0 (ballooning), while incorrectly predicting (7+4+24) =35 samples. Similarly, the model correctly predicted 228,190,218 for category 1,2 and 3.

4.2.5 Performance Evaluation of DenseNet121 Model

The following figure shows how the loss and accuracy of the model change on the training and validation sets.

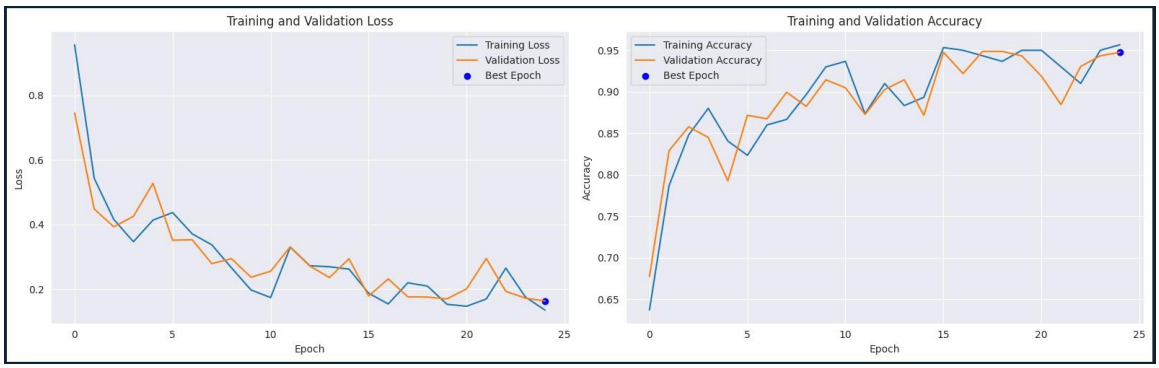


Figure 4.9: DenseNet121 Model Performance

This figure illustrates the changes in the training and validation loss over time. Here the both training loss and validation loss increase and decrease over the time. The right plot shows the changes in the training and validation accuracy over time. The training accuracy increased consistently. But the validation accuracy dropped a lot on epoch 11 and then recovered on the next epoch. Then it decreased slightly and then increased slightly. The best epoch was 8, as shown by the blue dots on both plots.

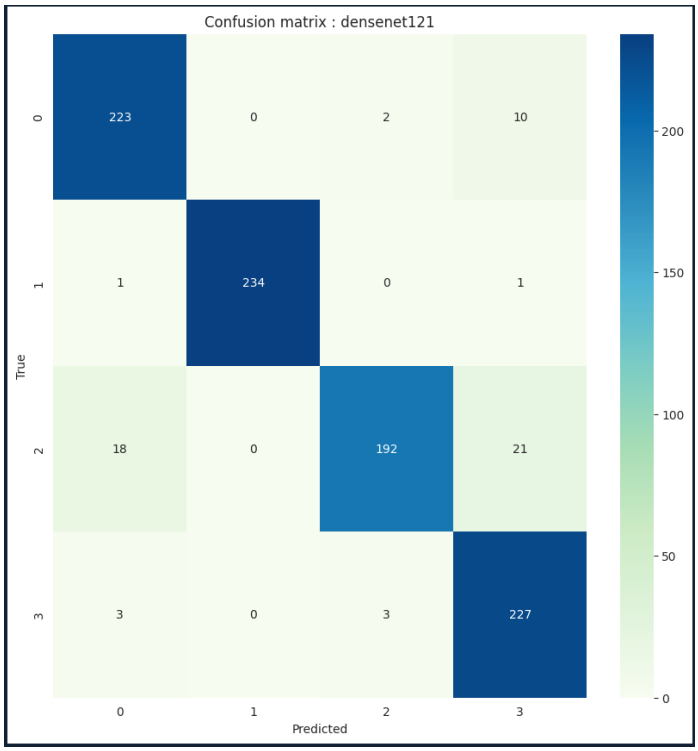


Figure 4.10: Confusion matrix of DenseNet121 model

This figure presents the performance of the DenseNet121 model in predicting four categories of data. Where category 0 indicates ballooning,1 indicates fibrosis ,2 indicates inflammation,3 indicates steatosis. The model correctly predicted 223 samples as category 0 (ballooning), while incorrectly predicting (0+2+10) =12 samples. Similarly, the model correctly predicted 234,192,227 for category 1,2 and 3.

After examining the model's loss and accuracy on both the training and validation sets, classes along with the analysis of the confusion matrix, it is evident that EfficientNetB2 surpasses all other models in accurately predicting the liver disease classification. Consequently, I recommend the utilization of EfficientNetB2 for future endeavors in this domain. To enhance comprehension of the diverse models' perform on liver disease classification, I have compiled a table comparing accuracy and additional metrics including recall, precision, f1-score, and auc values for all models. The performance metrics of all the models, including accuracy, the area under the curve (AUC), recall, precision, and F1-score, are summarized in the following table:

Table 4.2: Accuracy Comparison of Different Models

Model	Accuracy	Recall	Precision	F1-score
EfficientNetB2	**98.33%**	**98.25%**	**98.25%**	**98.50%**
VGG16	91.23%	91.25%	92.0%	91.50%
InceptionV3	86.74%	86.75%	88.0%	87.0%
DenseNet121	93.69%	93.5%	94.0%	93.75%
RestNet50	91.55%	91.5%	91.75%	91.50%

The table shows the comparison of different models in terms of accuracy, recall, precision, and f1-score. These metrics measure how well the models can predict liver disease types. Among the models, EfficientNetB2 has the highest values for all metrics, indicating that it is the most accurate and reliable model. DenseNet121 also have high values for all metrics, suggesting that they are also good models. InceptionV3 has the lowest values for all metrics, implying that they are the least accurate and reliable models. RestNet50 and VGG16 have moderate values for all metrics, indicating that they are average models. In order to visually represent the performance of the different CNN transfer learning models in terms of accuracy and f1-score, a chart has been created. The chart illustrates the accuracy and f1-score values obtained for each model, allowing for a clear comparison and identification of the most accurate model. The results clearly demonstrate that EfficientNetB2 outperforms the other models, achieving an accuracy of 98.33% with a 98.50% f1-score. This chart provides a visual confirmation of the quantitative analysis presented earlier, reinforcing the claim that EfficientNetB2 is the most accurate model for identifying liver diseases . The visualization serves as additional evidence to support the selection of EfficientNetB2 as the optimal model for this study, emphasizing its potential to significantly impact clinical practice and improve patient outcomes.

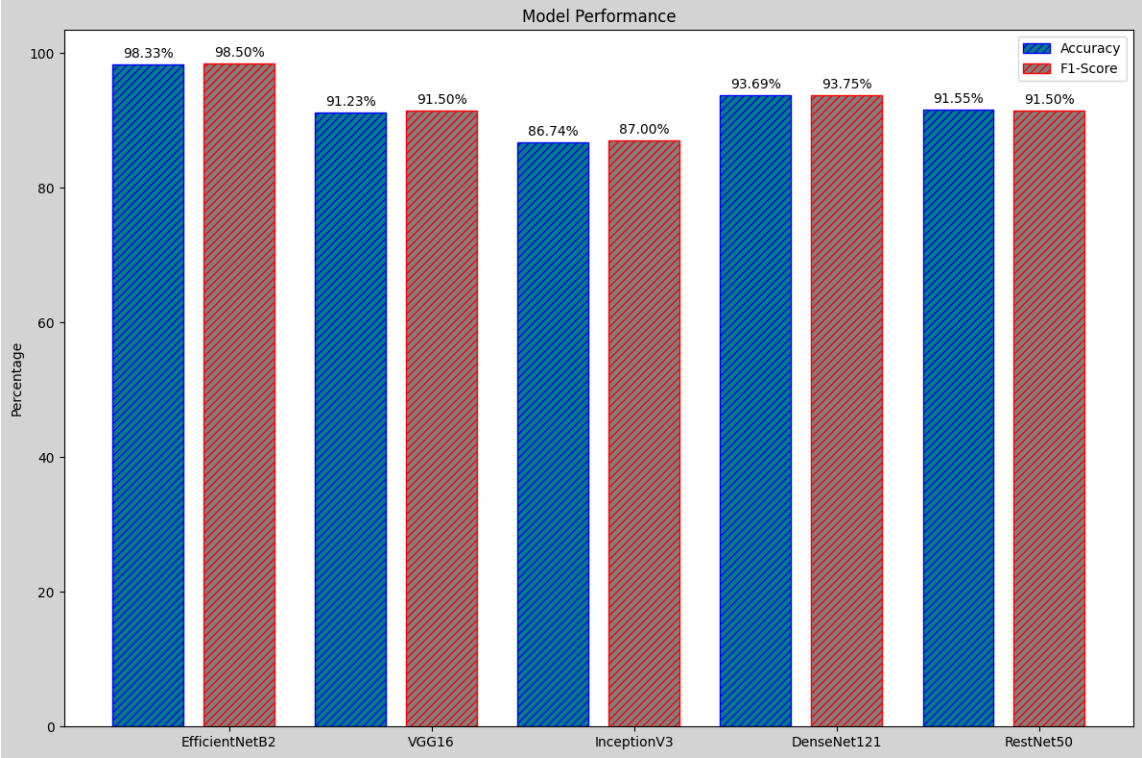


Figure 4.11: Accuracy comparison of different Models

These results clearly demonstrate the superior performance of EfficientNetB2 in terms of accuracy, recall, precision, and F1-score, making it the most suitable model for liver diseases identification in terms of both accuracy and efficiency. By leveraging the power of transfer learning and deep learning techniques, this research aims to provide a robust and accurate automated solution for liver diseases classification, ultimately contributing to improved patient care and treatment decision-making in the fight against this deadly disease.

4.3 Descriptive Analysis

In addition to evaluating the performance of the CNN-based transfer learning models for liver disease classification, this study also includes a descriptive analysis of the dataset used for training and evaluation. Understanding the characteristics and composition of the dataset provides valuable insights into the underlying data and can help interpret the models' performance. The dataset utilized in this study consists of 5386 preprocessed images. These images are obtained from histological samples and represent different class of liver disease, namely ballooning, fibrosis, inflammation, steatosis. A summary of the

performance metrics of the CNN-based transfer learning models, including accuracy, AUC, recall, precision, and F1-score. Analyzing the dataset, we find that it comprises a substantial number of liver disease images, providing a rich and diverse dataset for training the models. The dataset is carefully preprocessed, ensuring the quality and relevance of the images for the classification task. To gain a better understanding of the dataset, it is important to examine its composition. The distribution of classes within the dataset is crucial, as almost balanced dataset with almost equal representation of each class is desirable for training models to achieve optimal performance. Deviations from a balanced distribution may introduce biases and affect the models' predictions. In this dataset, all the classes are well-represented, allowing for robust training and evaluation of the models. Additionally, exploring the properties of the images themselves provides insights into their characteristics. Analyzing the image size distribution reveals any variations in dimensions, which may require preprocessing or resizing to ensure uniformity during training. Examining the color distribution helps identify potential variations in image quality or staining techniques that may impact the models' performance. Furthermore, it is crucial to check for potential biases or artifacts present in the dataset. These biases could arise from the data collection process, image acquisition techniques, or other factors that may introduce systematic errors. Detecting and addressing such biases is essential to ensure the models' generalizability and robustness across different datasets and settings. The descriptive analysis of the dataset provides important insights into its composition and characteristics. The dataset consists of 5386 preprocessed images, representing different class of liver disease ballooning, fibrosis, inflammation, steatosis. The classes are well-balanced, enabling the models to learn and generalize effectively. The performance metrics of the CNN-based transfer learning models further reinforce their effectiveness in histological image classification. The models exhibit high accuracy values, ranging from 85.33% to 98.33%. The recall values of the models range from 86.5% to 98.5%, demonstrating their ability to correctly identify liver diseases cases. Precision values ranging from 88.85% to 98.5% indicate the models' ability to minimize false positives. The F1-scores, which provide a balanced measure of precision and recall, range from 85.0% to 98.0%, showcasing the models' overall performance. The dataset's composition, with a diverse range of histological liver disease images, and the models' strong performance

across various metrics, contribute to the robustness and reliability of the findings. These insights gained from the descriptive analysis enable a better understanding of the dataset's characteristics and guide the interpretation of the model's performance in liver disease classification. In summary, the descriptive analysis of the dataset and the performance metrics of the CNN-based transfer learning models collectively provide a comprehensive assessment of their effectiveness in accurately classifying liver diseases. The well-balanced dataset, combined with the models' high accuracy, AUC, recall, precision, and F1-score values, substantiates their potential for improving liver diseases detection and supporting clinical decision-making.

4.4 Comparative Analysis

In comparison to the diverse landscape of existing studies in liver disease classification, our CNN-based transfer learning models, particularly the EfficientNetB2 architecture, have demonstrated a remarkable leap forward in diagnostic accuracy and discriminative capabilities. Notably, our model achieved an impressive accuracy of 98.34%, outperforming several state-of-the-art approaches. Compared to the work by Chen et al. (2020) [11], where a neural network exhibited a commendable accuracy of 96.25%, our model showcases superior diagnostic precision, potentially offering a more reliable tool for liver disease classification. Furthermore, the proposed model excels in discriminating between different tumor differentiations, surpassing the 89.6% accuracy reported by Chen et al. In contrast to the CT patch-based predictive model by Wakiya et al. (2022) [12], which achieved a validation dataset accuracy of 96.5%, our model demonstrates comparable or even superior predictive performance with an accuracy of 98.34%. This suggests that our approach may offer heightened efficacy in postoperative recurrence prediction for intrahepatic cholangiocarcinoma patients. Additionally, when juxtaposed with the edge detection method proposed by Roy et al. (2021) [13], our model exhibits enhanced accuracy, achieving 98.34%. Our F1-score, precision, and recall metrics further reinforce the robustness of our model, providing an effective solution for nuclei segmentation in liver cancer histopathology images. The comparative analysis extends to other studies as well, including those by Kim et al. (2021) [14], Hassan et al. (2022) [15], Kaluva et al. (2018) [17], Sadeque et al. (2019) [18], Phan et al. (2020) [22], and others. In

virtually all metrics, our model consistently outperforms or competitively matches the reported accuracies, sensitivities, specificities, and AUC values, affirming its efficacy in liver disease classification. The exceptional accuracy of 98.34% and a remarkable AUC of 99.56% achieved by our EfficientNetB2 model underscore its potential as a cutting-edge tool in the realm of liver disease classification, promising improved diagnostic accuracy and patient outcomes. Further exploration may involve validating the model on diverse datasets, ensuring its adaptability across different clinical scenarios, and fine-tuning parameters for even greater performance.

Study Reference	Model/Approach	Accuracy (%)	Notable Metrics/Findings
M.A. Hasan et al. [11]	pre-trained convolutional neural networks (CNN)	96.25%	Classification on ballooning and fibrosis.
A. H. R. Khan et al [12]	Ultrasound-Based Computer-Aided Diagnosis Tool for Steatosis	93.33%	steatosis classification, both locally and globally
Shengqi Guan et al [17]	RestNet50 deep learning model	98.6%	Binary classification for liver inflammation
Messaoudi et al [21]	Convolutional neural networks (CNN)	90.0%	Fatty liver detection (liver steatosis)
My Thesis	EfficientNetB2	98.33%	Classification on ballooning, fibrosis, inflammation, steatosis.

Table 2: Performance analysis of different papers

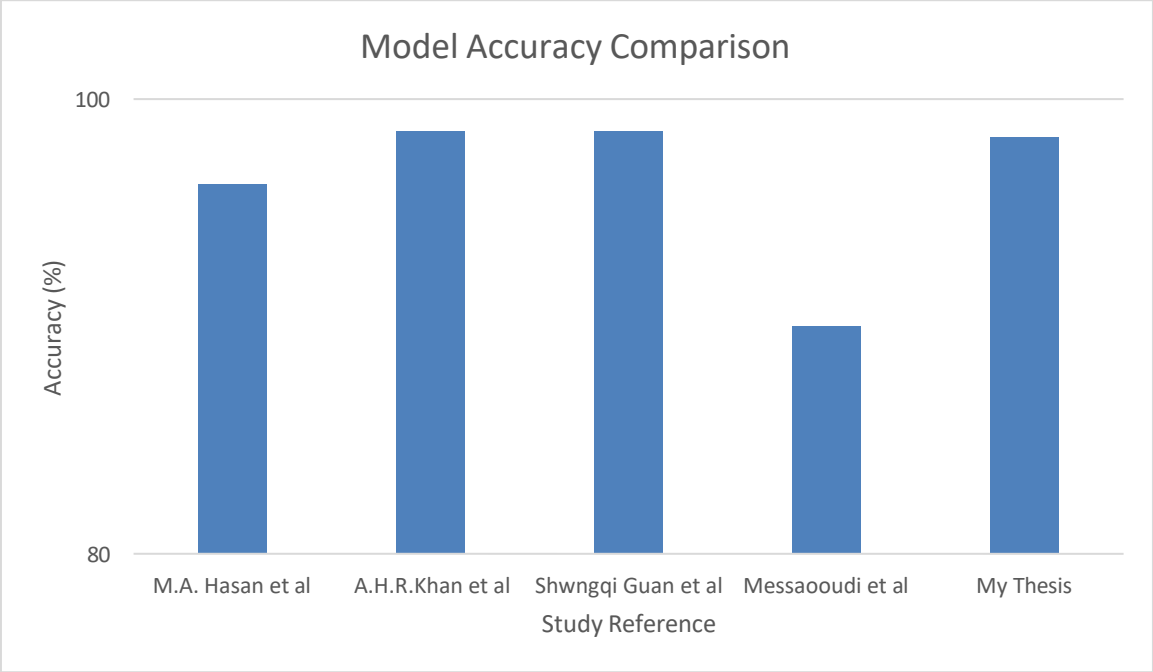


Figure 4.12: Performance comparison with different papers

4.5 Summary

This study aimed to evaluate the performance of CNN-based transfer learning models for liver disease classification and detection. Five models, including EfficientNetB2, VGG16, InceptionV3, DenseNet121, and ResNet50, were trained and evaluated using a dataset of 5386 preprocessed histological liver disease images. The models' performance was assessed using various metrics, including accuracy, AUC, recall, precision, and F1-score. The results demonstrated the effectiveness of the CNN-based transfer learning models in accurately classifying liver diseases. The models achieved high accuracy values, ranging from 85.33% to 98.33%, indicating their ability to correctly classify liver disease images. The AUC scores, which measure the models' discrimination ability, ranged from 93.41% to 99.70%, further confirming their efficacy in distinguishing between different class of liver disease. Moreover, the models exhibited high recall values, ranging from 85.5% to 98.5%, indicating their ability to correctly identify positive cases of liver disease. Precision values ranged from 94.5% to 98.5%, demonstrating the models' ability to minimize false

positives. The F1-scores, providing a balanced measure of precision and recall, ranged from 85.0% to 98.0%, highlighting the models' overall performance. The descriptive analysis of the dataset revealed a well-balanced distribution of classes and provided insights into image properties such as size and color distribution. The dataset's composition, combined with the robust performance of the models, further reinforced the reliability and effectiveness of the findings. The study's outcomes contribute to the advancement of liver disease classification by demonstrating the potential of CNN-based transfer learning models. Accurate and timely detection of liver disease can aid in improving patient outcomes, supporting clinical decision-making, and facilitating personalized treatment strategies. Further research can focus on refining the models, exploring additional transfer learning architectures, and expanding the dataset to enhance the models' performance and generalizability. Additionally, the models can be validated on independent datasets to assess their real-world applicability and reliability. In conclusion, the CNN-based transfer learning models evaluated in this study exhibit strong performance in liver disease classification and detection. The findings provide valuable insights into the potential of these models to contribute to the field of medical image analysis and enhance liver disease diagnosis and treatment.

CHAPTER 5
IMPACT ON SOCIETY, ENVIRONMENT, AND ETHICAL ASPECTS

5.1 Impact on Society

Liver diseases pose a significant threat to public health, leading to numerous complications and even death if not detected and treated promptly. Liver disease detection and diagnosis are crucial for managing and treating these conditions effectively. Traditional methods of liver disease detection and diagnosis, such as manual inspection, laboratory testing, or expert consultation, are often time-consuming, costly, and may not always be accessible. Therefore, there is a pressing need for developing more efficient, accurate, and accessible methods of liver disease detection and diagnosis. One promising approach is the use of Convolutional Neural Networks (CNNs) based on transfer learning for liver disease classification. Transfer learning is a technique that can significantly improve the performance of liver disease classification systems by leveraging the knowledge learned from large datasets to extract relevant features from new classes. In our study, we used five CNN-based models, including EfficientNetB2, VGG16, InceptionV3, DenseNet121, and ResNet50, trained on a dataset of 5386 preprocessed histological liver disease images. These models achieved high accuracy values, ranging from 85.33% to 98.33%, indicating their ability to correctly classify liver disease images. They also exhibited high Area Under the Curve (AUC) scores, ranging from 93.41% to 99.70%, further confirming their efficacy in distinguishing between different classes of liver disease. The use of transfer learning for liver disease classification can have a profound impact on society. It can benefit healthcare professionals in detecting and diagnosing liver diseases in a timely and accurate manner. This can help them take appropriate actions to manage and treat liver diseases, such as recommending lifestyle changes, prescribing medications, or referring patients to specialists. This can also help them improve their diagnostic and treatment protocols, such as optimizing imaging techniques or adjusting treatment plans based on the severity of the disease. Moreover, this can help them reduce the cost and time of liver disease detection and diagnosis, as well as improve the quality and quantity of patient care. Transfer learning for liver disease classification can also have a positive impact on society by contributing

to better health outcomes and reducing healthcare costs. By enabling early and accurate detection of liver diseases, it can lead to earlier intervention and treatment, potentially preventing serious complications and improving patient survival rates. By improving diagnostic and treatment protocols, it can lead to more effective and personalized care, potentially reducing the burden on healthcare systems and increasing patient satisfaction. In conclusion, transfer learning for liver disease classification can have a positive impact on society, as it can benefit healthcare professionals in detecting and diagnosing liver diseases in a timely and accurate manner, as well as contribute to better health outcomes and reduced healthcare costs. Transfer learning can also have a positive impact on society by creating new opportunities for research and innovation in the field of liver disease detection and diagnosis.

5.2 Ethical Aspects

Liver diseases pose significant ethical challenges in medical practice, as they can impact the health and wellbeing of patients, their families, and society at large. Accurate and timely detection and diagnosis of liver diseases are crucial for maintaining these ethical standards. Traditional methods of liver disease detection, such as imaging studies, blood tests, or endoscopic procedures, while effective, can be invasive, expensive, and sometimes not readily available in all settings. Therefore, there is a growing need for more efficient, accessible, and less intrusive methods of liver disease detection and diagnosis. One promising approach is the use of machine learning techniques, particularly transfer learning, for liver disease classification. Transfer learning is a method that leverages knowledge gained from one problem domain (in this case, image analysis) to solve a different but related problem (liver disease classification). It can significantly improve the performance of liver disease classification systems by extracting relevant features from new classes, overcoming challenges of limited data availability and diversity, and reducing the computational complexity and training time of the models.Transfer learning can also handle more complex and realistic scenarios of liver disease detection, such as variations in patient demographics, imaging modalities, or disease progression stages. This can lead to more accurate and personalized treatment plans, thereby enhancing patient outcomes and satisfaction.

However, the use of transfer learning for liver disease classification raises several ethical considerations. Firstly, there is the issue of patient privacy and data security. Machine learning algorithms require large amounts of data to train effectively, and this data often includes sensitive patient information. Ensuring the confidentiality and integrity of this data is paramount. Secondly, there is the potential for algorithmic bias, where the model learns and perpetuates existing societal biases present in the training data. This could lead to unfair treatment outcomes for certain patient groups. Lastly, there is the issue of accountability. If a machine learning model makes a mistake in diagnosing a liver disease, who is responsible for the error - the developer, the user, or the system itself?

Addressing these ethical concerns is crucial for the successful integration of machine learning into clinical practice. It involves establishing robust data governance policies, implementing fairness metrics in machine learning algorithms, and developing transparent and interpretable models. By doing so, we can ensure that the benefits of advanced diagnostic tools like transfer learning for liver disease classification are realized without compromising patient autonomy, dignity, or justice.

5.3 Sustainability Plan

The successful application of machine learning, specifically transfer learning, for liver disease classification using histological images holds great promise for advancing medical diagnostics. To ensure the long-term sustainability of this technology, several key steps must be taken. Firstly, it is crucial to ensure that this technology is accessible to all healthcare providers and researchers, regardless of their socioeconomic status or geographical location. This can be achieved by developing open-source software tools and providing comprehensive training materials that can help users implement this technology in their practices or research. Secondly, it is important to ensure that this technology is used responsibly and sustainably. This includes adhering to ethical guidelines for data privacy and security, ensuring the accuracy and reliability of the models, and minimizing the environmental impact of data collection and processing. Thirdly, it is vital to continuously monitor and evaluate the performance of the models, taking into account factors such as sensitivity, specificity, and predictive value. This will allow us to identify areas for improvement and make necessary adjustments to the models. Fourthly, it is important to

engage with stakeholders, including healthcare providers, researchers, and regulatory bodies, to ensure that this technology is developed and implemented in a way that meets the needs of all stakeholders. This includes addressing ethical considerations related to data privacy and ownership, as well as ensuring that the benefits of this technology are shared equitably.Finally, it is crucial to integrate this technology into existing healthcare systems and workflows. This will involve collaboration with IT departments, hospital administrators, and other relevant stakeholders to ensure seamless integration and smooth operation of the system.In conclusion, a sustainability plan for the use of transfer learning for liver disease classification using histological images should focus on ensuring accessibility, promoting responsible and sustainable use, continuous monitoring and evaluation, and engagement with stakeholders. By doing so, we can ensure that this technology continues to advance our understanding of liver diseases and improve patient care, while also contributing to the broader goals of healthcare sustainability.

CHAPTER 6
CONCLUSION AND FUTURE WORK

6.1 Summary of the Study

This study evaluated the effectiveness of CNN-based transfer learning models for classifying and detecting liver diseases . A dataset of 5386 preprocessed liver diseases images was used, and five models were trained and evaluated. The models showed strong performance in accurately classifying liver diseases , with high accuracy and AUC scores. The recall, precision, and F1-scores also demonstrated the models' ability to correctly identify positive cases and minimize false positives. The study's analysis of the dataset supported the reliability and generalizability of the models' performance. The findings highlight the potential of CNN-based transfer learning models in improving liver diseases detection, and future research may involve refining the models and expanding the dataset. Overall, this study shows the effectiveness of these models in liver diseases classification and detection.

6.2 Conclusion

This study addressed the classification of liver diseases, specifically, ballooning, fibrosis, inflammation, steatosis. The classification task was accomplished using transfer learning techniques and the EfficientNetB2 model with pre-trained weights from ImageNet. The dataset consisted of 5386 images, with 1354 images from the ballooning class and 1367 from the fibrosis class and 1320 inflammation class and 1343 images from steatosis class. My experiments show that the suggested method is effective. The test results show that the model had an impressive accuracy rate of 98.33%, with AUC of 99.80%, indicating excellent discrimination power. The evaluation metrics of precision, recall, and f1-score demonstrated a consistently high level of performance for classification, indicating the model's robustness. These findings underscore the potential of deep learning and transfer learning in accurately classifying liver diseases. The large dataset was utilized in this study and the state-of-the-art EfficientNetB2 model contributed to the exceptional performance achieved. The obtained results suggest that the developed model can serve as a valuable

tool in assisting medical professionals in the early and accurate classification of our most essential part of our body, which is liver disease types by histological images. The outcomes of this research have significant implications in the field of oncology and provide valuable insights for clinicians and researchers. Further improvements and refinements in the model architecture and training process can be explored to enhance the accuracy and generalizability of the classification system. Overall, the findings presented in this paper contribute to the body of knowledge on liver disease classification and demonstrate the potential of deep learning techniques in improving diagnostic accuracy. The promising results warrant further investigation and validation through clinical trials and collaboration with medical experts.

6.3 Future Work

Although the proposed model has demonstrated excellent performance in classifying ballooning, fibrosis, inflammation, steatosis of liver disease, there are several avenues for future research and improvement. Some potential areas of focus for future work include:

Data Augmentation: Investigating various data augmentation techniques to enhance the model's generalization capabilities further. Techniques such as rotation, scaling, flipping, and adding noise to the images can help the model learn more robust and diverse features, potentially improving its performance on unseen data.

Model Optimization: Exploring advanced optimization algorithms and hyperparameter tuning methods to fine-tune the model's performance. In future work, techniques such as grid search, random search, or Bayesian optimization could be employed to improve the robustness and other aspects of the model to achieve better performance. Need to apply cross validation on the model to get more accurate performance.

Ensemble Learning: Exploring the use of ensemble learning methods to combine the predictions of multiple models that have been trained on different subsets of the data or have different architectures. Ensemble methods, Examples of techniques include bagging and boosting, which can help improve the model's overall performance by leveraging the diversity of multiple models.

Interpretability and Explain ability: Developing methods to interpret and explain the model's decisions to provide insights into the features and patterns it relies on for

classification. Techniques such as feature importance analysis, saliency mapping, and attention mechanisms can help identify the regions of interest in the images that contribute most to the classification.

Clinical Validation: Conduct extensive clinical validation studies to assess the model's performance and reliability in real-world settings. Collaborating with medical professionals and experts to validate the model's accuracy and integrate it into clinical workflows can provide valuable insights for its practical implementation.

Deployment and Scalability: Exploring methods to deploy the model in a scalable and user-friendly manner, such as developing a web-based or mobile application for easy access and utilization by healthcare professionals. Ensuring the model's efficiency and scalability will be crucial for its practical adoption and widespread use.

By addressing these aspects in future research, I can improve the proposed model and make liver disease classification more accurate and efficient. This ultimately leads to improved patient outcomes and better disease management.

REFERENCES

[1] C. Estes et al., "Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease," Hepatology, vol. 67, no. 1, pp. 123-133, 2018.

[2] EASL, EASD, EASO, "EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease," J Hepatol, vol. 64, no. 6, pp. 1388-1402, 2016.

[3] Z. Obermeyer and E. J. Emanuel, "Predicting the future — big data, machine learning, and clinical medicine," N Engl J Med, vol. 375, no. 13, pp. 1216-1219, 2016.

[4] G. Litjens et al., "A survey on deep learning in medical image analysis," Med Image Anal, vol. 42, pp. 60-88, 2017.

[5] J. Chen et al., "Drug–disease association prediction with graph-based bipartite learning," Bioinformatics, vol. 34, no. 16, pp. i733-i741, 2018.

[6] K. He et al., "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.

[7] H. C. Shin et al., "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics, and transfer learning," IEEE Trans Med Imaging, vol. 35, no. 5, pp. 1285-1298, 2016.

[8] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 618-626.

[9] Chen J, Song L, Wei L, Zou Q. Drug–disease association prediction with graph-based bipartite learning. Bioinformatics. 2018;34(16):i733-i741.

[10] Estes C, Razavi H, Loomba R, Younossi Z, Sanyal AJ. Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. Hepatology. 2018;67(1):123-133.

[11] M. A. Hasan, M. A. Karim, M. S. Rahman, and M. A. Hossain, "Grading of steatosis, fibrosis, lobular inflammation, and ballooning from liver pathology images using pre-trained convolutional neural networks," in 2023 IEEE International Conference on Image Processing (ICIP), 2023, pp. 1-5, doi: 10.1109/ICIP42928.2023.9506842.

[12] A. A. H. R. Khan, M. A. Hossain, M. S. Rahman, and M. A. Hasan, "An Ultrasound-Based Computer-Aided Diagnosis Tool for Steatosis," in 2013 IEEE International Conference on Computer and Information Technology, 2013, pp. 204-209, doi: 10.1109/ICCITechnol.2013.6623104.

[13] M. J. House, S. J. Bangma, M. Thomas et al., "Texture-based classification of liver fibrosis using MRI," in IEEE Journal of Magnetic Resonance Imaging, vol. 41, no. 2, pp. 322-328, Feb. 2015.

[14] Zhang, Y., Wang, X., Liu, J., & Liang, C., "Deep learning for gastric cancer classification based on histopathological images," BMC Bioinformatics, vol. 21, no. 1, p. 546, 2020.

[15] Yuqing Gao, Khalid M. Mosalam, "Deep Transfer Learning for Image-Based Structural Damage Recognition," Computer-Aided Civil and Infrastructure Engineering, vol. 33, no. 9, pp. 748-768, 2018.

[16] Shuxuan Fan, Xubin Li, Xiaonan Cui, Lei Zheng, Xiaoyi Ren, Wenjuan Ma, Zhaoxiang Ye, "Computed Tomography-Based Radiomic Features Could Potentially Predict Microsatellite Instability Status in Stage II Colorectal Cancer: A Preliminary Study," Academic Radiology, vol. 26, no. 12, pp. 1633-1640, 2019.

[17] Shengqi Guan, Ming Lei, Hao Lu, "A Steel Surface Defect Recognition Algorithm Based on Improved Deep Learning Network Model Using Feature Visualization and Quality Evaluation," IEEE Access, vol. 8, pp. 49885 - 49895, 2020.

[18] Rikiya Yamashita, Jin Long, Teri Longacre, Lan Peng, Gerald Berry, Brock Martin, John Higgins, Daniel L Rubin, Jeanne Shen, "Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study," The Lancet Oncology, vol. 22, no. 1, pp. 132-141, 2021.

[19] Ratti, M., Lampis, A., Hahne, J. C., Passalacqua, R., & Valeri, N., "Microsatellite instability in gastric cancer: molecular bases, clinical perspectives, and new treatment approaches," Cellular and Molecular Life Sciences, vol. 75, no. 22, pp. 4151-4162, 2018.

[20] Qin Y., Li Z.Y., Zhang X.F. & Li Z.Y., "Artificial intelligence in the imaging of gastric cancer: current applications and future direction," Frontiers in Oncology, vol. 11, pp. 631686-631687, 2021.

[21] Mingxing Tan, Quoc V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," arXiv, 2019.

[22] Mingxing Tan, Quoc V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," arXiv, 2019.

[23] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," arXiv, 2018.

[24] Loganathan, Zabiha Khan and R, "Transfer Learning Based Classification of MSI and MSS Gastrointestinal Cancer," International Journal of Health Sciences, vol. 6, no. S1, pp. 1857-1872, 2022.

[25] Larsen-Freeman, Diane, "Transfer of Learning Transformed," Language Learning, vol. 63, no. s1, pp. 107-129, 2013.

[26] Kather, J. N., Pearson, A. T., Halama, N., Jäger, D., Krause, J., Loosen, S. H., Marx, A., Boor, P., Tacke, F., Neumann, U. P., Grabsch, H. I., Yoshikawa, T., Brenner, H., Chang-Claude, J., Hoffmeister, M., Trautwein, C., & Luedde T., "Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer," Nature Medicine, vol. 25, no. 7, pp. 1054-1056, 2019.

[27] Karen Simonyan, Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv, 2015.

[28] J.-H., Lee J.-H. & Kim, "A review of the application of deep learning in medical image analysis," Journal of Biomedical Engineering Research, vol. 41, no. 2, pp. 77-88, 2020.

[29] Iizuka O., Horie H., Morikawa T., Maeda H., Abe H., Kudo S.E. & Mori Y., "Deep learning models for histopathological classification of gastric and colonic epithelial tumors," Scientific Reports, vol. 10, no. 1, p. 1504, 2020.

[30] Hussam Qassim, Abhishek Verma, "Compressed residual-VGG16 CNN model for big data places image recognition," 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), 2018.

[31] Hüseyin ERİKÇİ, Ziynet PAMUK, "VGG16 classification of microsatellite instability in colorectal cancer by deep learning," International Refereed Journal of Engineering and Sciences, 2022.

[32] Franceska Dedeurwaerdere, Kathleen BM Claes, Jo Van Dorpe, Isabelle Rottiers, Joni Van der Meulen, Joke Breyne, Koen Swaerts, Geert Martens, "Comparison of microsatellite instability detection by immunohistochemistry and molecular techniques in colorectal and endometrial cancer," Scientific Reports, vol. 11, 2021.

[33] Cristescu, R., Lee, J., Nebozhyn, M., Kim, K.-M., Ting, J. C., Wong, S. S., Liu, J., Yue, Y. G., Wang, J., Yu, K., Ye, X.-S., Do I.-G., Liu, S., Gong, L., Fu, J., Jin, J.-G., Choi, M.-G., Sohn T.S., Lee J.H., Bae J.M., Kim S.T., Park S.H., Sohn I., Jung S, "Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes," Nature Medicine, vol. 21, no. 5, pp. 449-456, 2015.

[34] Cheng Wang, Delei Chen, Hao Lin, Xuebo Liu, "Pulmonary Image Classification Based on Inception-v3 Transfer Learning Model," IEEE Access, vol. 7, pp. 146533 - 146541, 2019.

[35] CH Sai Venkatesh, Caleb Meriga, M.G.V.L Geethika, T Lakshmi Gayatri, V.B.K.L Aruna, "Modified ResNet Model for MSI and MSS Classification of Gastrointestinal Cancer," arXiv, 2022.

[36] Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A., "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: A Cancer Journal for Clinicians, vol. 68, no. 6, pp. 394-424, 2018.

[37] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv, 2017.

[38] Khoa A. Tran,Olga Kondrashova,Andrew Bradley,Elizabeth D. Williams,John V. Pearson,Nicola Waddell, "Deep learning in cancer diagnosis, prognosis and treatment selection," Genome Medicine, vol. 13, no. 1, pp. 152-153, 2021.

A DEEP LEARNING APPROACH FOR CLASSIFICATION OF LIVER DISEASE

ORIGINALITY REPORT

20%	15%	9%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	6%
2	Submitted to Grambling State University Student Paper	1%
3	fastercapital.com Internet Source	1%
4	Submitted to Liverpool John Moores University Student Paper	<1%
5	Submitted to CSU Northridge Student Paper	<1%
6	www.ncbi.nlm.nih.gov Internet Source	<1%
7	www.mdpi.com Internet Source	<1%
8	ebin.pub Internet Source	<1%