



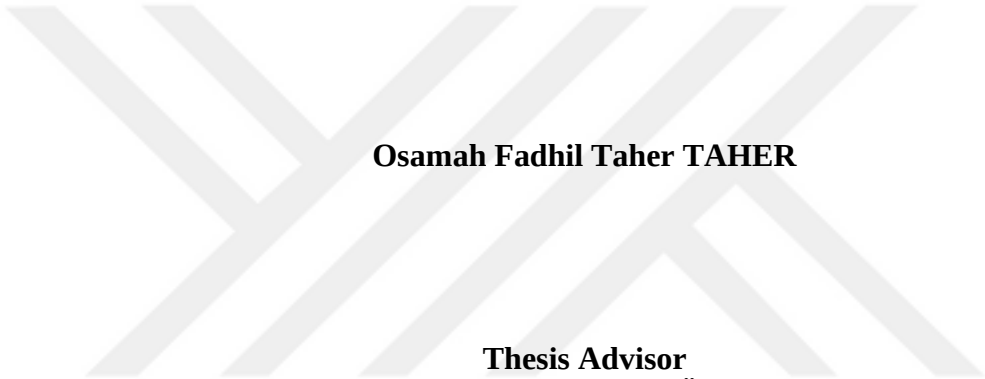
**AN ENHANCED CAPSULE NEURAL NETWORK
MODEL FOR HEEL DISEASE DIAGNOSIS
USING LATERAL FOOT X-RAY IMAGES**

**2025
PhD. THESIS
COMPUTER ENGINEERING**

Osamah Fadhil Taher TAHER

**Thesis Advisor
Assist. Prof. Dr. Kasım ÖZACAR**

**AN ENHANCED CAPSULE NEURAL NETWORK MODEL FOR HEEL
DISEASE DIAGNOSIS USING LATERAL FOOT X-RAY IMAGES**



Osamah Fadhil Taher TAHER

**Thesis Advisor
Assist. Prof. Dr. Kasım ÖZACAR**

**T.C.
Karabuk University
Institute of Graduate Programs
Department of Computer Engineering
Prepared as
PhD. Thesis**

**KARABUK
January 2025**

I certify that in my opinion the thesis submitted by Osamah Fadhil Taher TAHER titled “AN ENHANCED CAPSULE NEURAL NETWORK MODEL FOR HEEL DISEASE DIAGNOSIS USING LATERAL FOOT X-RAY IMAGES” is fully adequate in scope and in quality as a thesis for the degree of Doctorate of Computer Engineering.

Assist. Prof. Dr. Kasım ÖZACAR

.....

Thesis Advisor, Department of Computer Engineering

This thesis is accepted by the examining committee with a unanimous vote in the Department of Computer Engineering as a PhD. thesis. Jan 6, 2025

Examining Committee Members (Institutions)

Signature

Chairman : Assist. Prof. Dr. Yasin ORTAKCI (KBÜ)

.....

Member : Assoc. Prof. Dr. Evrim GÜLER (BARÜ)

.....

Member : Assoc. Prof. Dr. Caner ÖZCAN (KBÜ)

.....

Member : Assoc. Prof. Dr. Ümit ATİLA (GÜ)

.....

Member : Assist. Prof. Dr. Kasım ÖZACAR (KBÜ)

.....

The degree of PhD by the thesis submitted is approved by the Administrative Board of the Institute of Graduate Programs, Karabuk University.

Assoc. Prof. Dr. Zeynep ÖZCAN

.....

Director of the Institute of Graduate Programs



“I declare that all the information within this thesis has been gathered and presented in accordance with academic regulations and ethical principles and I have, according to the requirements of these regulations and principles, cited all those which do not originate in this work as well.”

Osamah Fadhil Taher TAHER

ABSTRACT

PhD. Thesis

AN ENHANCED CAPSULE NEURAL NETWORK MODEL FOR HEEL DISEASE DIAGNOSIS USING LATERAL FOOT X-RAY IMAGES

Osamah Fadhil Taher TAHER

Karabük University

Institute of Graduate Programs

Department of Computer Engineering

Thesis Advisor:

Assist. Prof. Dr. Kasım ÖZACAR

January 2025, 135 pages

Heel bone conditions, such as heel spurs and Sever's disease, pose significant challenges to patients' daily activities. While orthopedic and traumatology doctors rely on foot X-rays for diagnosis, there is a need for more AI-based detection and classification of these conditions. This thesis addresses this need by proposing two deep learning models: HeCapsNet and MedCapsNet.

This thesis first presents a novel dataset of 3,956 annotated lateral foot X-ray images and uses the original capsule network (CapsNet) as a baseline. The low accuracy of CapsNet (73.99%) led to the development of an enhanced CapsNet, HeCapsNet, which adjusts the feature extraction layers, adds extra convolutional layers and utilizes improved initialization and padding schemes. HeCapsNet achieved higher accuracy rates, including 97.29% for balanced data, 94.19% for imbalanced data, an AUC of 98.69%, and a 5-fold cross-validation accuracy of 95.77%. While our findings aligned

with those reported in the literature for relatively more uncomplicated, non-medical datasets like MNIST and F-MNIST, we achieved better results with more complex, medically focused datasets, including brain MRI, knee, and our novel dataset.

In addition, this thesis introduces MedCapsNet, a new hybrid capsule model that combines the modified DenseNet201 model with the original CapsNet. The architecture of this model is designed to detect and classify cases of heel bone pathologies using lateral heel X-ray images of the foot. The model has been applied to a heel data set and other medical datasets (John Cheng's MRI brain, Breast BreakHis v1, and HAM10000 melanoma). MedCapsNet's versatility in different types of images, including X-ray, MRI, and microscopy images, where it achieved high-performance metrics without requiring a unique architecture for each set or tuning the hyperparameter for each dataset, made it superior to state-of-the-art models.

HeCapsNet and MedCapsNet offer the potential of AI-based methods to detect and categorize heel bones, fulfilling the need for diagnosis and more reliable management of these conditions.

Key Words : Capsule network, Deep Learning, Foot X-ray images, Heel spur, Sever Disease

Science Code : 92432

ÖZET

Doktora Tezi

LATERAL AYAK RÖNTGEN GÖRÜNTÜLERİNİ KULLANARAK TOPUK HASTALIKLARI TEŞHİSİ İÇİN GELİŞTİRİLMİŞ KAPSÜL SİNİR AĞI MODELİ

Osamah Fadhil Taher TAHER

Karabük Üniversitesi

Lisansüstü Eğitim Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

Tez Danışmanı:

Dr. Öğr. Üyesi. Kasım ÖZACAR

Ocak 2025, 135 sayfa

Topuk dikenli ve Sever hastalığı gibi topuk kemiği rahatsızlıkları, hastaların günlük aktivitelerinde önemli zorluklar yaratmaktadır. Ortopedi ve travmatoloji doktorları teşhis için ayak röntgenlerine güvenirken, bu durumların daha fazla yapay zeka tabanlı tespitine ve sınıflandırılmasına ihtiyaç vardır. Bu ihtiyacı karşılamak için bu çalışma, HeCapsNet ve MedCapsNet iki derin öğrenme modeli sunmaktadır:

İlk olarak 3.956 yan ayak röntgen görüntüsünden oluşan yeni bir veri kümesi sunmakta ve orijinal kapsül ağını (CapsNet) temel olarak kullanmaktadır. CapsNet'in düşük doğruluğu (%73,99), özellik çıkarma katmanlarını ayarlayan, ekstra evrişimli katmanlar ekleyen ve gelişmiş başlatma ve doldurma şemaları kullanan gelişmiş bir kapsül ağı olan HeCapsNet'in geliştirilmesine yol

açmıştır. HeCapsNet, dengeli veriler için %97,29, dengesiz veriler için %94,19, %98,69 AUC ve %95,77 5 kat çapraz doğrulama doğruluğu dahil olmak üzere daha yüksek doğruluk oranları elde etmiştir. Son teknoloji ürünü modifiye CapsNet modelleriyle yapılan karşılaştırmalar, HeCapsNet'in daha karmaşık tıbbi veri kümelerinde daha iyi performans gösterdiğini ortaya koymaktadır.

Ayrıca bu çalışma, modife edilmiş bir DenseNet201 ile bir kapsül ağını birleştiren yeni bir hibrit kapsül modeli olan MedCapsNet'i de önermektedir. Bu model, yan topuk ayak röntgen görüntülerini kullanarak topuk kemiği hastalıklarını doğru bir şekilde tespit etmek ve sınıflandırmak için tasarlanmıştır. Topuk veri kümesi ve diğer tıbbi veri kümeleri (Jun Cheng Brain MRI, Breast BreaKHis v1 ve HAM10000 cilt kanseri) üzerinde yapılan deneyler, veri kümesine özgü mimari veya hiperparametre ayarına ihtiyaç duymadan yüksek performans ölçümleri elde eden MedCapsNet'in çok yönlülüğünü göstermektedir.

HeCapsNet ve MedCapsNet, topuk kemiklerinin tespiti ve sınıflandırılmasına yardımcı olmak için yapay zeka tabanlı yaklaşımların potansiyelini ortaya koyarak, teşhis süreçlerinin iyileştirilmesine ve bu koşulların daha güvenilir bir şekilde yönetilmesine katkıda bulunmayı amaçlamaktadır.

Anahtar Kelimeler : Kapsül ağı, Derin öğrenme, Ayak röntgen görüntüleri, Topuk diken, Sever Hastalığı

Bilim Kodu : 92432

ACKNOWLEDGMENT

Bismillahir Rahmanir Rahim,

All praise be to Allah, the Most Gracious and the Most Merciful, for bestowing upon me the opportunity and strength to complete this doctoral endeavor.

From the depths of my heart, I express my heartfelt gratitude to my beloved family, to my esteemed father and mother, whose unwavering love, sacrifice, and constant prayers have been the guiding light throughout my journey, to my cherished wife, whose unwavering support, patience, and understanding have been the bedrock upon which I have thrived, to my dearest son, Özcan, whose innocent joy and pure heart have been a continuous source of inspiration and motivation.

I sincerely thank my dear brother and Mr. Ayad, whose steadfast companionship and encouragement have been invaluable to my trusted friends and whose unwavering support and camaraderie have enriched my life and bolstered my determination.

I would like to express my heartfelt appreciation to my esteemed supervisor, Assist. Prof. Dr. Kasım ÖZACAR, whose tireless guidance, endless patience, and unwavering belief in me have successfully completed this research. Your mentorship has been a true gift, and I am forever grateful.

Furthermore, I extend my gratitude to the academic and administrative staff in the Computer Engineering Department for their support and contribution to my academic journey, providing me with the resources, knowledge, and encouragement needed to achieve this milestone.

CONTENTS

	<u>Page</u>
APPROVAL.....	ii
ABSTRACT.....	iv
ÖZET.....	vi
ACKNOWLEDGMENT.....	viii
CONTENTS.....	ix
LIST OF FIGURES	xiii
LIST OF TABLES	xv
SYMBOLS AND ABBREVIATIONS INDEX	xvi
 PART 1	 1
INTRODUCTION	1
1.1. PROBLEM STATEMENT	4
1.2. RESEARCH OBJECTIVES	4
1.3. RESEARCH QUESTIONS.....	5
1.4. RESEARCH CONTRIBUTIONS.....	6
 PART 2	 8
THEORETICAL BACKGROUND.....	8
2.1. INTRODUCTION OF HEEL BONE DISEASES AND MOTIVATION TO AI-BASED DETECTION.....	8
2.2. OVERVIEW OF MACHINE LEARNING AND NEURAL NETWORKS .	12
2.2.1. Introduction to Machine Learning (ML)	12
2.2.2. Overview of Neural Networks and Key Components	14
2.2.2.1. Neural Network Structure.....	15
2.2.2.2. Neural Network Key Components (Neurons, Weights, Biases, Activation and Loss Functions).....	16
2.2.3. Overfitting and Underfitting	20
2.2.3.1. Overfitting.....	21
2.2.3.2. Underfitting.....	21
2.2.4. Evaluation Performance of Machine Learning Models.....	22

	<u>Page</u>
2.3. OVERVIEW OF DEEP LEARNING	24
2.3.1. Introduction to Deep Learning (DL).....	24
2.3.2. Convolutional Neural Networks (CNNs)	26
2.3.2.1. Convolutional Layer	27
2.3.2.2. Pooling Layer.....	30
2.3.2.3. Dropout Layer.....	31
2.3.2.4. Fully Connected Layer	32
2.3.2.5. Non-Linearity (Activation Functions)	33
2.3.2.6. Backpropagation	34
2.3.3. Limitations of CNNs.....	34
2.4. CAPSULE NETWORKS: CONCEPT AND KEY COMPONENTS.....	35
2.4.1. Introduction to Capsule Networks	35
2.4.2. Capsule Network Architecture.....	36
2.4.3. Limitations of Capsnet in Medical Image Processing	39
PART 3	41
RELATED WORK	41
PART 4	65
METHODOLOGY.....	65
4.1. DATASET CREATION AND ANALYSIS	65
4.1.1. Collection of Lateral Foot X-ray Images	65
4.1.2. Dataset Characteristics.....	66
4.1.3. Annotation Process	68
4.1.4. Data Preprocessing Techniques	69
4.2. BASELINE MODEL - ORIGINAL CAPSNET	70
4.2.1. Application of CapsNet to Heel Disease Dataset	70
4.2.2. Performance Analysis (Accuracy, AUC, Cross-Validation)	71
4.2.2.1. Accuracy	71
4.2.2.2. Area Under the Curve.....	72
4.2.2.3. Confusion Matrix Analysis.....	73
4.2.3. Limitations of CapsNet for Heel Disease Detection.....	74
4.3. AN ENHANCED CAPSULE NETWORK (HECAPSNET).....	76

	<u>Page</u>
4.3.1. Introduction to HeCapsNet	76
4.3.2. Architectural Modifications	77
4.3.3. Incorporating Yolo with HeCapsNet	79
4.3.3.1. Implementation of YOLO Steps	80
4.4. HYBRID CAPSULE NETWORK (MEDCAPSNET)	80
4.4.1. Introduction to MedCapsNet	80
4.4.2. Introduction to DenseNet201 Model	82
4.4.2.1. Dense Blocks	84
4.4.2.2. Transition Layers	84
4.4.2.3. Growth Rate Optimization.....	84
4.4.2.4. Output Layer	85
4.4.3. Integration Of DenseNet201 with Capsule Network.....	85
PART 5	91
EXPREMENTS AND RESULTS	91
5.1. APPLICATION OF HECAPSNET ON HEEL DISEASE DATASET	91
5.2. APPLICATION OF HECAPSNET ON DIFFERENT DATASETS.....	94
5.3. APPLICATION OF MEDCAPSNET AND OTHER MODELS TO HEEL DATASET.....	99
5.4. EXPERIMENTS OF MEDCAPSNET ON ADDITIONAL MEDICAL DATASETS	106
5.4.1. Breast BreaKHis v1 Dataset	106
5.4.2. HAM10000 Skin Cancer Dataset	109
5.4.3. Jun Cheng’s Brain Imaging MRI Dataset.....	112
5.5. ASPECTS OF COMPARATIVE MODELS OVER A SHARED DATASET	117
PART 6	119
DISCUSSION	119
6.1. ANALYSING OF PROPOSED MODELS RESULTS	119
6.2. ADVANTAGES AND INNOVATIONS OF PROPOSED MODELS	120
6.3. CHALLENGES AND LIMITATIONS	121

	<u>Page</u>
PART 7	123
CONCLUSION	123
REFERENCES.....	125
RESUME	135



LIST OF FIGURES

	<u>Page</u>
Figure 2.1. Heel spure.	8
Figure 2.2. Sever disease.....	9
Figure 2.3. Sever in X-ray.....	10
Figure 2.4. Paradigm of supervised learning.....	13
Figure 2.5. Paradigm of unsupervised learning.....	14
Figure 2.6. Neural network structure.	16
Figure 2.7. Artificial neuron model adapted from	17
Figure 2.8. Example of overfitting adapted from.....	21
Figure 2.9. Underfitting example. adapted from.....	22
Figure 2.10. Family of AI adapted from	24
Figure 2.11. Differences between machine learning and deep learning steps atapted from.....	25
Figure 2.12. CNN layers adapted from	27
Figure 2.13. Convolutional layer mechanism adapted from	28
Figure 2.14. Kernel size of 3×3 adapted from.....	29
Figure 2.15. A zero padding around the original input pixels adapted From	30
Figure 2.16. Pooling layer (Max Pooling) adapted From	31
Figure 2.17. (a) is a standard neural network, (b) after applying the dropout adapted from.....	32
Figure 2.18. FC layer as represented final result of CNN. Adapted from	33
Figure 2.19. Original CapsNet structure adapted from	38
Figure 2.20. CapsNet decoder structure adapted from.....	39
Figure 4.1. Normal heel.....	66
Figure 4.2. Heel spur.	67
Figure 4.3. Heel with sever disease.....	67
Figure 4.4. (a) Normal heel, (b) Heel spur, (c) Sever disease.	68
Figure 4.5. Samples of cropped ROI images.	69
Figure 4.6. Original capsule architecture with heel dataset.	71
Figure 4.7. Training and testing accuracy of CapsNet applied on heel dataset.	72
Figure 4.8. Training and testing loss of CapsNet applied on heel dataset.	72

	<u>Page</u>
Figure 4.9. Curve of AUC for original CapsNet applied on heel dataset.	73
Figure 4.10. Confusion matrix of original CapsNet applied on heel dataset.	74
Figure 4.11. Proposed enhanced CapsNet (HeCapsNet).....	79
Figure 4.12. YOLO is applied to a heel X-ray image to automatically crop the ROI.	80
Figure 4.13. Dense connectivity pattern adapted from	83
Figure 4.14. Architecture of MedCapsNet model.	90
Figure 5.1. HeCapsNet training - testing Accuracy over epochs on heel dataset. ..	92
Figure 5.2. HeCapsNet training - testing loss over epochs on heel dataset.	93
Figure 5.3. HeCapsNet AUC ROC curve on heel dataset.....	93
Figure 5.4. HeCapsNet AUC ROC curve on heel dataset.....	94
Figure 5.5. Distribution of balancing heel diseases dataset.	100
Figure 5.6. Training and testing accuracy of heel diseases dataset.....	101
Figure 5.7. Training and testing loss of heel diseases dataset.....	101
Figure 5.8. AUC curve for heel diseases dataset.....	102
Figure 5.9. Confusion matrix of heel diseases dataset.	102
Figure 5.10. Training and testing accuracy of cross validation.	104
Figure 5.11. Training and testing loss of cross validation.	104
Figure 5.13. Confusion matrix of 5-fold cross validation of heel dataset.	105
Figure 5.14. Training and testing accuracy of breast BreaKHis v1.	107
Figure 5.15. Training and testing loss of breast BreaKHis v1.	108
Figure 5.16. AUC curve of breast BreaKHis v1.	108
Figure 5.17. Confusion matrix of mreast BreaKHis v1.	109
Figure 5.18. Accuracy of HAM10000 skin cancer dataset.	110
Figure 5.19. Loss of HAM10000 skin cancer dataset.	110
Figure 5.20. AUC curve of HAM10000 skin cancer dataset.	111
Figure 5.21. Confusion matrix of HAM10000 skin cancer dataset.	111
Figure 5.22. Accuracy of Jun Cheng's brain imaging MRI dataset.....	113
Figure 5.23. Loss of Jun Cheng's braini MRI dataset.....	113
Figure 5.24. AUC curve of Jun Cheng's brain imaging MRI dataset.....	114
Figure 5.25. Confusion matrix of Jun Cheng's brain imaging MRI dataset.	114

LIST OF TABLES

	<u>Page</u>
Table 2.1. Confusion matrix overview.....	23
Table 2.2. Dynamic routing algorithm	38
Table 4.1. Class-Wise Performance Metrics of CapsNet on heel disease dataset....	74
Table 4.2. Comparison of original and modified DenseNet201.....	89
Table 5.1. Class-Wise Performance Metrics of HeCapsNet on Heel Disease Dataset.	94
Table 5.2. Accuracy comparison between HeCapsNet and original capsule models for knee and heel disease datasets.	97
Table 5.3. Performance (accuracy) comparison between proposed and state_of_the_art models.....	98
Table 5.4. Results in comparison with base and recent capsule network models for heel dataset.	103
Table 5.5. Class-Wise performance metrics of MedCapsNet on heel disease dataset.	103
Table 5.6. Class-Wise of cross vlaidation performance metrics of MedCapsNet on heel disease dataset.....	106
Table 5.7. Class-Wise performance metrics of MedCapsNet on breast BreaKHis v1.	109
Table 5.8. Class-Wise performance metrics of MedCapsNet on HAM10000 skin cancer.	112
Table 5.9. Class-Wise performance metrics of MedCapsNet on Jun Cheng brain tumor dataset.	115
Table 5.10. Results in comparison for various medical datasets with MedCapsNet vs. other models.	115
Table 5.11. Performance Comparison of Capsule Network Models	118

SYMBOLS AND ABBREVIATIONS INDEX

SYMBOLS

σ	: Standard deviation
$\hat{u}_{j i}$	: Predicted output vector for capsule j from capsule i
s_j	: Total input to capsule j
v_j	: Activation vector for capsule j
c_{ij}	: Coupling coefficient between capsules i and j
W_{ij}	: Weight matrix between capsules i and j
b_{ij}	: Agreement scores for routing between capsules i and j
a_{ij}	: Consistency scalar value, representing the agreement between capsules i and j
b	: Bias Value
x_j	: Input Values
max	: Maximum Function
$\sum$	: Summation Function

ABBREVIATIONS

AI	: Artificial Intelligence
ANN	: Artificial Neural Networks
AUC	: Area Under the Curve
CapsNet	: Capsule Neural Networks
CNN	: Convolutional Neural Networks
LSTM	: Long Short-Term Memory
FC	: Fully Connected
GPU	: Graphic Processing Unit

KNN	: K-Nearest Neighbor
MNIST	: Mixed National Institute for Standards and Technology
F-MNIST	: Fashion-MNIST, a dataset of Zalando's article images
CIFAR-10	: Canadian Institute for Advanced Research dataset with 10 classes
HAM10000	: Human Against Machine with 10,000 training images
ReLU	: Rectified Linear Unit
ROI	: Region of Interest
ROC	: Receiver Operating Characteristic
SVM	: Support Vector Machine
CT	: Computer Tomography
MRI	: Magnetic Resonance Imaging
PET	: Positron Emission Tomography
WSIs	: Whole-Slide Images
YOLO	: You Only Look Once
US	: Ultrasound
exp	: Exponential Function
ML	: Machine Learning
MLP	: Multi-Layer Perceptron
RGB	: Red, Green, Blue
ViTs	: Vision Transformers
NLP	: Natural Language Processing
MLP	: Multi-Layer Perception
CXR	: Chest X-ray
RSNA	: Radiological Society of North America
LIDC-IDRI	: Lung Image Database Consortium and Image Database Resource Initiative
CLAHE	: Contrast Adaptive Histogram Equalization
VGG	: Visual Geometry Group
ISIC	: International Skin Imaging Collaboration
RFMiD	: Retinal Fundus Multi-disease Image Dataset
TTDCapsN	: Tri Texton-Dense Capsule Network
et	

CBAM	: Convolutional Block Attention Module
COVID-19	: Coronavirus Disease 2019
SMOTE	: Synthetic Minority Over-Sampling Technique
DPafy-GCaps	: Denoising Patch-and-Amplify Gabor Capsule Network
GI	: Gastrointestinal
DWT	: Discrete Wavelet Transform
GHRO	: Genetic Hybrid Rice Optimization
MSCNet	: Multi-Scale Convolutional Capsule Network
ACNGCNN	: Adversarial Capsule Network - Graph Convolutional Neural Networks
GAN	: Generative Adversarial Networks
CAPSGAN	: Capsule Network - Generative Adversarial Networks
SSFO	: Sewer Shad Fly Optimization
RID	: Retrovirus Integration Database
LDA	: LncRNA-Disease Association
UDA	: Uniform Distributive Augmentation
HIV	: Human Immunodeficiency Virus
HIV ISs	: Human Immunodeficiency Virus Integration Sites
EF	: Ejection Fraction
DBN	: Deep Belief Network
SVD	: Singular Value Decomposition
HCNN-CN	: Hybrid Convolutional Neural Network-Capsule Network
AEPO	: Adaptive Emperor Penguin Optimization
GLCM	: Grey Level Co-occurrence Matrix
GRLM	: Gray Level Run-length Matrix
GNNs	: Graph Neural Networks
DICOM	: Digital Imaging and Communications in Medicine

PART 1

INTRODUCTION

The concept of Artificial Intelligence (AI) first originated during the mid-20th century, specifically in the range of the 1950s and the 1960s. In the year 1956, a conference at Dartmouth was held by John McCarthy and Marvin Minsky that outlined key features towards the emerging field of AI. In its earliest stages, traditional AI approaches were highly rule-based in which a computer would be forced to use an internal model to make decisions; however, this was also one of the up-and-coming fields of Deep Learning (DL) that began to change the world [1,2]. While in the past, programming rules were primarily used in AI, advancements in DL have introduced a wealth of possibilities that have revolutionized AI. New and enhanced intelligent systems are being developed to assist and transform healthcare as we know it. AI technology is rapidly advancing in the field of medical imaging, where CT scans, MRIs, ultrasounds, and X-rays remain difficult to interpret due to the overwhelming suite of images produced for specialists [3,4,5]. However, DL models have proven to be efficient in this domain, including the analysis of medical images [6,7,8]. The major types of AI methods suitable for use in conjunction with medical imagery include detection, classification and segmentation of the images [9].

DL, a subset of Machine Learning (ML), employs Artificial Neural Network (ANN) architectures with a distributed representation that consists of a multi-layered coding strategy system with multiple hidden layers to be learned [10]. In the early 2010's, neural networks known as Convolutional Neural Networks (CNNs) were extensively used in the field of medical imaging [11]. Then, in 2017, Hinton G. et al. introduced Capsule Networks (CapsNet), a more advanced technique for conducting medical imaging analysis compared to traditional networks, providing several benefits over other conventional techniques [12].

Heel bone disease affects a patient's performance of daily tasks very much due to reduced mobility, which results from pain on the lower side of the foot. An orthopedic specialist is involved who provides the necessary examinations and foot X-rays and diagnoses the condition, and thus, treatment is done for the patient. Two of the most common heel bone diseases are heel spurs and Sever's disease. Heel spurs are defined as bony growths that extend into the heel bone, which in return lead to inflammation that results in pain, which at times can be managed with physical therapy but at times necessitates surgery, while Sever's disease, which is common in youngsters, is inflammation due to overuse, wherein there is swelling of the posterior aspect of the heel bone, and this causes pain on the heel [13,14].

This research creates a base for lateral X-ray foot views and shows AI's role in medicine. At first, the original CapsNet was tested on the dataset. Still, as the performance was relatively poor, two improved CapsNet models, HeCapsNet and MedCapsNet, were designed to facilitate the identification and classification of heel bone diseases. Introducing advanced capsule models shortens the time and increases the accuracy with which satisfactory treatment can be proposed. This application of AI in medical technologies is one of the milestones that can place the patient's welfare and clinical practice at the next level in the forthcoming time.

This thesis contains eight parts, each focusing on various aspects of advanced Capsule Network (CapsNet) models for detecting and classifying heel bone diseases.

Part 1, Introduction, provides an overview of heel bone diseases like heel spur and severe disease, emphasizing the diagnosing challenges faced with these conditions. It points out how AI-based models, especially CapsNets, can improve detection and classification effectiveness. The thesis also outlined the research aims, highlighted the critical questions to address, and contributed to this dissertation.

Part 2, Theoretical Background, covers essential information connected with this study. This section includes such subtopics as the history of heel bone diseases, why artificial intelligence methods have become popular in their treatment, supervised and unsupervised learning notions, neural network topologies, accuracy criteria employed

in ML, and other concepts crucial to our understanding of DL like CNNs and factors leading to the creation of CapsNets.

Part 3, Capsule Networks: Concept and Key Components, explains how CapsNets can be used in medical image analysis by discussing their appearance and work. It looks at the difficulties faced when using CapsNets with complex datasets and new developments that have improved their efficiency in medical imaging tasks.

Part 4, the methodology, explains how a new dataset for heel disease was created and examined. The data collection process, annotation, and preprocessing procedures were described. This section presents the CapsNet model and discusses the improved HeCapsNet and MedCapsNet models used to check performance against the Heel disease dataset.

Part 5 provides the implementation and results of the baseline CapsNet model on the heel dataset. Architectural improvements and the inclusion of the You Only Look Once (YOLO) framework for automated heel region identification are some of the features that define this advanced convolutional capsule network, HeCapsNet, developed for detecting and classifying heel abnormalities. Furthermore, MedCapsNet, which is a hybrid model combining DenseNet201 with Capsule Networks, is introduced to address feature extraction and classification issues. It is then applied to additional medical datasets such as Breast BreaKHis v1, HAM10000 skin cancer, and Jun Cheng Brain MRI datasets.

Part 6 examines the strengths and contributions of HeCapsNet and MedCapsNet in medical image analysis. This section highlights the novelty of the proposed models, discusses their limitations, and offers recommendations for improvement.

Part 7, Conclusion, summarizes the main results and highlights the relevance of HeCapsNet and MedCapsNet in AI-based medical imaging. It also applies possibilities for further investigations, such as increasing data sets, multimodal imaging, and expanding the models to accommodate other medical uses. Moreover, it is provided

that several scholars are currently involved in the improvement of CapsNet architecture by modifying it for better performance.

1.1. PROBLEM STATEMENT

The conditions of heel spurs and Sever's disease disable the mobility and quality of life of the affected person. Orthopedic physicians are the ones who are qualified to accurately diagnose these conditions and treat them using lateral X-rays of the foot. In places lacking such doctors, the typical clinical examination usually takes a long time and is susceptible to error, which results in late and wrong treatment. Though there has been some success in AI in recent times, particularly in deep learning, the current AI techniques like CNNs or CapsNets are struggling to detect and classify such diseases from medical images properly.

Despite the potential demonstrated by recent encouraging advances in AI and medical image analysis toward contribution to medical diagnostic practice, the development of tailor-made datasets and better models specific to heel bone diseases remain entirely untouched. Using the original CapsNet network, the results showed an inability to achieve the required degree of accuracy for reliable diagnosis. Therefore, it is essential to create more accurate and efficient models for detecting and classifying heel bone diseases. The primary aim of this development is to enhance patient outcomes and minimize the time lost in diagnosis.

This work aims to fill in the gap by providing a new set of lateral foot X-ray images and developing two new enhanced CapsNet models named HeCapsNet and MedCapsNet. These models shall enable more accurate and timely diagnoses, allowing for better patient care through early intervention and better treatment strategies for patients with heel bone diseases.

1.2. RESEARCH OBJECTIVES

This thesis aims to overcome the difficulties in diagnosing diseases such as heel spurs and Sever's disease in heel bones by developing advanced AI-based models that use

Capsule Networks and hybrid architectures like MedCapsNet, including DenseNet201, for better extraction of features. A novel dataset of lateral foot X-ray images categorized and annotated by medical professionals will be created to address the lack of a publicly available dataset. At the same time, data quality has been improved using preprocessing techniques such as ROI cropping, resizing, and balancing. The study investigates methods to improve diagnostic accuracy without relying on data augmentation using enhanced models such as HeCapsNet and MedCapsNet with more convolutional layers, dropout mechanisms, and "he_normal" initializers. These models have also been tested on other medical datasets, including brain MRI and skin cancer, ensuring their versatility and robustness relative to current state-of-the-art architectures on the heel dataset. Moreover, for effective analysis purposes, these algorithms can include object detection schemes to automate ROI detection, such as YOLOv8. This research seeks to design a complete decision support system for automated, accurate diagnosis of heel diseases, particularly for areas with limited access to experts, thus bridging gaps in AI-based medical imaging diagnostics that exist today.

1.3. RESEARCH QUESTIONS

The following research questions are formulated to assist the undertaken study in resolving the existing challenges in the detection and classification of heel bone diseases by means of AI models. These queries delve into the role of advanced AI techniques such as CapsNets in enhancing the accuracy and efficiency of diagnosis in medical imaging. The investigation is also aimed at creating a new dataset and conducting performance evaluation of novel AI models in order to be able to more effectively assist towards the fight against heel and other diseases. The questions posed are meant to find out how the automation of feature extraction, classification, and detection of diseases could be accomplished better by means of the suggested models:

How can lateral foot X-ray images arranged into a dataset by qualified professionals help enhance the performance of AI models in laterally inverted foot imaging?

What improvements can be made to AI-based models so that heel bone diseases such as heel spurs and severe diseases can be accurately detected and classified?

What improvements in performance of the proposed models over the state-of-the-art models, and how do they cope with medical image datasets?

1.4. RESEARCH CONTRIBUTIONS

The following contributions summarize the key advancements made through this study. By addressing existing gaps in AI-based detection and classification of heel bone diseases, these contributions aim to enhance both the dataset availability and the performance of AI models in medical diagnostics. The study introduces novel models and methodologies that significantly improve the accuracy and efficiency of disease detection, particularly in medical imaging. These contributions are outlined below:

- **Novel Dataset Creation:** Both studies contribute to the medical field by creating a unique dataset of 3,956 lateral X-ray images, addressing the lack of sufficient labeled data for diagnosing heel diseases such as heel spurs and Sever's disease. This dataset was annotated by expert orthopedists, providing a solid foundation for AI-based disease detection models.
- **Development of Enhanced Capsule Network Models:**
- **HeCapsNet:** The first model introduced is an advanced CapsNet that targets heel bone diseases. This model overcomes the limitations of the original CapsNet architecture regarding feature extraction by including more convolutional layers, utilizing he_normal kernel initializer for better fitting in medical images, which resulted in 97.29% accuracy for balanced data, 94.19% for imbalanced data and an AUC of 98.69%.
- **MedCapsNet:** The second model introduced was a hybrid model that combines DenseNet201 and CapsNet. This model provides detection and classification of heel disease and proves adaptable to datasets such as breast cancer, brain MRI, and skin cancer, with an accuracy of 96.38% for heel disease and 98.40% for breast cancer datasets.

Automation of system: Both models aim to automate the detection of heel diseases by providing a decision support system that makes X-ray images more accessible and faster to analyze. These systems are particularly valuable in regions where access to orthopedic specialists may be limited and offer a more scalable approach to medical diagnosis.

Evaluation of performance: The models have been thoroughly evaluated on both simple and complex medical datasets, tested against other cutting-edge AI models, and found to outperform them in all factors. Many studies show the developed models can work on complicated medical images faster than conventional approaches.



PART 2

THEORETICAL BACKGROUND

2.1. INTRODUCTION OF HEEL BONE DISEASES AND MOTIVATION TO AI-BASED DETECTION

Conditions cause heel bones such as Sever's disease and heel spurs have been demonstrated to affect patients' quality of life and daily stress levels. Specifically, heel spurs, as depicted in Figure 2.1, are located just below the calcaneus. Such heel spurs are generally a result of excessive mechanical stress on the plantar fascia, a condition often associated with plantar fasciitis. The combined effect of these factors is the manifestation of discomfort, specifically at the plantar surface of the foot. This condition has been observed in 15% to 20% of the population, with a higher prevalence among runners. In some cases, surgical intervention is necessary when medication fails to address heel spurs in the heel bone [13,14].



Figure 2.1. Heel spure [13].

Heel spurs have been reported in several subsets of populations with the ugliest percentage rates being 11.0% and the highest being 55.1%. An individual can evolve with both plantar and dorsal heel spur and even have them in both of their heels. As heel spurs develop over an individual's lifetime, most of them are painless and remain unnoticed until people visit health care centers complaining of heel aches. Most of the heel spur research done by scholars has focused on the condition on one foot, with few attempts having been made to the two footed condition [15].

In the other part, they talked about a condition called Sever's disease that has been discussed in the images 2.2 and 2.3, the medical condition which carries the Sever's name was first noted in 1912 by an orthopedic doctor James Warren Severe. People between the ages of 11-15 are said to be highly affected by calcaneal apophysitis, which is in simpler terms called Sever's disease. These are direct results to excessive strain applied to the body so sort of high impact physical activities or a period a growth spurts. This directly involves an irritated region which is found in the heel region of the foot.

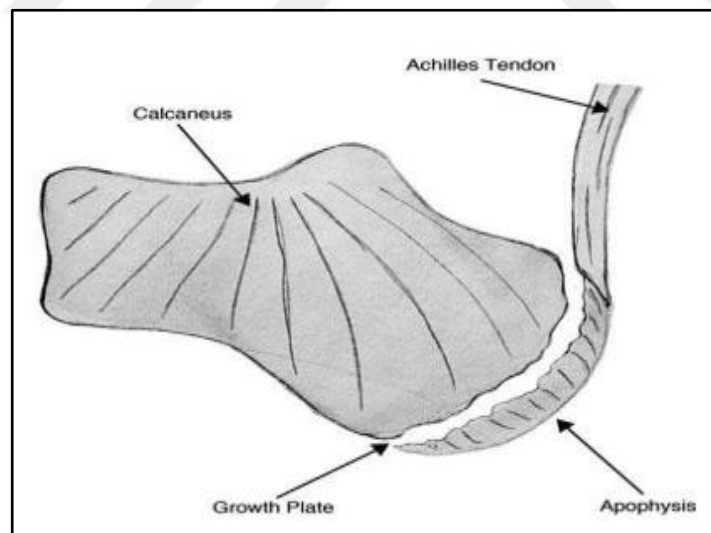


Figure 2.2. Sever disease [16].



Figure 2.3. Sever in X-ray [16].

The heel hurts if one has a medical condition of a significant nature and the heel pain increases when someone walks, jogs or skips. The heel pain is located at the back and the bottom of the heel which is above the calcaneal apophysis. If the pain is more severe, it tends to restrict the capacity of any person to physically carry out any activity causing them to limp.

The patient's heel does not appear to exhibit any remarkable findings, and there are no indications of local disease, such as edema or redness. A preliminary diagnosis can be made via a physical examination that involves palpating the calcaneus to identify discomfort when applying medial-lateral compression to the area of the heel bone. This method, known as the "squeeze test," is a valuable diagnostic tool but generally physicians ask patients to do X-ray radiograph imaging for more specific diagnoses. Consequently, the diagnosis of the higher form of the disease is applied clinically. Studies have demonstrated that heel pain is present in approximately 8.2% of children seeking medical attention [13,14,16].

Traditionally, physicians have diagnosed such conditions by taking X-rays of the foot. However, there is still a demand for AI-based solutions that provide accurate and rapid diagnoses [13,14]. The conventional approach to diagnostics entails a medical professional evaluating the disease through an assembly of academic and clinical knowledge. This is also the case for medical diseases that require diagnostic imaging. In these cases, the calculated interpretation performed by the physician is overly reliant

on personal insight, which makes the technique time-consuming and less reliable [17,18].

The recent rapid growth of AI technologies in many sectors including healthcare provides opportunities to improve accuracy and efficiency of results including accurate detection and classification of diseases. With the help of AI systems, a large amount of data can be analyzed and patterns that may be difficult for people to see are able to be fully explored and thus speeding up the time taken and improving the accuracy of the diagnosis. Speed and accuracy undergoing automation and analysis of complex medical images, including X-rays, MRIs, and CT scans, greatly improves the outcomes of such procedures [19].

The training of AI models offers prospective means of enhancing efficiency with respect to the identification and treatment of diseases [20]. Approaches that utilize AI can assist specialists provided in remote or rural medical clinics where a physician professional is not present [21]. Even when there is a delay in transferring the patient to that location the wellbeing of the patient will still be improved with the help of AI as it saves costs and improves patient outcome [21]. Additionally, AI provides a timely and precise diagnosis of the disease which significantly reduces the expense in the health care system along with improving the efficiency with which the diseases are classified and diagnosed [21].

There is a space which still exists even with multiple AI-based studies focused on disease detection and classification, that systems developed remain incapable of detecting heel bone diseases. An appropriate model and a corresponding dataset have been missing as well, to further supplement this space in the literature. In response to these concerns, the current research seeks to fill this gap by proposing advanced CapsNet models for the accurate detection and classification of heel bone conditions. The dataset of lateral foot X-ray images were collected from a medical institution, and an expert orthopedist and traumatologist labeled and classified the dataset. The constellation of the original model of CapsNet, was first introduced as a newer model taking into consideration this dataset. The outcome was not all that satisfying, the first

model that was presented was not all that enhanced and so this point was the place that we started thinking about developing enhanced models.

2.2. OVERVIEW OF MACHINE LEARNING AND NEURAL NETWORKS

2.2.1. Introduction to Machine Learning (ML)

With the help of data, ML, which is a subset of AI, strives to construct intelligent devices such as computers and more, that are able to automatically make optimizations without the need for manual programming. It aims to provide efficient automation in recognizing and predicting patterns as well as performing decision making. Over time, computing power and access to immense datasets has experienced excessive growth and so have ML methods, which gives us an insight that every day new technologies aid in advancing the methods even more [22].

The core activity of ML is to construct a model on a specified dataset and use that model to make some decision on an unseen data instance. The model is a work in progress and undergoes fine-tuning based on its accuracy. These algorithms include models that are prerequisite to classify whether their input is an image, text or number. More advanced models however take the input and link it to the correct output regardless of the data type. The uses of ML are endless, and from various domains including medicine, banking and self-driving cars. In medicine, ML models are used in many areas such as medical X-rays, diagnostics and in the development of new drugs [23].

The domain of machine learning can be vast in its two main types. It can be subdivided into four main categories:

- **Supervised Learning:** In this case, the data is labeled and so each training sample has an output specified with it. Now the model makes generalizations based on its predictions of the label for data never seen before. For example, one of the tasks for supervised models is image classification- determining if an email is spam or not is a common image

classification use case as Figure 2.4. shows the paradigm of supervised learning. Something similar can be said for supervised tasks that take regression, the main goal is still prediction. This technique is popular in predicting the likelihood of loan repayment in banking and identifying disease markers in medicine [24,25].

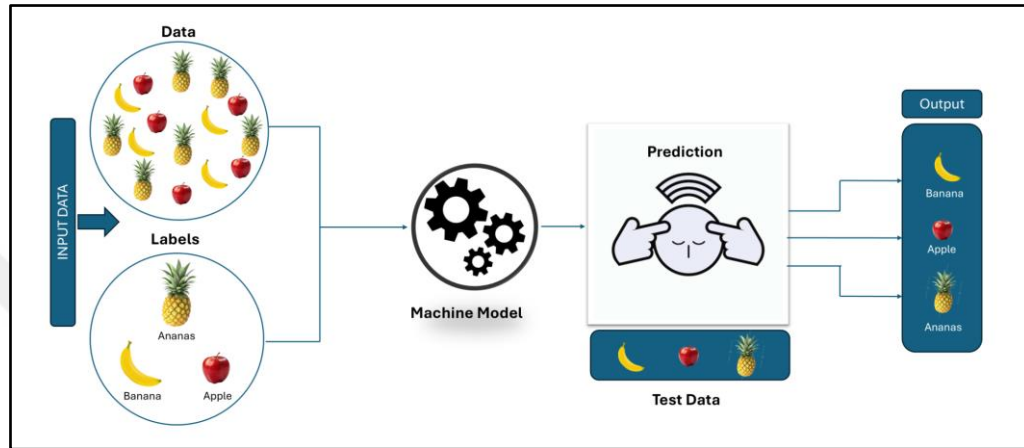


Figure 2.4. Paradigm of supervised learning.

Unsupervised Learning: As opposed to supervised learning, unsupervised learning is trained with unlabeled data, with the aim of finding unseen structures in the data as the paradigm of unsupervised learning is illustrated in Figure 2.5. Clustering algorithms such as k-means and Hierarchical clustering are developed to group data points which have relevant traits in common. Its application unfolding during market segmentation is more concerned with grouping of customers by purchase patterns, and in genomics with the search for unknown molecular disease subtypes with genetic data. Examples of applications of unsupervised learning today are as follows: clustering, density estimation, feature extraction, dimensionality reduction, discovery of association rules, and anomaly detection. [24,25].

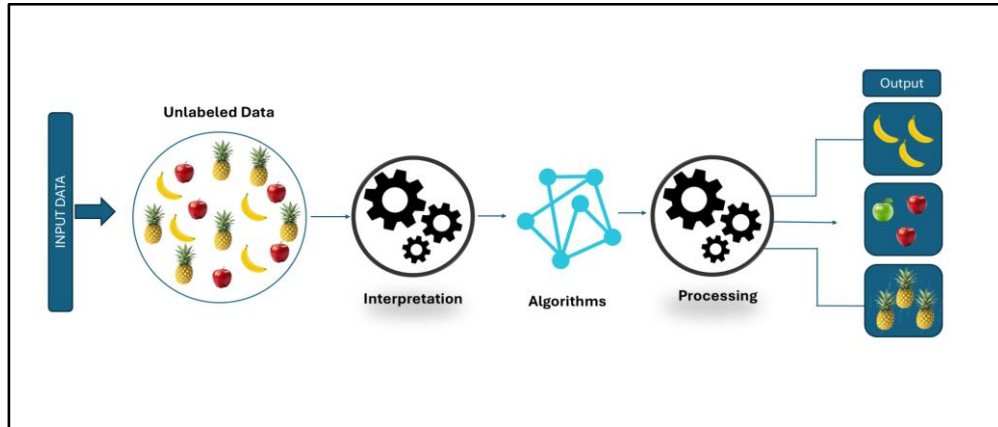


Figure 2.5. Paradigm of unsupervised learning.

The variety of algorithms like support vector machines (SVM), decision trees, and ANN used to be the focus of machine learning [26]. The rapid emergence of, which is a subset of machine learning, but more sophisticated, more advanced models are capable of handling complex, including robotics, video games, natural language processing and computer vision [27]. Medical images have also been analyzed fairly well using deep learning models. Among them are various imaging techniques including ultrasound (US), X-ray, computed tomography (CT) scans, magnetic resonance imaging (MRI) scans, and positron emission tomography (PET) scans, retinal photography, whole slide images (WSI) obtained during histopathology and cytopathology as well as images from endoscopes and dermoscopes [28]. CNNs and CapsNets are common examples of deep learning tools [29].

2.2.2. Overview of Neural Networks and Key Components

Neural networks were introduced by McCulloch and Pitts in the early 1940s, and they made a significant development in relation to the workings of the human brain [30]. Although, they rose to prominence with the advancement of computing in the late 20th century. Neural Networks were described as computational models, made after inspiration from the Human brain in 1999 by Abdi et al. biological neurons are replaced with interconnected units which form these networks. Neural networks are designed to detect associations, comprehend them, and then use the resultant knowledge to anticipate situations as per the defined requirements [31].

Now a days, there has been growth in the application of deep learning strategies and AI, which is a clear advancement of technology developed on the foundation of neural networks, accomplished in distinct areas such as pattern detection, photo recognition, general language understanding and treatment selections.

2.2.2.1. Neural Network Structure

Neural network systems comprise networks of cells called neurons, which form interconnected structures. Such cells discover patterns within the input data and work towards training them using complex computations. In general, three primary types of layers are what compose the structural organization of the neural networks, they include: input layer, hidden layers and output layer. Each layer processes the information received from the previous layer and sends it to the next one. It is explained below as to what the function of these layers is:

Input Layer:

In a neural network, the first layer is the input layer. Every neuron in this layer corresponds to a single feature or attribute within the data set. For example, in image processing, the input layer would encompass the pixel values of an image. The input layer establishes a parameter for entering the network, but it does not carry out any calculations; it only propagates the data to the next hidden layers [32]. Figure 2.1 illustrates the neural network with layers.

Hidden Layers:

Computation takes place in layers between the input and output layers and is termed as the hidden layers. The neurons in these layers receive signals from the input layer, perform some convolution and calculate several steps of the function in a complex [32,33] process. The existence of multiple hidden layers is a characteristic feature of a neural network which enables the model to learn more advanced features and representations of the data [34]. These layer depth highly influence the network's

power, that is its ability to learn complex relationships in data. Parts of a neural network listed mark a graph of Figure 2.8.

Output Layer:

The last layer, which is the output layer is the layer which gives the final results or even the prediction classes and is based on the input data that was already transformed, which was processed by the hidden layers. The quantity of neurons contained in the output layer is a function of the number of classes which the network was targeted for the tasks [33]. For example, the output layer for a classification task where three classes have been defined would comprise three neurons. The layers of the neural network are shown in Figure 2.6.

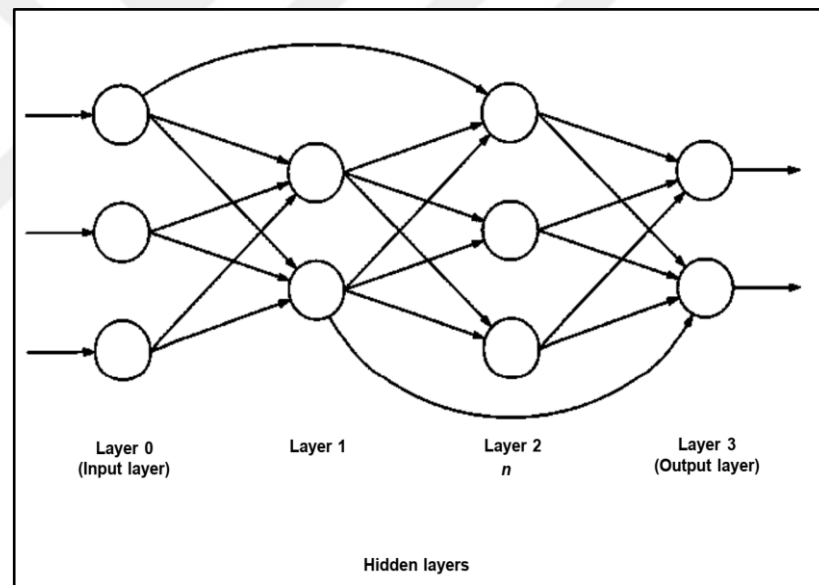


Figure 2.6. Neural network structure [32].

2.2.2.2. Neural Network Key Components (Neurons, Weights, Biases, Activation and Loss Functions)

In order to understand how neural networks work, it is critical to understand the basic building blocks that allow them to receive inputs which in turn helps them make decisions, classify information or forecast events. These are components are represented as follows:

Neurons:

A neuron is a biological neural representation and simulation unit that constitutes an ANN [35]. For every input aggregate, one or more signals are given to the neuron which processes each of them by multiplying with values called weights that influence the degree of the connections. The internal output of a neuron is also denoted as the weight. Such output is sent to subsequent layers of dots where the sigmoid function determines the transfer output along with the weighted connections (as illustrated in Figure 2.7) to construct an interactive or active neuro. The transfer function of choice converts the weighted input into the output for the given neuron. The authors describe how information in the network is transmitted using these connection weights.

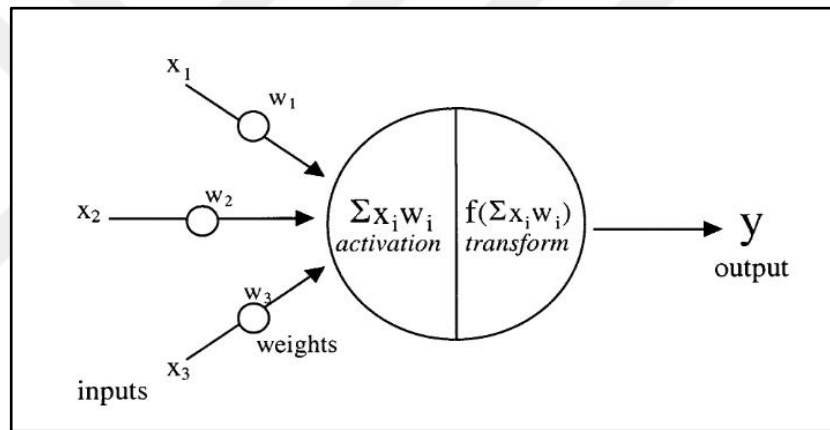


Figure 2.7. Artificial neuron model adapted from [35].

Weights:

Neural networks employ weights as the tunable parameters that establish the effect of the neuron's input towards its output. The operation at the output of a neuron is a weighted sum of the weighted inputs. As illustrated in Figure 2.7 w_1 , w_2 and w_3 are the weights assigned to the inputs x_1 , x_2 , and x_3 respectively. These weights essentially determine the importance of each of the inputs in the calculation of the neuron's output signal. To be specific, the first step occurs in the training mode, where modeling employs backpropagation and other methods to adjust the weights of a neuron to minimize the difference between the desired and actual outputs. For a given input x , its weight interacts with its input to determine whether the particular input has

a significant influence on the output of the neuron. Weights are quite instrumental in pattern recognition from the data since they determine how much importance is attached to each of the input variables in the prediction and what associations have been developed by the network [35].

Biases:

The bias in neural networks aids the model in identifying intricate trends and correlations within the data. Bias is the constant added to the dot-product of the weights and inputs prior to the activation function. This enables the activation function to move horizontally, which expands the location of the decision boundary. This shift enables the network to modify its output without being compelled to originate from the same point. Each neuron possesses its own bias, which determines their firing status according to the input signal and bias level. The inclusion of bias during the computation of neural nets enhances their performance and adaptability for learning during training resulting in better learning outcomes [36,37].

Activation Functions:

ANNs utilize the weighted sum approach in which a neuron's inputs along with weights and biases are estimated via mathematical functions known as activation functions. This is what encodes a neuron's response to the input and hence the output that the network must produce. Among others, the most common where the sum of the inputs is transformed to a value between zero and one, is the sigmoidal function, enabling the functional output of the neuron to lie between 0 and 1. This transfer function also helps to introduce a non-linear element to the network, allowing it to effectively model complex patterns in the data as demonstrated in Figure 5. In a neural network architecture, this operating property is achieved as a neural network comprises many units termed interneurons which process in parallel each input, modifying it linearly and subsequently altering the processed input through an activation function. If there were no activation functions present, it would simply translate the value of the inputs without sufficiently enabling generalization which is necessary for a neural network [35,38,39]. The role of activation functions can be summarized as follows:

- **Introduce non-linearity:** To satisfy additional conditions, incorporation of more general activation functions into the basic form enables the model to learn to approximate more complex non-linear correspondence between the input and the output.

Increase the capacity of the model: Non-linear activation functions allow the model to learn more complex patterns and relationships in the data.

Improve generalization: By introducing non-linearity, activation functions help the model to generalize better with new, unseen data.

There are diverse kinds of activation functions, each of them has its own pros and cons. Most of the time, activation functions are:

Sigmoid:

The mapping behind the sigmoid function is appropriate in the case of binary classification problems since it helps to squish the input into an output that is between 0 and 1. The specific formulation to the sigmoid function is given as below:

$$f(x) = \left(\frac{1}{1 + e^{-x}} \right) \quad (2.1)$$

ReLU (Rectified Linear Unit):

The ReLU function applies a non-linear transformation on output where negative values become all zeros and the positive values become one value, which also makes it easy to compute and frequently employed [38]. The ReLU activation is given below:

$$f(x) = \max(0, x) = \begin{cases} xi, & \text{if } xi \geq 0 \\ 0, & \text{if } xi < 0 \end{cases} \quad (2.2)$$

SoftMax:

The SoftMax function is suitable for multi-class problems since it maps the input to a probability vector across the class which is stochastic as well [38]. The SoftMax function is given as below:

$$f(xi) = \frac{\exp(xi)}{\sum \exp (xj)} \quad (2.3)$$

Loss Functions:

A loss function is a crucial aspect of machine learning models. It is the most interesting loss function in regression analysis, allowing to gauge the effectiveness of the model, as well as setting the bar for minimizing the difference between estimated and real-world results. In machine learning, loss functions are vital for optimizing a model, since they allow the model weights to be learnt at each training iteration in a way that not only makes the model better but also predicts better outcomes. It is very important for the loss function to be set up correctly, since otherwise, its misuse must lead to the poorer quality of the system being developed [40].

2.2.3. Overfitting and Underfitting

Unquestionably, overfitting and underfitting are crucial problems which arise in machine learning and deep learning, due to their strong effect on model performance and their ability to generalize. The most common situation is when a model suffers from underfitting or overfitting: these terms relate to the degree of correlation a model has between the training and test or validation datasets. Any task of the machine learning modeling is to solve the problem of good generalization on unseen data. This implies that the model should not just overfit the training data set or just underfit and learn nothing useful from it. Overfitting and underfitting are core issues that have to be resolved in order to achieve this equilibrium [41,42,43].

2.2.3.1. Overfitting

Overfitting occurs when a model focuses on the noise and details of the data rather than just the patterns. The performance of the model is perfect after training but fails on testing. Being able to fit several parameters causes deep learning models to always overfit. Such models are so complex that they no longer learn; they just find and remember the data. This causes excessive training and test errors. Overfitting occurs in cases where the model is either too complex, the training data is too small, or too much noise is present in the training data. An example of overfitting is illustrated in Figure 2.8. The model is too tightly fitted to the training data which causes it to pick the noise instead of the actual patterns. There are various strategies to cope with overfitting; weight decay, dropout, and batch normalization, as well as data augmentation are some. These include where additional constraints or variations which improve generalization are added [41,42,43].



Figure 2.8. Example of overfitting adapted from [42].

2.2.3.2. Underfitting

The inverse of fitting is the underfitting, the situation whereby a fitting is too simple relative to what is needed to explain the relevant information within the data. When the model lacks sufficient capacity or complexity to grasp the underlying structure of the data, it can also lead to high errors during training and test of the model. It is quite an easy notion to understand how and why underfitting may occur. This is use models

that have insufficient layers or number of parameters or they haven't been trained sufficiently or simply the wrong model has been selected for the data given. Understanding the difference between overfitting and underfitting is important in understanding AI. Overfitting is when a model knows the data too well or 'memorizes' it; simply put the model learns too much to be useful in the analyzing of new data. What this implies is a situation of high bias but low variance. What figure 2.9 presents is the case of underfitting and its 'classic' symptoms. It demonstrates the consequences of the over-simplistic model, which failed to consider factors that are not too insignificant in the data. As a result, the model accuracy turns out to be low on both the test and training datasets.

There are a few measures we can take in countering an underfit model. For example, model complexity can be increased, more time can be devoted to the training, or a better model can be chosen for the complexity of the task [41,42,43].

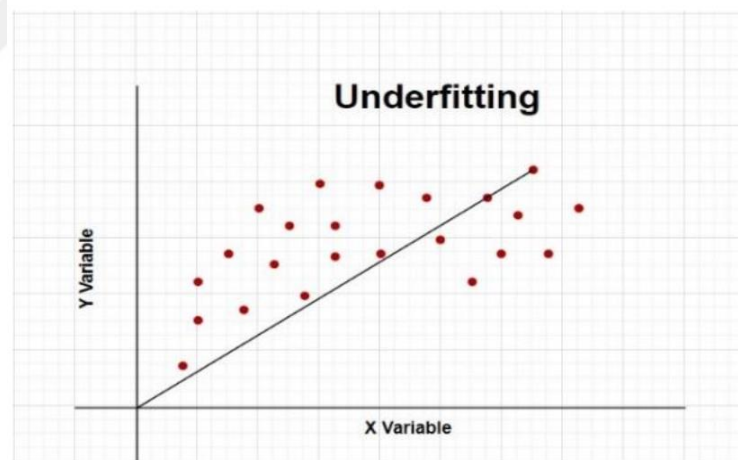


Figure 2.9. Underfitting example. adapted from [42].

2.2.4. Evaluation Performance of Machine Learning Models

In machine learning, particularly in classification tasks involving datasets with target labels, confusion matrices are a standard way of determining the effectiveness of trained models. These matrices help to see the extent to which a model can correctly classify the positive and negative cases in a given test dataset. Table 2.1 represents the Confusion matrix overview.

Table 2.1. Confusion matrix overview.

Actual Positive	Actual Negative
True Positive	False Positive
False Negative	True Negative

The confusion matrix provides a detailed breakdown of a model's predictions:

- True Positive (TP): Correctly predicted positive instances.

True Negative (TN): Correctly predicted negative instances.

False Positive (FP): Incorrectly predicted as positive when the actual class is negative.

False Negative (FN): Incorrectly predicted as negative when the actual class is positive.

To measure the efficiency of classification models, various metrics such as precision, recall, F1-score and accuracy are common. Such constraints assist in evaluating the performance of a model in predicting the positive and negative cases. Below is a detailed explanation of these criteria:

- Precision (P): Precision measures the proportion of true positive predictions among all positive predictions made by the model.

$$P = \frac{TP}{TP + FP}$$

Recall (R): Recall calculates the proportion of actual positives correctly predicted by the model.

$$R = \frac{TP}{TP + FN}$$

F1-Score: The F1-Score combines precision and recall providing a balanced measure of performance.

$$F1 - score = 2 \cdot \frac{P \cdot R}{P + R}$$

Accuracy (A): Accuracy measures the overall correctness of the model's predictions.

$$A = \frac{TP + TN}{TP + TN + FP + FN}$$

2.3. OVERVIEW OF DEEP LEARNING

2.3.1. Introduction to Deep Learning (DL)

Deep learning, an advanced subset of machine learning, is transforming the field of AI through its ability to enable computers to train on and learn from large data sets in ways that emulate human perception [44]. The defining characteristic of deep learning is its ability to automatically extract features from data without the need for manual intervention. This capability enables systems to learn complex patterns, which differentiates it from machine learning, which is essentially multi-layer neural networks. This technique has yielded enhanced capabilities for detecting intricate representations in images and processing natural language [45,46,47].

AI encompasses ML and DL concepts, as depicted in Figure 2.10. The concept of DL is a subdivision of ML which empowers its ideas [48]. In classical approaches to machine learning, algorithms utilize manual feature extraction frequently and more simple neural networks are limited on layers. However, with deep learning, the neurons, activation functions, loss functions, and optimization methods used in the model enable it to learn from the data.

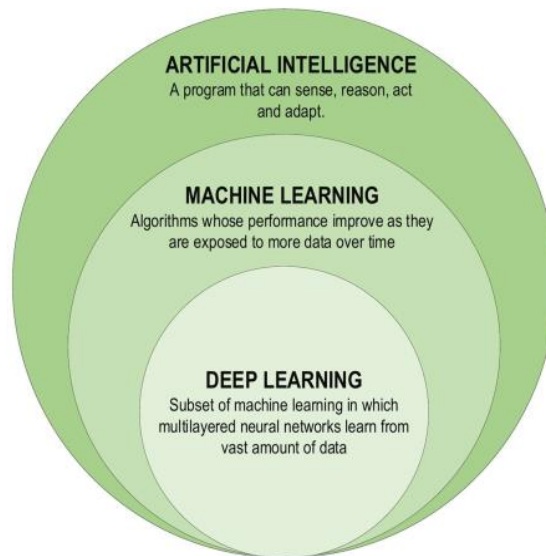


Figure 2.10. Family of AI adapted from [41].

It's worth noting that deep learning is even more effective as it incorporates a multilayered neural network structure allowing the model to detect intricate patterns in unprocessed information [41]. Deeper structures improve performance especially when large amounts of data are involved and motivate more recent developments in many areas, for example image processing, natural language understanding, bioinformatics, control systems, and medical information processing, etc. On the other hand, deep learning is distinguished from traditional machine learning ML techniques in that it can do the job of automatically learning sets of features that describe the tasks at hand. With DL such a complication is avoided as learning and classification can take place in a single operation. Figure 2.11. The differences between ML and DL as levels of processing boxes are portrayed in the map [41].

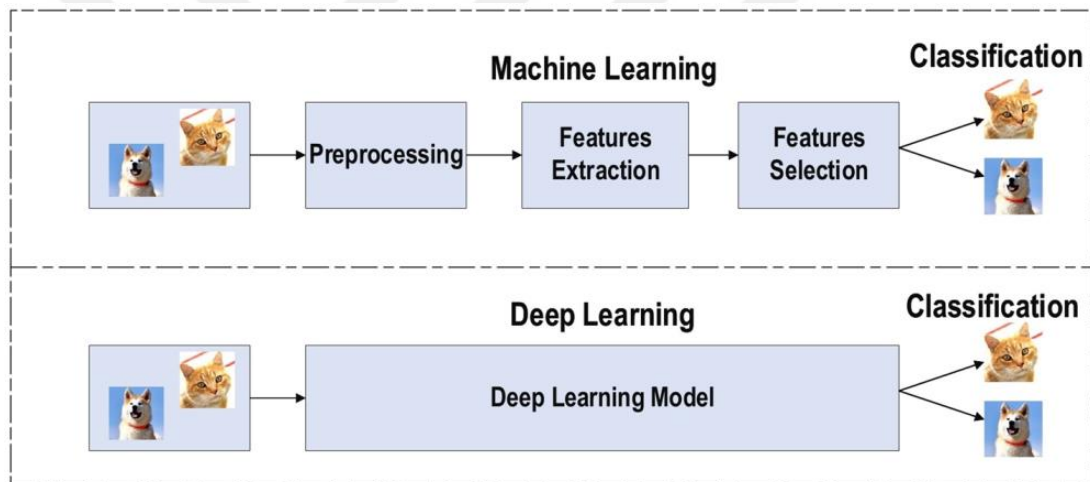


Figure 2.11. Differences between machine learning and deep learning steps adapted from [41].

In deep learning, multiple models of deep learning have been developed with the aim of solving different types of applications and tasks. These architectures include CNNs for image-related tasks, recurrent neural networks (RNNs) for sequential data, long short-term memory networks (LSTMs) for time series and language processing etc. [41]. However, in light of the fact that our research is concerned with the processing of images, we will be concentrating more on the explanation of CNNs as they are defined and used in the present work.

2.3.2. Convolutional Neural Networks (CNNs)

CNNs are a class of deep learning models which are well suited for applications involving data that can be represented in a matrix form, for example images [49]. The emergence of a follow-up, which now appears to be misleadingly termed, a convolutional neural network, seems to be genetically related to the functioning of the neurons of visual cortex of animals [41]. CNNs were initially introduced to the world of science by Fukushima in 1998 [50] and gradually, these have found applications in a gamut of fields such as image and speech recognition, natural language processing (NLP), and many more [41,51,52].

The architecture of a CNN consists of multi-layered structures that serve the role of learning and extracting features from the data in a hierarchical manner. This includes convolutional layers, pooling layers, as well as fully connected (FC) layers as depicted in Figure 2.12. [41,49,50]. Convolutional layers are vital in feature extraction. These layers utilize several filters, or kernels, that are moved over the input data in order to identify spatial patterns, thus, enabling the extraction of local relationships within the data [47,49,50]. These operations yield a set of feature maps depicting the directions in which various patterns found in the input data are oriented [47].

A fundamental advantage of the CNNs is their use of local connections and shared weights [41,47]. So, any filter existing within a convolutional layer can be used for many differing locations of the input, hence making it possible for the CNNs to recognize features irrespective of their location [41,47]. This property of being translation invariant, together with the fact that parameters are shared, allows for efficient pattern recognition with less memory and less computation, unlike the standard FC networks [41].

The structure of the CNNs is hierarchical and adaptive which facilitates the extraction of both low-level features and high-level features and is one of the reasons as to why they have become a core technology in the field of computer vision. Their validity has been proved across multiple areas including, but not limited to, face recognition systems, autonomous vehicles, medical imaging, etc. [41,52].

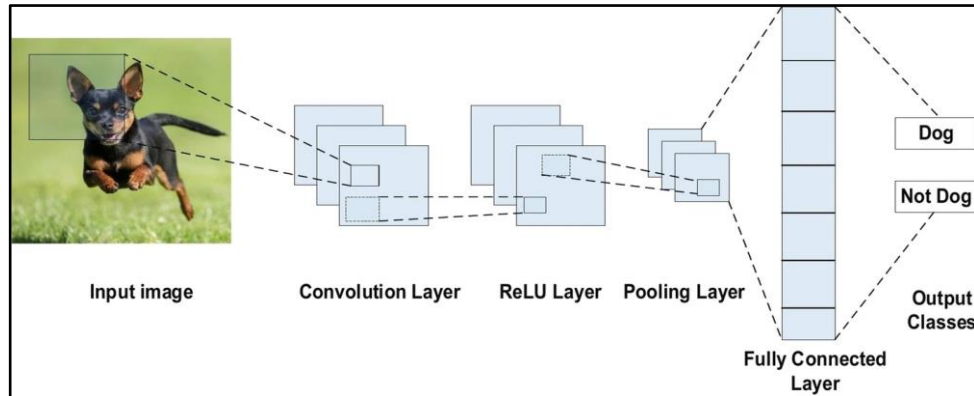


Figure 2.12. CNN layers adapted from [41].

2.3.2.1. Convolutional Layer

The typical structure of CNN incorporates a convolutional layer which performs the important role of feature extraction by means of applying trainable filters or kernels across the spatial dimensions of the input data [49,50,53]. This layer is preceded by a sequence of linear and non-linear operations wherein said layer begins with a convolution operation that takes the inner product of the input and filter values resulting in the production of a 2D activation map [50,54]. The function of the convolutional layer is illustrated in Figure 2.13.

Various regions of the input are processed by a convolutional filter via a sliding window in a specified stride and padding which leads to the formation of localized feature representations which are also referred to as receptive fields as a result of the convolution process [54,55]. Such filters are generally intended for targeting particular features or edges within the input at specific locations so that the CNN can later identify such features in the images or other forms of structured data [49,50,56].

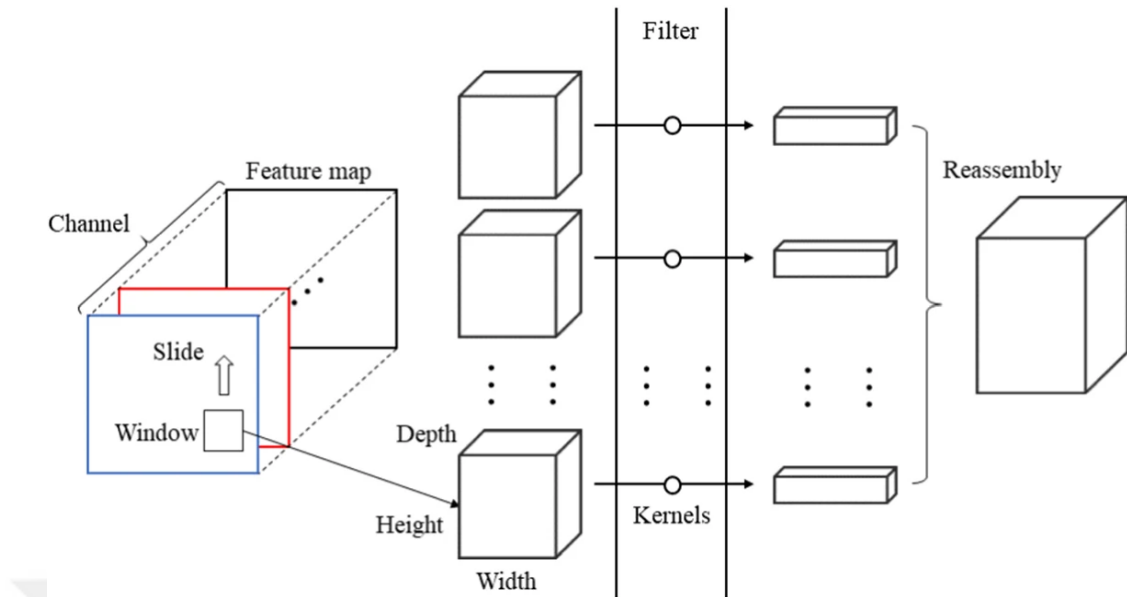


Figure 2.13. Convolutional layer mechanism adapted from [53].

The parameters of the convolutional layer include hyperparameters such as filter depth, stride and padding, which modify the size and the number of output feature maps [55]. Each convolutional layer is often followed by an activation function, e.g., ReLU, which adds nonlinearities and permits the network to discover complex correlations in the data [54]. Due to the prominent performing hierarchical decompositions of feature representations associated with the repetitive use of convolution and activation functions, CNNs are superior to pattern recognition [50,53,55]. The following definitions for each component from CNNs: Filters (or kernels), Strides, Padding and feature map generation permeates how a convolutional layer works and what output it produces.

Filters/Kernels

A kernel in CNN is a set of discrete values representing a weight which is small on the grid and is used for feature extraction [41,49]. Initially, the weights located within the kernel are assigned randomly. However, as the network trains, these values are altered as the network learns what patterns or features are critical in the input data [41].

The operation of the kernel involves moving across the input tensor which is a structured number array and calculates the scalar between its values and corresponding

values in the input. Such processes produce output in the form of the 2D activation map which indicates areas where the kernel was able to feature find specific elements within the data [49,55]. Kernels are small usually in spatial dimensions, but their depth suits the input data which allows the kernels to find both shallow and deep level features in the layers [55]. The kernel with a dimension of 3×3 is depicted in Figure 2.14 as it moves on top of the input tensor.

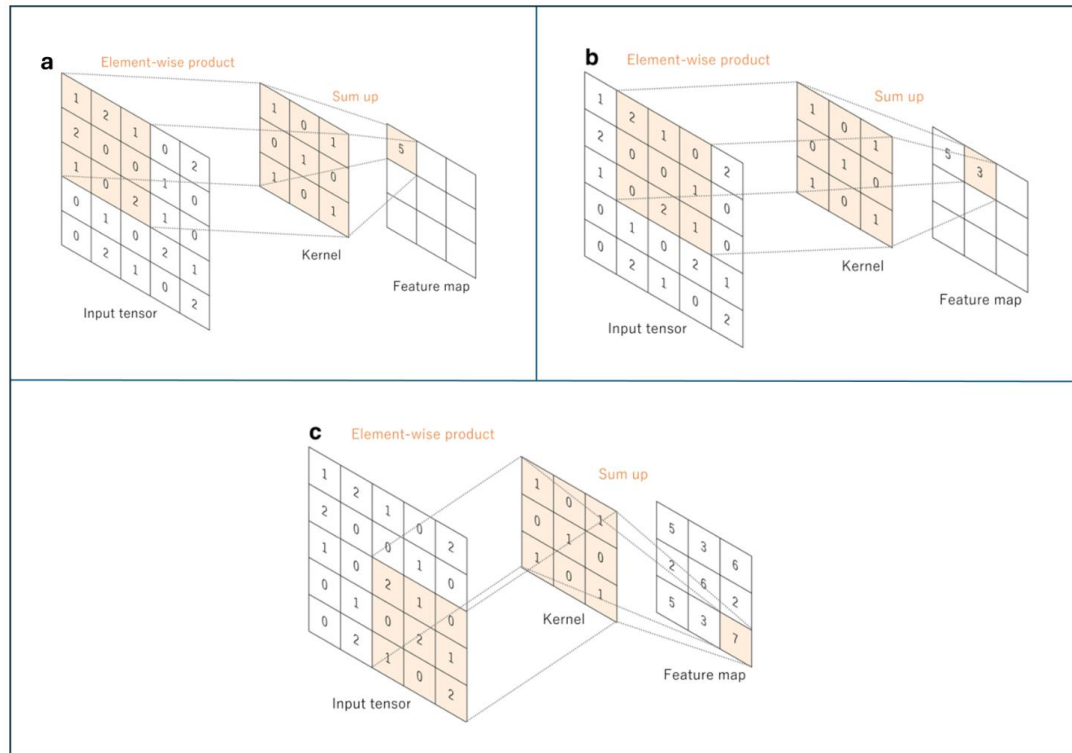


Figure 2.14. Kernel size of 3×3 adapted from [49].

Strides

The stride indicates the advancement or length of a single step performed by the filter while accessing the input data. A stride of 1 means the filter moves a single pixel of maximum intensity as depicted in Figure 2.16 from (a) to (b) [49]. In a stride of 2, a two-pixel intensity is advanced at a particular point. A big stride increases the output feature map's size properties but decreases its detail level and lowers the processing requirements. Thus, it seems straight forward to determine the optimum value of computational efficiency and detail by selecting an appropriate stride.

Padding

Padding is the operation of extending input data with additional rows and columns of pixels around the border before filtering. This ensures that the spatial dimensions in the resulting feature map are equal to or only slightly less than those of the input depending on the type of the chosen padding. The practice of convolution making addition of zeros around edges which is known as zero padding is the finest technique for preserving information owing edge focus. If padding is not done the filters tend to compress the spatial dimensions which leads to a tremendous decrease in the size of the feature map after several layers [49,55]. Figure 2.15 illustrates the zero padding around the original input pixels.

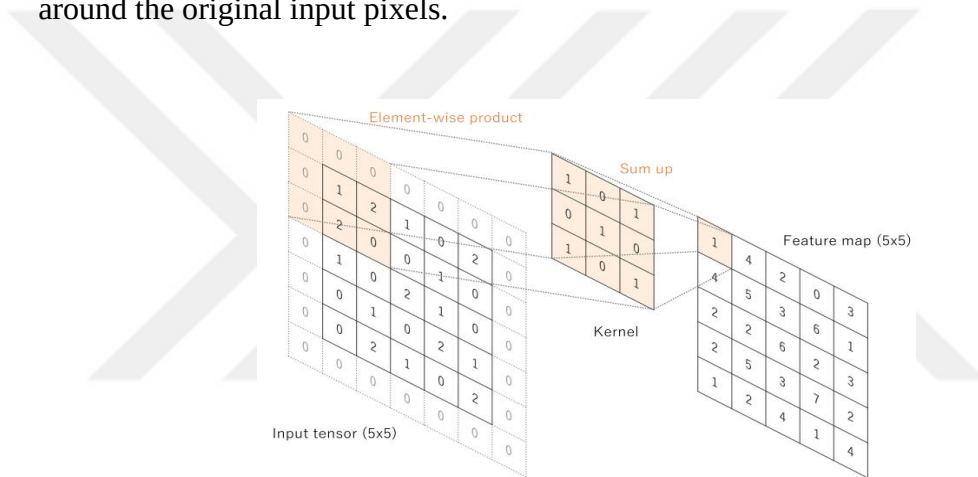


Figure 2.15. A zero padding around the original input pixels adapted From [49].

Feature Map Generation

The responses delivered by every single filter throughout its pass over the Input are represented visually in what is referred to as a Feature Map, which is produced after a convolution layer. The results of kernel scanning for the input that generates feature maps are observable in Figure 2.14.

2.3.2.2. Pooling Layer

A pooling layer is an important concept in the architectural design of CNNs since they help to decrease the number of spatial dimensions of the feature maps that are produced

by the convolutional layers. Having so many parameters and computations in the network is other no. less demanding as it economizes the necessary resources [49,50,54,55].

A pooling layer gathers information from neighboring neurons in proximity and coverage of an area and compresses the area's outputs into one number. Max Pooling that takes the maximum value plus average pooling are common pooling functions. It was also noted that pooling layers allow the model to work with small variations in the input data and displacement of features. This eliminates overfitting by making it easier to detect common features thus decreasing the complexity of the model. Figure 2.16. shows the Pooling Layer (Maxpooling).

It was suggested that most pooling operations involve a $[2 \times 2]$ kernel size happening at a stride of two, thereby maintaining the depth of the input and at the same time shrinking the activation map by 25% [54,55]. However, there are also limitations in pooling, especially spatial information, hence trade of effective rate and level of details available [41,56].

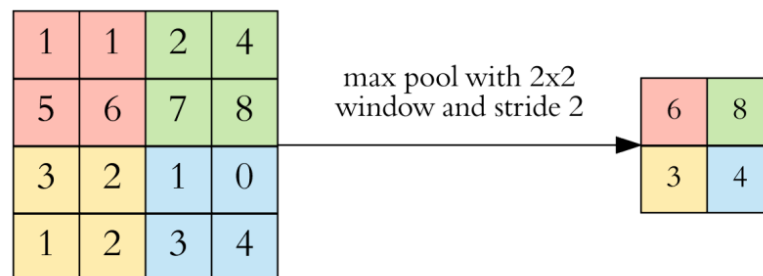


Figure 2.16. Pooling layer (Max Pooling) adapted From [56]

2.3.2.3. Dropout Layer

Dropout regularization is a method of regularization in a neural network which consists in withholding every random group of neurons – set of fractions of neurons to be ignored or turned off when training the graph. In Figure 2.17 (a) the standard neural network is displayed, while the (b) shows after applying the dropout. This process helps in overfitting by less dependence of the network on the particular sets of neurons

and helps in the learning of more common and less sensitive features instead [57]. The process referred to as dropout, the removal of both hidden and visible units along with their connections, effortlessly creates an ensemble of different network topologies. This step also helps towards the generalization of the network, addressing a key point of machine learning [57]. To strengthen the model along the neural network at each epoch the model drops as a group increasing its stresses multi paths within it's extremely structured unit [51,57].

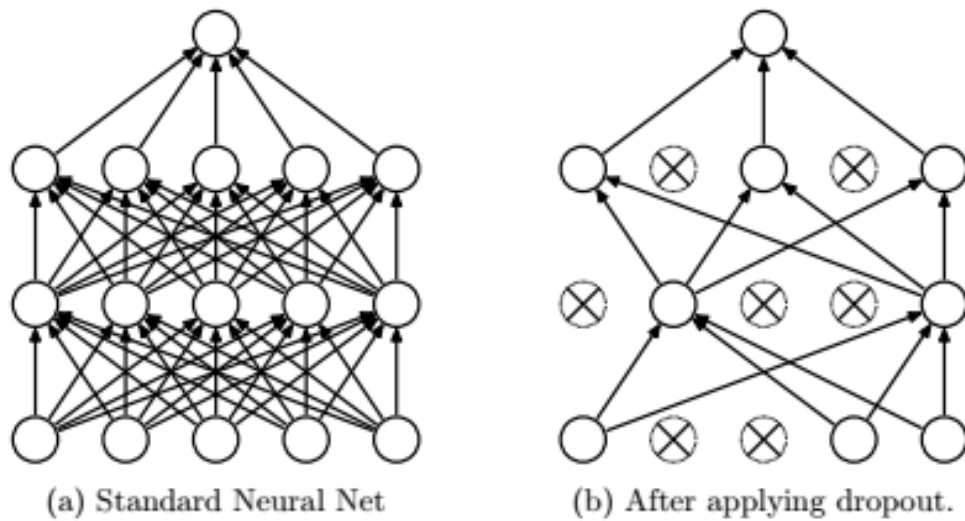


Figure 2.17. (a) is a standard neural network, (b) after applying the dropout adapted from [57].

2.3.2.4. Fully Connected Layer

Fully connected layer works as a classifier which fuses or combines inputs originating from previous layers outputs and is usually located at the end of a CNN. For every neuron from the previous layer, there is a connection or link between every neuron from the current layer; this is what is generally referred to as the FC layer [41,50,51]. Similarly, the FC layer is similar to a conventional multi-layer perception (MLP) in that it too uses a feed-forward structure to perform activations via matrix multiplication followed by the addition of a bias [50,53].

As the name suggests, the last pooling or convolution layer of a neural network produces the feature vector which serves as the input of the FC layer, this input

undergoes a process termed flattening [41]. It involves taking away the spatial dimensionality of multi-dimensional feature maps provided by any of the previous layers and forming a single classification vector of one dimension containing the mapping [53]. Finally, what comes out of the FC layer is what gets the final touch or is the final output or classification of the model which is the CNN as seen in the discourse of [41,50]. The output generated by the FC layer constitutes the final result of the CNN, as depicted in Figure 2.18.

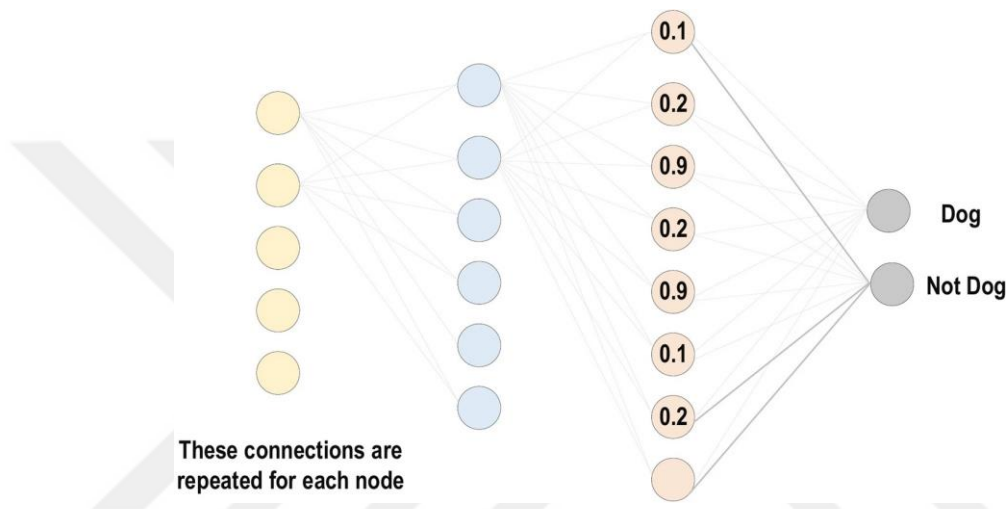


Figure 2.18. FC layer as represented final result of CNN. Adapted from [41].

2.3.2.5. Non-Linearity (Activation Functions)

Non-Linearity (Activation Functions) Neural networks, particularly CNNs, need nonlinear functions. They permit the model to convert the input signals into output signals in non-linear fashion. This is very important; or else, many processes in reality which are complicated cannot be modeled using linear transformations [51]. The nonlinear functions allow the model to learn more complex interactions and dependencies in the data, improving its expressiveness and capabilities to model more functions in general [41,51].

A nonlinear function is designed to follow the layers which have trainable weights like FC layers and convolutional layers – they constitute the non-linearity of the whole CNN architecture [41]. Non-linear functions however add more complexity for them

to work. They possess the special feature of being differentiable which is important in facilitating error back-propagation in network training [41].

The most widely used nonlinear activation functions are the sigmoid (or logistic) function, the hyperbolic tangent (tanh) function and the rectified linear unit (ReLU) [50]. The clear winner among the activation functions used in CNNs is the rectified linear unit (ReLU) owing to its simplicity and being very effective. ReLU enables to cut down all values less than zero to those which are zero; this assists in reducing the restrained gradient issue and enhances the speed of the training procedure [38].

2.3.2.6. Backpropagation

Backpropagation is a significant training technique it allows the network to self-learn by changing its parameters called weights and biases so that the errors in its outputs are minimized [49,50]. The approach proceeds by first deriving the gradient of the loss function in relation to every weight in the network. It then applies those weights with the direction given by the calculated gradient so that the model is enhanced in many rounds [49].

2.3.3. Limitations of CNNs

CNNs have numerous advantages; however, there are salient limitations that should be taken into account while implementing them. One of the major concerns is their applicability only in situations where a huge amount of labeled data is available. CNNs have millions of parameters, so a large amount of variation in data is needed to achieve minimum risk of overfitting. Gathering and labeling this enormous data set can be a daunting and expensive exercise, especially in instances of fine-grained classification or segmentation tasks [51]. Further, it has been noted that due to the down-sampling operation (pooling), CNNs lose important spatial information, as they discard information about an object's precise pose and location [52]. They also have problems with a lack of internal rotation, therefore making it hard for them to correctly label the objects when they are in different orientations from those trained. Also, since CNNs are sensitive to even the slightest perturbations of pixels, they can easily misclassify

and suffer adversarial attacks. These challenges raise the areas where the use of CNNs poses limitations, more so in areas where spatial relations, efficiency and robustness matter [51,52].

But as reinforcement of learning becomes easier, there have emerged deep learning techniques such as CapsNets that offer promising insights to go above the traditional cons of CNNs, including the loss of important spatial structures between the elements in the hierarchy [12]. In the next part, the concept of CapsNets will be described in detail.

2.4. CAPSULE NETWORKS: CONCEPT AND KEY COMPONENTS

2.4.1. Introduction to Capsule Networks

As described in 2017 by Sabour et al., CapsNet are capable architectures which can be handy for problems that require significant amount of spatial attention such as image recognition [12]. This is because they address some of the limitations that are naturally present in the classical neural networks.

It can be viewed as a state-of-the-art neural network structural design in which multi-dimensional spatial hierarchies in relation to the data are compiled and encoded in “capsules” instead of the conventional neurofunctional units that has a single output. In the case of CapsNet, the core concept refers to an image representation method that aims to overcome this limitation by treating visual information as a complex hierarchical structure composed of cosets. In general, the traditional neural networks operate on the principle that its basic functional unit is the neuron which provides scalar outputs. The output is a multi-dimensional vector that represents a cluster of neurons and the unit cell at this level is a capsule. This vector provides information concerning various characteristics of the given object or a component of that object such as its position, orientation and scale. Thus, the network can understand the different level of image components discrimination and image segmentation. For instance, the length of the capsule’s output vector will denote the probability that an object is present while the direction of the vector carries other spatial information [12].

Capsules are considered to assist these networks, known as capsule nets or real-world nets, in perceiving aspects in an orderly fashion and preventing the effect of changes in pose, texture or any distortion. For example, in the case of object recognition tasks, it is not only the features that caps networks sense but also the spatial position of those features, hence the strain of the degree of positional accuracy is relieved. Because of the parts of an object were moved around, there would not be as many confusions caused by object misclassifications. The strength of this capability is further magnified with the introduction of the dynamic routing mechanism that allows the nesting of pods with the lower-level output sent to a higher-level pod when there is a strong spatial relation between them. This improves the sensibility of the network [12].

2.4.2. Capsule Network Architecture

A base structure of a CapsNet proposed by Hinton Geoffrey et al. in 2017 relies on the developments that appeared concerning the representation of an object or an object part by activity vectors [12]. They employ the length of the activity vector to indicate the existence of the entity and the direction of the vector to indicate the parameterized formation of the objects. This is done to ensure that the output vectors are in the range of 0 and 1 which is affected by the equation which is Non-linear squash. As given below:

$$v_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \frac{s_j}{\|s_j\|} \quad (3.1)$$

In accordance with equation 3.1, the output vector of capsule j is denoted by v_j , and s_j is its total input. Although, s_j remains the same for the first one, the prediction value of all pillars for that case is as shown in (3.2) a weighted summation of the prediction vector $\hat{u}_{j|i}$ from the last pillar. Also note that the output of u_i from the lower-level capsule is multiplied by the weight matrix W_{ij} and the resultant vector is the prediction vector $\hat{u}_{j|i}$.

$$s_j = \sum_i c_{ij} \hat{u}_{j|i}, \quad \hat{u}_{j|i} = W_{ji} u_i \quad (3.2)$$

In the model described, the linking coefficient, referred to as c_{ij} , is crucial and is determined by means of a dynamic routing algorithm which is undertaken in a recurrent manner over time. Specifically, these link factors, which evaluate the strength of a link between a certain lower-level capsule, i and all higher-level capsules, are bound by the SoftMax function to add up to one. The probability that capsule i will interconnect with capsule j , named as b_{ij} , plays an important role in establishing these link factors and is naturally a part of the model as expressed in the equation (3.3).

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \quad (3.3)$$

The two capsules' position and type establish the initial probability which is then updated with the weights during the learning phase, which is defined as the input image [12]. Therefore, the initial link of factor is changed to incorporate the output v_j as of the higher-level capsule, to $\hat{u}_{j|i}$ be the prediction made regarding the capsule i . Each feature is then recognized, and parts of the object are organized into a hierarchy using dynamic routing., $a_{ij} = v_j \cdot b \hat{u}_{j|i}$ programmed in a scalar value which is later added to b_{ij} so that the new link factors can be computed. Dynamic routing can be seen in Table 2.2.

The initial architecture of the capsule neural network had a Conv2D layer consisting of 256 filters of size $[9 \times 9]$ with a stride of 1. The output produced $[20 \times 20 \times 256]$ was as a result of feature maps produced by the “ReLU” activation function. Thereafter, a primary capsule layer utilizes 256 filters which consist of a convolutional kernel of size $[9 \times 9]$ with a $[2 \times 2]$ stride resulting into 32 channels of 8D capsules. The outputs were slotted into the $[6 \times 6 \times 32]$ dimensional space, where it consisted of 32 units with each unit having an 8D output capsule. The third layer consists of a digit capsule layer whose neurons are FC to all the capsules present in the next layer. This layer has 10 capsules, one for each digit and each capsule is represented by a $[16 \times 1]$ vector which was obtained from the capsules obtained in the previous layer in order to do image classification as shown in Figure 2.19. This layer executes the function of determining the bias of each capsule, which is useful to indicate the likelihood of the

presence of the object and can be applied as the probability for the classification result [12].

The authors mentioned, the CapsNet model focuses on a peculiar decoder structure, which aims at reconstructing the digit images from the representations contained within the DigitCaps layer. The MNIST database has long enjoyed the position of an industry standard in the sphere of machine learning and consists of 70,000 scanned, gray-scale images of handwritten digits from 0 through 9, each one having a resolution of $[28 \times 28]$ pixels. This dataset is suitable for testing the capabilities of models trained to classify images which involve recognizing images of handwritten digits [12].

Table 2.2. Dynamic routing algorithm [12].

Procedure 1 Routing algorithm

```

1: procedure ROUTING ( $\hat{\mathbf{u}}_{ij..}$ ,  $r$ ,  $l$ )
2: for all capsule  $i$  in layer  $l$  and capsule  $j$  in layer  $(l+1)$ :  $\mathbf{b}_{ij} \leftarrow 0$ .
3: for  $r$  iterations do
4: for all capsule  $i$  in layer  $l$ :  $\mathbf{c}_i \leftarrow \text{softmax}(\mathbf{b}_i)$ 
5: for all capsule  $j$  in layer  $(l+1)$ :  $\mathbf{s}_j \leftarrow \sum_i \mathbf{c}_{ij} \hat{\mathbf{u}}_{j|i}$ 
6: for all capsule  $j$  in layer  $(l+1)$ :  $\mathbf{v}_j \leftarrow \text{squash}(\mathbf{s}_j)$ 
7: for all capsule  $i$  in layer  $l$  and capsule  $j$  in layer  $(l+1)$ :  $\mathbf{b}_{ij} \leftarrow \mathbf{b}_{ij} + \hat{\mathbf{u}}_{ij} \cdot \mathbf{v}_j$ 
return  $\mathbf{v}_j$ 

```

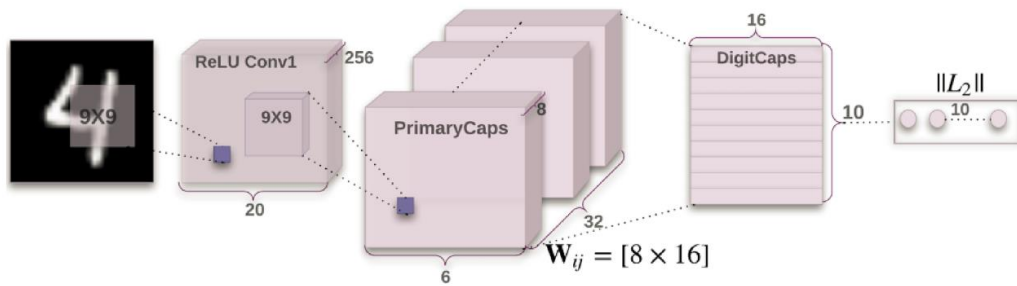


Figure 2.19. Original CapsNet structure adapted from [12].

Before the Decoder section there is a section called The DigitCaps volume which contains 10 capsules, each of which is responsible for the prediction of a single digit class. Each capsule is a directional 16 vector which is responsible for capturing the basic characteristic of its digit class. At the time of decoding, all other non-predicted

capsules are made zero and the only non-zero capsule used corresponds with the predicted class. This selective masking ensures that the decoder bears emphasis on the reconstruction of features associated with the predicted number only, with less disturbance from other classes thus improving the quality of the reconstructed image [12].

The decoding itself is composed of three FC reconstruction layers, this means that layer one in the decoder has 512 and layer two has 1024 units respectively, with all layers using “ReLU” activation so that’s growth and refinement feature space, letting the network get esoteric details of the digit. And the last layer is 784 units which correspond to the input of $[28 \times 28]$ image dimension with the use of Sigmoid activated function in order to output only values close to 1 and 0 making it suitable for binarization. During training the model induces the length of the instance vector with the help of a custom loss function such that for the correct capsule its representation can be close to 1 while for the rest the ideal is close to 0, which is valid as the input data is binary. This structure for combined reconstruction and classification helps the CapsNet form better and more useful and explainable capsule encodings. In the same sense it’s worth noting the caption appearing on Figure 2.20. constructing a digit from the digit caps layer representation [12].

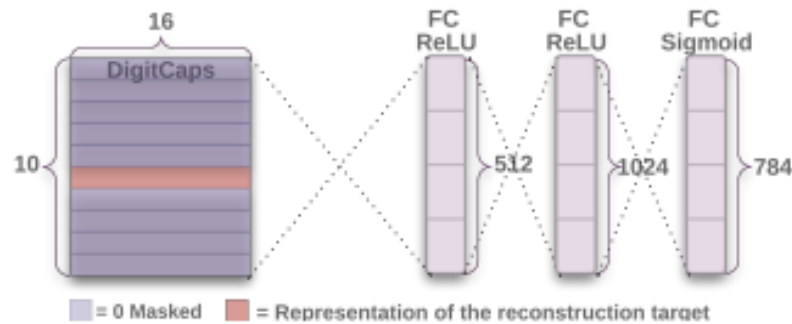


Figure 2.20. CapsNet decoder structure adapted from [12].

2.4.3. Limitations of Capsnet in Medical Image Processing

Even though CapsNet indicates a shift in deep learning and solves some of the drawbacks of the existing CNN models, this still has some constraints which makes it

difficult for medical images or images with complicated features to get stronger outputs. Such factors are identified and illustrated as follows:

- **Incompatibility with Medical Images:** Initially, CapsNet was designed to work on the MNIST dataset, which consists of basic binary images of low resolution such as handwritten digits [12]. This dataset is simple and does not have the deep and conveyance representation abstraction that other kinds of images, especially medical images, contain. In general, medical images have richer feature representations, high-quality images, and complex pixel information which again is in stark contrast with the MNIST dataset. Without adaptations, CapsNet fails on the other hand, because the complex representation has not been intended for the operation with such rich feature datasets in the first place [58,59].

Need for Design Adjustments: Medical datasets tend to have a small number of high-resolution images and high feature complexity compared to MNIST which contains an excessive number of simple images. Several experiments point out that original CapsNet was not developed for small size intricate datasets and was required to be modified in order to acknowledge the intricate characteristics of medical imaging [59].

However, numerous researchers are presently looking for ways to enhance the structure of CapsNet. Medical image processing is more effective when enhanced modifications and changes have been made as per several research articles in this area and in order to overcome these challenges

PART 3

RELATED WORK

The focus of recent work has been on improving the CapsNet architecture and training algorithms for the purpose of medical imaging. These advancements are intended to expand the scope of issues that the CapsNet can solve, such as those involving high resolution images, intricate textures, and small differences that are common in medical data sets. Strategies like adding hybrid architecture, changing routing approaches, and fusion of multiple mechanisms have been employed to overcome these hurdles, thus making the CapsNets more versatile and resilient for the Clinical purposes. This subsection analyzes the recent developments in CapsNets that have led to better results and performance in the task of medical image processing.

Toraman et al in their work incorporated convolutional CapsNet which is supposed to reveal the abnormalities of COVID-19 from the X-ray edges. They augmented the original CapsNet by adding multiple convolutional layers in front of the primary capsule layer to improve on feature extraction. The network architecture therefore involved four convolutional layers arranged in the following order of internal components: Conv1 has 16 $[5 \times 5]$ filters with a stride of 1 and a $[2 \times 2]$ max-pooling layer afterwards; Conv2 has 32 $[5 \times 5]$ filters with a stride of 1 and a $[2 \times 2]$ max-pooling layer afterwards; Conv3 has 64 $[5 \times 5]$ filters with a stride of 1; and Conv4 has 128 $[9 \times 9]$, stride 1 filters. After the convolution layers, the primary capsule layer contains a total of 32 capsules, each with a $[9 \times 9]$ kernel and a stride of 1. The LabelCaps consists of 2D and 3D classification tasks in two sets of 16 capsules while applying ReLU activation function. Due to the small images size, the model was fitted on augmented images on a binary and multi-class settings that reached the accuracies of 97.24% for the binary classification and 84.22% for the multi-class classification respectively [60].

A model for classification and detection of lymph nodes in medical images was proposed by Zhang et al., they called it Camelyon16 capsule network model. They improved the routing process by feature decomposition whereby the feature vector and feature value obtained from decomposition stood in form of output vector and predicted value in the standard routing process making it a new routing schema. As they knew the shortcoming of a kernel of fixed size, they contacted a multi scale feature extraction module just before the first capsule layer. This module includes several convolution and pooling layers with the aim of obtaining different spatial features. Firstly $[1 \times 1]$ dimension reducing convolution is applied and then $[3 \times 3]$ and $[5 \times 5]$ convolutions are used to capture larger scale patterns later. Also, because the dominant features need to be focused on and the input down-sampled, a $[2 \times 2]$ max-pooling layer was added. These outputs are then concatenated aimed at yielding a shifted multi-scale representation which is more informative. They reported accuracy results of 82.31% for CIFAR-10, 99.65% for MNIST, and 88.23% for Camelyon16 when the model was tested on these three datasets [61].

Afriyie et al. conducted a study which sought to test the first version of CapsNet on several datasets namely R-MNIST, CIFAR-10, MRI images of the brain tumors, and blood cells. From their experiments, the results achieved are as follows: profiles of 62.98% in the case of CIFAR-10, 78.69% in the case of brain tumor MRI, 92.49% in the case of the F-MNIST, and 82.00% in the case of blood cell images. The results show the shortcomings of the original CapsNet in general, and in particular, its ability to work with complex picture dataset-medical image dataset. Because of this, the focus of many researchers is on the improvement of technology, so that it better corresponds to the requirements of different data types. The subsequent sections summarize the studies that have aimed to alter and adapt the CapsNet for use in medical imaging based on various other datasets. [62].

Afshar et al. have created a modified CapsNet for the purpose of diagnosing brain tumors using brain MRI scans. In their case, the changes were not restricted only to the architecture but also implemented in the loss function. The authors resized the brain MRI scans from $[512 \times 512]$ to $[128 \times 128]$. The architecture begins with a convolution layer that has 64 filters with a kernel of $[9 \times 9]$ which aims at feature map extraction.

They then constructed a primary capsule layer which contained 32 capsules with each having eight dimensions (8D) and a $[9 \times 9]$ kernel. Rather than the standard methods of CapsNet which utilize Margin Loss and a reconstruction loss, Authors suggested using a novel loss function that was constructed by integrating capsule loss and cross entropy loss in order to boost the overall classification accuracy. Based on their conclusions, the designed CapsNet surpassed other methods aflame in terms of accuracy reaching 90.89% in correctly diagnosing brain tumor [63].

In a recent research work, Afshar et al, improved the CapsNet for classification of the tumor types further. Their improved version of model contained only one convolutional layer with 64 filters instead of 256 as it was in the previous versions and took images which were resized to $[64 \times 64]$ pixels. The initial capsule layer, which consisted of 32 capsules of eight dimensions each was unchanged and had kernels of size $[9 \times 9]$ with a stride of 2, areas of which were pooled to form a $[24 \times 24]$ feature map. Their CapsNet modification outperformed the original CapsNet accuracy of 82.30% and achieved 86.56% accuracy. Furthermore, they applied models with two convolutional layers filter, each having 64 models, which had an accuracy of 81.97 percent. In order to obtain better results, they have improved the reconstruction Internal Neural Network by augmenting it with an additional FC layer of the reconstructed neurons large 4096. They have also tested three-layer models in the decoder with 1024, 2048 and 4096 neurons to achieve accuracy of 83.93% [64].

A number of researchers have explored the fusion of pre-trained CNNs with CapsNet aiming at developing hybrid frameworks that resolve certain weaknesses of the primary CapsNet while obtaining better results. In this context, Mandic et al. have proposed a hybrid model - DenseCapsNet that integrates DenseNet121 model and a capsule neural network (CapsNet) with the aim of Covid-19 infection detection from chest X-ray (CXR) images. This approach achieved an accuracy of 90.7%. The segmentation network TernaUSNet was used to mask the lung region from each image and the subtracted images were input into DenseNet for lung feature extraction. In order to enhance the performance of the model further, random data augmentation was performed. During the span level using CapsNet, the spread loss function was utilized in order to decrease the training sensitivity. The paper claimed that the coverage of

this aspect of the issue is still quite poor as it was focused on the results of experiments when data augmentation and model training and coupling of CapsNet and CNNs were present [65].

As reported in a study by Sezer et al., a novel twist to the CapsNet was advanced for rotator cuff pathologies classification in the computer-aided analysis of shoulder MRI. This enhanced architecture incorporated two convolutional layers into the original CapsNet. The first convolutional layer had 128 filters with kernel size $[7 \times 7]$ and padding stride 2, yielding a reduced output of $[29 \times 29]$ without any formal padding. The output from these filters served as the input in the second CNN layer which comprised 256 filters with $[9 \times 9]$ convolution kernel and a stride of 1. The convolutional networks comprised of two layers where the first one was linked with the other which was passed to the primary capsule layer. The parameters of the primary capsule layer were different to the previous layers which were altered by the authors. They employed a total of 64 capsules of 8 dimensions each with $[9 \times 9]$ conv kernel and stride of 4. As far as the reconstruction is concerned, a decoder block was applied with 512, 1024, and 4096 layered neurons that aim to restore the given images. Inception along with new appendages applied to the word-oriented MRI data in coronal plane obtained 95.36% accuracy whereas a modified version of the CapsNet associated with no inception deleted 94.75%, together with sustenance thumb etc. [66].

As part of their work, Anant et al. developed a reconfigured CapsNet explicitly targeted at pediatric bone age assessment. They adapted the model for the content comprising the RSNA Pediatric Bone Age Challenge (2017) data set of hand X-ray images of children which are over 10 GB in size. The researchers have introduced a number of improvements over the originally designed CapsNet. First, they reduced the loss function to three using a specific formula. Moreover, they sought to address the ‘dying ReLU’ issue by substituting the standard ReLU with Leaky ReLU. ReLU has mathematical properties which mean that when a neuron has an output of zero, then that neuron is inactive since there is no way for the weights to change, which can be referred to as ‘turning off’ the neuron. Instead, if Leaky ReLU is applied, then the researchers hoped this challenge would be overcome. The paper does not indicate

specific accuracy but rather these changes were made to reinforce that, over different age group model accuracy is not biased [67].

DenseCaps, is another different capsule architecture which employs dense connections, makes use of cross-capsule feature concatenation, which as a matter of fact, was motivated by the dense net's cross layer feature binding strategy. The model was tested on a variety of well-known canonical datasets with accurate performance of 99.70 % on MNIST, 94.93% on Fashion-MNIST, 89.41% on CIFAR-10 and 95.99% on SVHN. As mentioned, these data sets have a great number of images, therefore data augmentation was not needed. On the other hand, the fact that it is reliant on a large dataset suggests it may be difficult for DenseCaps to attain such a high level of accuracy as these in smaller ones [68].

ResCaps was developed to aid in the problem of papillary thyroid cancer classification by extending the CapsNet approach. In contrast to the ResNet architecture which replaces the standard CNN layer, CapsNet architecture only conducts convolution of its input without introducing additional complexities leading to increased errors. ResCaps was trained using a smaller dataset consisting of 1956 training images and 424 testing images, the model attained an 81.06% classification ratio. In contrast, the same dataset ResCaps using CNNCaps was applied with a final accuracy rate of 79.17, indicating a significant benefit from the residual unit in the processing of advanced medical images [69].

VGG-CapsNet is an extension of CapsNet which focusses integrating its capabilities with the VGG-16 architecture so as to be able to classify lung cancer tissues within CT images with better accuracy and efficiency. The authors of this paper put emphasis on the performance of VGG-CapsNet by benchmarking its performance against other models on the LIDC-IDRI and Kaggle datasets. In the case of the LIDC-IDRI dataset, VGG-CapsNet got AUC score of 0.980 and F1-Score of over 98% meaning it had a precision, recall, specificity and accuracy score of 99.07%, 98.16%, 99.07% and 98.61% respectively. The VGG-CapsNet achieved a precision of 98.14%, a specificity of 99.16%, accuracy of 98.07% an AUC score of 0.98 and F1-Score of 98.14% for the Kaggle dataset which showcases how perfectly the model was developed [70].

The IRCM-Caps model integrates CNN and CapsNet architectures for Covid 19 diagnosis in X-ray images, Improving its performance with the aids of an attention mechanism and a multi-branch lightweight module. To enhance the image contrast, the images are first preprocessed by the Contrast Adaptive Histogram Equalization (CLAHE) algorithm. The model has been validated with a dataset of 1200 standard X-ray images including 400 positives for Covid 19, 400 positive for viral pneumonia and 400 negative cases. Results indicate that the IRCM-Caps model achieved 99% accuracy which was an improvement over the 93% accuracy of the initial CapsNet model [71].

S-VCNet, a hybrid model combining VGG and CapsNet, enhances the accuracy of lumbar spondylolisthesis identification in X-ray images. The study utilized a dataset of 466 private radiographs, with 186 images showing spines affected by spondylolisthesis and 280 images depicting normal spines. The proposed model was compared to other diagnostic approaches and demonstrated superior performance, achieving an accuracy of 98% [72].

Ensemble Model can be classified as a skin cancer detection method that utilizes VGG, CapsNet, and ResNet in order to aid in achieving higher accuracy level. The authors in this research utilized the ISIC (International Skin Imaging Collaboration) dataset, which contains 25000 skin images belonging to several categories with the main concentration on binary classification involving 3000 malignant images and 2800 benign images. Experimental results indicate that the integrated model achieves higher performance than the single models i.e. an accuracy of 93.5% as opposed to 75% for CapsNet, 79% for ResNet and 69% for VGG [73].

The model RNNinCaps consists of modified CapsNet and some RNN module for recognition of 3D vertebral images. The vertebral image dataset used in this study contains a total of 4000 images beside the CIFAR-10 database. The model was evaluated against the CNN, CapsNet, Baseline-1 and Baseline-2 models which were similarly scaled with the same dataset and was found to be more accurate than these comparative models. [74].

The combination of Xception and CapsNet into a model termed XCapsNet is aimed at the accurate and efficient detection of diabetic retinopathy (DR) from fundus images. When this model was used on the APTOS2019 dataset, it obtained a classification accuracy of 83.06% for multiclass DR classification and 98.91% for binary-class DR classification. On Messidor dataset, XCapsNet provides classification accuracy of 98.33% for DR and Normal classes of fundus images which is better than existing models in terms of classification accuracy and computational cost [75].

The FixCaps model is particularly intended for classifying dermoscopic images in the skin cancer context. The model increases the workload easing detection by including a large kernel convolution layer, with a kernel size of 31×31 , increasing the receptive field of the CapsNet, rather than the standard 9×9 kernel, which is the standard norm. The results of the experiments fully support the supposition that FixCaps is superior to all available as well as dominant IRv2-SA models developed for dermoscopic images classification. The Zip files containing the HAM10000 dataset output file observed that 96.49% accuracy using FixCaps which outperformed IRv2-SA results [76].

The BoostCaps model, which was first described by Ben et al. is the first CapsNet which has been able to implement a boosting mechanism. It is specifically created to perform tumor classification tasks. The model accepts images of the human brain as well as rough boundary boxes of the tumor, which enables it to analyze both the target tissues and tissues surrounding it. An internal boosting mechanism is used to iteratively enhance weak learners which makes building large architecture scalable. Evaluated on the Jun Cheng brain dataset, BoostCaps obtained a classification rate of 92.45%, exceeding the accuracy of the original capsule model of 89.83% [77].

Khikani et al., developed a powerful capsule neural network for the classification of breast cancer using histopathological images, by adding a Res2Net block and four other convolutional layers in order to aid in multi-scale feature extraction and minimize. The model was evaluated on BreakHis dataset, and it was able to achieve an overall accuracy of 95.6% and a recall rate of 97.2%, thus outshining other previous models. The study also performed image rotation and flipping to increase the size of

the dataset and improve the performance, and as it has been pointed out, augmentation should be performed whenever small databases are utilized [78].

Kumari et al. synthesized a hybrid model by integrating CapsNet and SVM for classifying five stages of diabetic retinopathy and reached an accuracy rate of 96%. A scan of current work which has employed CapsNet in retinal image classification indicates that while there are works aimed at the detection of common retinal diseases such as diabetic retinopathy and glaucomatous neuroretinal atrophy, no work has specifically sought to classify varied retinopathies themselves [79].

HMedCaps, a proposed hybrid CapsNet architecture, is designed to enhance the analysis of complex medical images by integrating Residual Block and FractalNet modules within the feature extraction layer. This innovative architecture aims to deepen the network while effectively addressing the vanishing gradient problem through the implementation of skip connections. Additionally, a novel Sigmoid-Squash function is employed to improve capsule activation, emphasizing the distinctive features of the capsules. To address dataset imbalances, data augmentation was utilized to equalize the number of images across selected cases. The model's performance was assessed using the CIFAR10 dataset for complex images, the RFMiD dataset for multiclass retinal images, and the Blood Cell Count Dataset for blood cell analysis. The results highlighted that HMedCaps notably surpassed the performance of traditional CapsNet and other contemporary methods, achieving higher accuracy in classifying complex and multiclass medical images, thereby showcasing its potential for more precise diagnoses in medical image analysis [80].

The focus of the work of Kalyani et al. revolves around the usage of a modified architecture of the capsule neural network to achieve the detection and classification of diabetic retinopathy. The aim of the work is to develop a highly accurate and early detection technique by making use of a CapsNet which is free from loss of information which is a characteristic drawback of pooling layers in standard convoluted neural networks. This technique uses the Messidor data set which contains over 1200 RGB fundus picture dispersed into 4 groups where 0 means absent of diabetic retinopathy and 1 to 3 means as the numbers increase so does the severity of the condition. The

model performs enhancement preprocessing of the fundus image to improve the image contrast based on the green channel, press decomposed averaging followed by a set of sequences comprising of convolutional layer, primary capsule, and class capsule and ended by softmax classification layer. The results are quite pleasing with a high accuracy rate of 97.98% for images of healthy retina and there is 97.65%, 97.65% and 98.64% for after the first, second and third stages of diabetic retinopathy. This indicates that the method in discussion has high accuracy in the diagnosis and gradation of diabetic retinopathy [81].

A novel preprocessing pipeline, a fine-tuned Vision transformer (ViT) and CapsNet are integrated together in the algorithm developed by Oulhadj et al. to gauge the progression of diabetic retinopathy. To begin with, power law transformations are applied to change the contrast of the retinal fundus images. The same channels are then processed with contrast-limited adaptive histogram equalization to a certain extent which intensifies the microaneurysms. This results in processed images with one interface providing an input to ViT while the second interface providing an input to the CapsNet. The classification model here takes the output of the two models in determining the extent of development of diabetic retinopathy. This methodology has been implemented and tested on four publicly available datasets like APTOS, Messidor-2, DDR and EyePACS. Results indicated that the test accuracy for the mentioned datasets were 88.18%, 87.78%, 80.36%, and 78.64% respectively. Such results indicate a substantial improvement with respect to the existing best techniques solutions and thus emphasize the importance of the model in enabling a timely diagnosis of diabetic retinopathy [82].

In another study Kumar et al., this study proposes a method that utilizes a modified CapsNet (CapsNet) model which is optimized to detect malaria parasites in blood smear microscopic images. According to the authors, this model consists of a fully functional CNN integrated with CapsNet thereby increasing the accuracy of diagnosis. As a part of improving image quality, the image processing pipeline consists of rotation and normalization techniques before inputting data into hybrid model. The study uses the Malaria Cell Imaging Dataset from the National Institutes of Health which contains more than 27,558 images which consist of infected and non-infected

malaria blood cells. Furthermore, it was found that the hybrid CapsNet model performed satisfactorily well with effective results of detection rate of 99% and accuracy of 99.08% with a false acceptance rate of 0.97%. These results are significant compared with the previously existing models and further prove the effectiveness of the hybrid model for malaria diagnosis [83].

The new capsule model which has been designed in theory includes the foundation of three separate capsule approaches including Tri Texton-Dense Capsule Network (TTDCapsNet), TDCapsNet Dense Stage and TDCapsNet Core Stage which aims at assimilation of three core building pillars towards improved understanding of intricate medical images. The voting scheme provides a simple procedure for feature extraction along with a dedicated eight-layer u shaped dense us encased in a luminance saliency as its last stage. The authors started image classifications in the TDCapsNet's final layer by using the first PC but then may consider altering the second-level TDCapsNet configurations in accordance with proper integration criteria sense-assisted TS quite different from Probability Cubic which influenced TDCapsNet. The author's jury cited over 85% accuracy for facial recognition while noting that this in some study set samples of the model including having achieved over 97% accuracy with specific parameters, accuracy levels that are positioned around 94.90% level for Fashion-MNIST unmatched clothing datasets, CIFAR-10, and Breast cancer. It could therefore be argued that from the arguments made, TTDCapsNet is a fifth-generation CapsNet as it raises the bar for latest CapsNet configurations among others targeting advanced three-dimensional image recognition settings [84].

Veni et al., acute clinical conditions like acne are commonplace in our society today, thus a comprehensive understanding of the condition through its grading and classification is very crucial. In this regard, utilizing a vast amount of data, we need advanced systems and architectures capable of classification and detection. To solve some of the issues within the papers, such as, overly reliant and non-robust models, we propose a hybrid system consisting of a VGG16 which has been reinforced by the inclusion of a CapsNet network which we have called convolutional block attention module (CBAM) to be able to classify acne into the three levels of severity including mild moderate and severe acne. Due to the model being built on some of the recent

datasets, we were able to train and adapt the model successfully. The results achieved by the model were impressive, demonstrating high performance, high confidence clinical tools including a precision of 100%, an F1 score of 99%, recall of 100%, accuracy of 99%, specificity of 100%, and a kappa score of 97.87%. etc. As such, the model shows great promise in the field of medical diagnostics [85].

Saif et al., CapsCovNet et al., gives a second-generation capsule neural network which helps to assist diagnosis of COVID-19 from multiple forms of medical imaging that involve chest X-ray and ultrasound scans using a dataset. CapsCovNet differs CapsNet for diagnosing COVID-19 from more basic ones in that it employs some major features. It uses many multi convolutional blocks whilst varying the kernels so as to extract information at different dimensions, it also deepens the number of the capsule neural network layers to improve the capturing of complex data, finally it employs twisted capsules to mostly improve the learning process. With an application of a refined routing algorithm the information retention is also greatly improved, furthermore, a pre training technique based on a large source of data greatly aids in overcoming secondary data problems. All these changes help improve the models overall predictive ability and efficacy and enable it to perform better than the current practices available in multimodal medical imaging. CapsCovNet, three standard benchmark datasets were used to evaluate: A chest X-ray dataset of COVID and pneumonia patients, a chest X-ray dataset and a point of care ultrasound dataset. The findings were of remarkable success since the model used in the investigation was able to achieve better results in accuracy when put against the current best models, most notably a classification accuracy of 100% of the videos taken of ultrasound scans of COVID-19 cases and great improvement in chest X-ray datasets that improved all other metrics such as precision and recall rates and therefore this was a good pointer in the automatic diagnosis of COVID-19 disease [86].

In this work, an image segmentation method that utilizes capsule neural networks (CapsNet) for segmenting corneal images is proposed. The method consists of several steps; the first one deploys preprocessing in order to remove speckle noise via a hybrid gaussian, anisotropic and bilateral filters. This stage is important in improving the quality of the images before the classification process. There are 579 OCT images in

the dataset which were obtained from the noor hospital, Iran using SD-OCT imaging technology, of which 384 images were used for training while the rest were reserved for testing. The system is specifically trained to classify and segment three key layers of the cornea, outer epithelium, upper stroma and lower endothelium. The modified capsnet's accuracy level was 96.41% as well as high sensitivity (96.83%) and specificity (83.33%). On top of that, the segmentation performance was also analyzed using the Dice Similarity Coefficient and it was revealed that the modified SegCaps outperformed the existing methods and once more emphasized the effectiveness of the suggested development in regard to corneal images segmentation in pre-surgical imaging tasks [87].

Gupta et al. have developed a multimodal neural network called COVID-19 Capsule WideNet which can automatically diagnose COVID-19 from chest X-ray pictures. Their model consists of two convolutional layers and three capsule layers which solve the conundrum of conventional CNNs preserving spatial context while simultaneously solving feature detection. This model has further been tested on the COVIDx dataset that has 13942 images of which there were 507 images of COVID-19 patients and 13435 images of non-covid patients. Such high metric results as 91% accuracy, sensitivity, and specificity, together with AUC equals 0.95. This underlines the ability of the COVID-WideNet framework to address the given task of the timely diagnosis of COVID-19 based on X-ray images [88].

The deep learning framework which utilizes CapsNet based structures of various CNNs such as DenseNet, ResNet, VGGNet, and MobileNet have been proposed for the detection of COVID-19 from lung CT scans. This hybrid architecture, termed as DenseCapsNet, ResCapsNet, VGGCapsNet, and MobileCapsNet, employs the best of both worlds where ConvNets are utilized for feature extraction and CapsNets for higher degree of spatial awareness. The model is trained on SARS-COV-2 CT-Scan dataset containing 1,252 CT images of COVID-19 positive patients and 1,230 non-COVID CT images thereby making a total of 2,482 images. The results show that the three models VGGCapsNet, DenseCapsNet and MobileCapsNet achieved an accuracy of up to 99%, thus proving the effectiveness of the proposed architecture in the diagnosis of COVID-19 from the lung CT scans [89].

Madhu et al., proposed method in the article introduces an Imperative Dynamic Routing mechanism using fully trained CapsNet for malaria classification, addressing the limitations of traditional CNNs. This model is specifically designed to enhance the extraction of features from thin blood smear images, which contain samples of both parasitized and healthy erythrocytes. The dataset utilized consists of 27,558 images sourced from the National Institutes of Health (NIH) repository, ensuring a balanced representation of both classes without class imbalance. The CapsNet architecture includes a series of convolutional blocks followed by dynamic routing between primary and secondary capsules, effectively capturing spatial relationships and reducing information loss by eliminating pooling layers. The results achieved by the model are noteworthy, with an accuracy of 98.82%, an Area Under the Curve (AUC) of 99.03%, and a specificity of 99.30% on the test samples, demonstrating its robustness and effectiveness in accurately classifying malaria-infected cells [90].

Recently, Goceri et al., presented AFConv-CapsNet which is a new deep learning architecture designed for the purpose of diagnosing skin cancers by use of adjustable and fully convolutional capsule layers. This architecture incorporates an adjustable capsule layer that encodes spatial relations between capsule vectors with a learned bias feature, allowing the model to perform better in maintaining orientation and location information. For training and evaluating the proposed model, the authors used the HAM10000 dataset which includes a considerable number of dermoscopic images under seven forms of skin lesion. To achieve a balanced diverse coverage, they used a total of 805 images with the dataset enhanced through the use of image augmentation techniques. Achieved results obtained by the AFConv-CapsNet model are quite good, registering the recognition accuracy at the level of 95.24% with sensitivity diagnosed at 95.37% and specificity at 99.23%, which indicates offline effective classification of multiple skin cancer types [91].

Behara et al. introduced an innovative model for automated skin lesion classification that encompasses a combination of an active contour segmentation, ResNet50 based feature extraction and a lightweight attention driven CapsNet. To begin with, the method in the first step utilizes active contour models to properly segment the skin lesions thus increasing the degree of relevance to the features contained in the images.

After segmentation, important features are extracted from the images through the use of the ResNet50 model, and then the CBAM which reads together each of the channels in the feature map enhances the attention to important parts in the attention spatial feature maps. The obtained features were further concatenated and forwarded to CapsNet that employs dynamic routing for classification. The model has been tested on available databases; the HAM10000 and ISIC 2020 and it has obtained a record of 98% accuracy and 97.3% AUC-ROC. These outcomes demonstrate the considerable prospect of the model to generalize well in the context of automated skin diagnosis systems [92].

The InceptionCapsule approach put forth by Sadeghnezhad et al., combines Inception-ResNet with CapsuleNet within a self-attention framework to address the challenges pertaining to medical image classification. This strategy employs a form of transfer learning, wherein the model is initialized with weights from ImageNet to avoid having to deal with random weight selection problems. The procedure includes a CapsNet augmented with attention that processes feature rich vectors derived from the Inception middle layers and may aid in highlighting useful features. The model was tested against the Kvasir dataset and BUSI with GT dataset. It was able to yield an accuracy of 97.62% and 94.30% for 5 and 8 class respectively in Kvasir dataset. The model achieved an accuracy of 98.88%, 95.34% precision and an F1-score of 93.74 in the BUSI with GT dataset, affirming its superiority over existing techniques as highlighted in the literature [93].

The approach used by Aydin et al. was to address the limitations of conventional CNNs using an enhanced Capsule Neural Network (CapsNet) architecture to classify bone marrow (BM) cells. The specific objectives were achieved through the implementation of three pre-trained models; VGG16-CapsNet, Residual Capsule Network (RES-CapsNet) and Google Network Inception V3 (GN-CapsNet). These models aim at enhancing feature representation by employing dynamic routing and hierarchical feature comprehension. The dataset used in the research consisted of more than 170,000 images showing BM cells with more than 21 cell classes that had been annotated by experts. The application of the Synthetic Minority Over-Sampling Technique (SMOTE) was able to reduce and if not eliminate class imbalance making

classification much better. These results were remarkable, with GN-CapsNet recording a 99.45 % accuracy rates while RES-CapsNet followed with a 99.24 %, VGG16-CapsNet a 98.95 % and the basic CapsNet 96.99 % all achieved impressive performance rates. Such performance results of the models contribute to accurate diagnosis for various hematological diseases [94].

In their study, POKUAA et al. presented their system which is the Denoising Patch-and-Amplify Gabor Capsule Network (DPafy-GCaps), aimed at the diagnosis of gastrointestinal (GI) diseases. The proposed method extends the traditional architecture of the Capsule Networks by using a technique which enhances certain features while suppressing the noise embedded in images. This technique applies to Gabor filters that will be used to selectively obtain the frequency components critical for the diagnosis of the disease. The model was tested on the Kvasir-V2 dataset which consists of 8000 colored images of gastrointestinal diseases. The effectiveness of this model was demonstrated by recognition accuracy that was raised to a remarkable 96.80% on Kvasir-V2 dataset which means its ability to detect diseases such as esophagitis and polyps is greatly enhanced. Moreover, it has also demonstrated comparatively good results with Fashion-MNIST and CIFAR-10 datasets with 95.10 and 85.50 accuracy respectively. All the results obtained show that the model is quite robust and has a broad applicability to medical imaging even in conditions of limited data [95].

A novel method which is named as improved capsule neural network (CapsNet), was introduced by Remya et al in their work, with the objective of being able to correctly identify and quantify the thickness of the brain tumor regions. The method proposed is a three-step process, namely: pre-processing, classification and segmentation. First, the method adopts an enhancement of Discrete Wavelet Transform (DWT) filtering system to reduce Rician noise from movies of the 461 brain tumors of various grades in the BRATS2018 data set. After the noise cancelation step, classification which is also a feature extraction step is performed, only that this time the images are already cleaned. The classification step uses the CapsNet, which can construct useful features even with meager samples. The last step involves segmentation of the images by use of modified Seg Caps in order to distill the abnormal areas of the tumor for accurate

measurements. The accuracy of the tumor segmentation was noted to be 95.36%, an improvement of 1.5% from other algorithms. Moreover, the estimated areas of the tumor thickness were the best, when compared with the other estimates that were provided by the specialists in the field, suggesting that the device can be suitable for use in clinical settings [96].

As per the article by Ye et al., a Genetic Hybrid Rice Optimization (GHRO) algorithm that utilizes a hybrid approach to the search for optimum values in hyper systems for capsule is introduced by the authors. This approach combines genetic search techniques with the hybrid rice optimization algorithm to improve global search capabilities while preventing local optima. GHRO was used to automatically optimize hyper parameters for the CapsNet model which entailed setting up appropriate batch size, the number of kernels as well as suitable activation functions. The methodology was rigorously evaluated with several datasets including MNIST, Chest X-ray (pneumonia) and Chest X-ray (COVID-19 & pneumonia). The experimental outcomes clearly illustrated that the application of GHRO in optimization of parameters of the CapsNet improved its performance in image classification tasks as they achieved more accurate results and a reduction in time taken to train the models relative to the traditional methods for parameter estimation. More specifically, the GHRO-CapsNet model, outperformed the other models in classifying the images in the datasets which attested to the success of the hyperparameter optimization component in the model [97].

The method developed by Chaudhari et al. focuses on the development of an automated bacterial recognition system that incorporates a histogram oriented optimized CapsNet. The approach uses histograms as one of the features which helps in reducing the dimensionality of the input data while still preserving essential color information. The CapsNet on the other hand is supposed to take care of such factors as spatial orientation and pose of a bacterial image. This is further said to enhance the classification ability of an object as compared to conventional CNNs. The model was evaluated using the DiBAS database which consists of a collection of 33 bacterial species. Unfortunately, the paper did not discuss KNN or methods of finding nearest neighbors. The results came out to be quite optimistic in that the accuracy achieved by

the CapsNet was quite impressively high at 95.08, thus beating numerous benchmark ML algorithms such as KNN, Decision Trees, Naïve Bayes and SVM. This implies that the proposed technique is capable and efficient in the accurate classification of the studied genera of bacteria having very few training parameters [98].

The research of Mursalim et al. proposes a model named BiCaps-DBP, a combination of Bi-directional LSTMs and ID-Capsule which aims to create a model that predicts DNA-binding proteins. It contains a one-hot encoding layer to transform protein sequences into binary vectors so that the Bi-LSTM layer can extract contextual features. Later on, the features obtained by relational information from the ID-Capsule would help in classification. The model was tested against three datasets namely, PDB2272, PDB186, and PDB20000. This model achieved the results of 83.11%, 73.49%, and 94.58% respectively, enabling it to be witnessed as a superior model, being 1.05%, 5.79%, and 0.40% more accurate than the other predictors, marking an important step in DBPs [99].

Chen et al. developed a multi-scale convolutional capsule network (MSCCNet). This network has been trained for the classification of cervical cells using a deep learning framework. A cross-layer attention-based feature fusion module has been integrated within the architecture of the model that unites features of different levels (low, middle, and high) to endorse the representation capabilities. Moreover, an attention-assisted spatial relationship modeling module that's been added to the architecture of the model in order to model the differences between the cervical cells of different types that may look similar to each other. The model was tested on two databases the in-house created DSCC data set that has 15,509 cervical images and the publicly available SIPaKMeD containing 4,049 images. Particularly, for the DSCC dataset the accuracy rate was approximately in the range of 87.88% while for the SIPaKMeD data set the scores reached nearly 97.90%. The success rates surpassed many of the existing classification techniques and demonstrated high accuracy performance in cervical classification [100].

To address the limitations of traditional CapsNets used in complex background images, Adu et al. designed the SqueezeCapsNet model by merging SqueezeNet and

CapsNets aims to improve the feature extraction while reducing the number of parameters. The architecture employs sigmoid activation functions for better normalization of coupling coefficients and proposes a new embedding squash function to modify the distribution of the mass of the capsule activations. The model was applied to diverse medical images set such as the brain tumor dataset, lung and colon cancer dataset, alongside with standard datasets, that are MNIST, Fashion-MNIST, and CIFAR-10. Conclusively, the model performed impressively well with 99.87% accuracy on the MNIST dataset, 94.85 on the brain tumor dataset, 99.76% in lung and colon and 82.45 percent on CIFAR-10 which justify that SqueezeCapsNet increases the accuracy of diagnosis in medical images and is robust to many complex background images [101].

According to Pallapu et al., the method describes a new model of detection for breast cancer as well as its spread which includes adversarial capsule network and Adversarial Capsule Network with Graph Convolutional Neural Networks (ACNGCNN). This allows a new technique that utilizes a Generative Adversarial Network (GAN) for data augmentation by constructing artificial images to supplement the training set, further increasing its realism. Such a model applies to a Non-Local Means (NLM) filter that effectively reduces noise from the image without removing any unique features from the image itself. To perform the classification task LSTM with Bidirectional Gated Recurrent Units within a Recurrent ShuffleNet V2 architecture is employed making it easy to tackle long dated data. The method was extensively evaluated on the BreaKHis dataset and there were tangible increments in the detection rates: a 4.9% increase in precision, 3.5% in accuracy, and 3.4% in recall, and a 2.5% boost in the AUC. Moreover, such a model, as a bonus, achieved improvement at specificity of 1.9% as well as reduced the duration required for breast cancer stage detection by approximately 3.4%, confirming the model's potential on early detection and accurate staging of the breast cancer disease [102].

As noted by Dubey et al., the Golden Hawk Optimization based Distributed Capsule Neural Networks is the name of the method that they developed. The aim of this research is to increase the performance of the detection of epidermis lesions specifically focused on melanoma using computer techniques. This technique uses an

optimized clustering-based segmentation approach using Sewer Shad Fly Optimization (SSFO) in order to obtain better boundaries and features of the lesion. The model uses a dataset which comprises 10,000 dermatoscopic images, namely, the HAM10000 dataset. The performance metrics as reported show that the GH0-DCaNN model training had a specificity of 97.53%, a sensitivity of 99.05%, and an overall accuracy of 98.83%. For a k-fold cross-validation with 10 folds, it recorded 97.83% specificity, 99.50% sensitivity, and 99.06% accuracy affirming its effectiveness in the early diagnosis of epidermis lesion [103].

In their proposal, Boruah et al. present a model called CaDenseNet that combines a CapsNet with attention mechanism in order to provide HIV integration site (IS) prediction. The architecture also includes an encoding module, CapsNet module, and other attention layers which help in the processing and learning of genomic sequences. To prevent any bias from antiretroviral therapy due to data stemming from patients on such therapy already, the models work from clean data acquired from the Retrovirus Integration Database (RID), which precisely includes data from HEK293T cells containing HIV. According to the authors, CaDenseNet outperformed all the previously known cutting-edge techniques in the field, demonstrating better improvements on the evaluation metrics of AUC-ROC, AUC-PR, and F-score, thus providing greater insight on the capability of the model to be utilized in HIV research, as well as coming up with intervention strategies [104].

Manikandan et al., effort is identified as CAPSGAN. It is a glaucoma detection method which is a deep learning hybrid model which integrates Generative Adversarial Networks (GAN) and CapsNet. The GAN portion is responsible for creating artificial retinal images to complement the database, which tackles the issue of limited data in medical imaging. A glaucoma dataset was taken from Kaggle with a total of 400 fundus images “normal” classification containing 300 images while “glaucoma” classification consisted of 100 images. Once synthetic images were generated, the CAPSNET was used for classification because it contains various properties pertinent to diagnosis of the disease, namely positional and orientation information. The outcomes showed a very high level of performance with the accuracy of 99.29% registered by the CAPSGAN model, also high precision, recall and F1-score metrics

indicating that the model exceeded several modern ones, such as ResNet-50 and VGG-19 [105].

The research method adds another feature to the existing methods targeting LDA (lncRNA-disease association) using a novel architecture comprised of a stacked autoencoder, attention mechanism and CapsNet, which Zhang et al. termed as CapsNet-LDA. Starting with building similarity matrices for lncRNAs and diseases across N number of views, the method then performs initial feature extraction via a stacked autoencoder in order to obtain lower dimensional representations. Attention mechanism is often utilized as an auxiliary technique to assign the importance variably across these representations before feeding them into a capsule neural network-based manner. Further, studying the work, CapsNet-LDA has been tested with four benchmark datasets which are Dataset1, Dataset2, Dataset3 and Dataset4 with AUC scores of 0.9722, 0.9645, 0.9610 and 0.8770 respectively. These results illustrate that the model is more competent than five other contemporary methods, and also emphasize such robustness and generalization abilities of the model in predicting the lncRNADisease associations [106].

Ravindran et al. proposes an enhanced two-level smart data augmentation in conjunction with Capsule Neural Networks (CapsNet) to forecast the occurrence of cancer diseases. The study employs a cancer gene expression dataset with 14,124 features into six classes such as lung cancer, breast cancer, brain cancer, endometrial cancer as well as prostate cancer. With respect to this issue of lack of sufficient training samples, the researchers resorted to using Uniform Distributive Augmentation (UDA) and created dense synthetic samples using a Wasserstein Generative Adversarial Network (W-GAN). Such an augmentation technique not only achieved statistically balanced class distributions but also augmented the ability of the models to learn. The CapsNet model after being trained on such augmented datasets produced very good results, with measures of performance such as accuracy, precision and recall all above 98% thereby minimizing the error margin typical to such neural network deep learning approaches significantly [107].

According to Zhang et al., the CN2A-CapsNet model integrates attention mechanisms with CapsNet and CNNs to facilitate the automatic diagnosis of COVID-19 from chest X-ray images. As a contribution to the model, CBAM is integrated to give attention to certain areas in the image to improve the extraction of features without spacial dependency that constrains traditional tensile recurrent neural nets. For validation purposes, a dataset of 33,920 chest X-ray images which were subdivided into 3 categories was employed, where the categories included: positive for COVID-19 pneumonia and non-pneumonic patients. The results obtained for the CN2A-CapsNet model were encouraging, where 98.54% accuracy and 99.01% recall were achieved for the binary classification (normal and COVID-19) and 96.71% accuracy and 98.34% recall for the multi-class classification (COVID-19, pneumonia, and normal) illustrating its performance in small size datasets [108].

The work of Boruah et al., MLCapsNet+, consists of human immunodeficiency virus (HIV) integration sites (HIV ISs) identification mechanism that employs multi layered CapsNets integrated with attention mechanism. To allow genomic data to be more comprehensible, the said architecture adds an attention layer to multiple class-capsule layers. The model was trained with data sourced from the RID which consists of positive samples of HIV ISs from hg19 genome build and matched random control sites as negative samples. The dataset was partitioned into training, validation and test sets, and a robust figure of 37,664 positive samples and a corresponding number of negative samples were generated. The results confirmed that MLCapsNet+ surpassed the previously established methods, scoring high metrics on numerous evaluation criteria which included AUC-ROC and F β scores which are indicative of HIV ISs for this method of working [109].

According to the work of Han et al. which is proposed as CardioCaps, CardioCaps is an attention-based dynamic routing capsule network developed with a view to mitigating the problem of class imbalance in echocardiogram images. The structural design includes a new margin-based loss function which encodes a weighting factor towards the positive instances of that class, combined with an additional regression loss that looks at Ejection Fraction (EF) to make the model less sensitive to class imbalance problem. CardioCaps adopts attention mechanism paradigm instead of

dynamic routing which is supposed to enhance training and reduce the burden of satisfying translation-equivariance aspects in echocardiographic images. The model performed consistently well on the EchoNet-LVH dataset that contains around 12,000 videos of echocardiograms yet this dataset is known to be rather skewed with 80% of the content being normal and 20% abnormal videos. Thus, CardioCaps was able to optimally classify 85% of the submitted video samples which surpassed the results achieved by the baseline algorithms i.e. conventional ML and a variety of DL models suggesting that the model is robust to class imbalance in medical images [110].

The method introduced in this study incorporates a Capsule Neural Network (CapsNet) with transfer learning to further improve the classification of thoracic pathology in the chest X-ray images. The researchers created a dataset which consisted of 3043 chest X-ray images from four open access repositories and divided them into four classes being pneumonia, COVID-19, tuberculosis and healthy. The model combines three CNNs including VGG16, VGG19, and Inception V3 as the feature extractors and CapsNet as the classifier with the help of augmentation tactics to solve data imbalance problems. The results were outstanding as the model achieved a training accuracy of 97.29% and the validation accuracy was 96.47%. The sophisticated design of CapsNet model was achieved with F1- score of 97.08%. Which shows that it is capable of diagnosing multiple thoracic diseases [111].

According to Almarshad et al., the SOADL-GCC method proposed by them is applicable in detection and classification of gastrointestinal cancer by way of employing an all-inclusive approach consisting of bilateral filtering for noise reduction, a CapsNet structure for feature extraction, and Deep Belief Network (DBN) for classification. The model also uses a newly developed Snake optimisation algorithm to optimize the hyperparameters of the CapsNet model. The system was assessed using the Kvasir dataset which has 5000 endoscopic images of different gastrointestinal states. Experimental results demonstrated that SOADL-GCC achieved an impressive accuracy rate of 99.72% with sensitivity, specificity and F-score values of 99.29%, 99.82% and 99.29%, respectively. These results prove that the model performs well in detecting and classifying gastrointestinal cancers which ceteris

paribus suggest an improvement of computer-aided diagnostic systems in the health sector [112].

The method introduced by Lei and his colleagues, BP-CapsNet, has also proved powerful and effective for the medical image analysis tasks by integrating the use of a singular value decomposition (SVD) module into Capsules Networks. By performing noise reduction, the SVD function allows the input to make more relevant contributions towards the final decision. It includes more layers including convolutional layers to extract features, a primary capsule layer which transforms those features into vector capsules and A routing module which is designed with a new strategy called Bayesian–Pearson routing. This algorithm responsible for capsule routing manages to alleviate some of causes of deterioration of performance in the case of decaying and noise infestation of multi-dimensional vectors. The performance of BP-CapsNet was comprehensively tested using seven different datasets of medical images drawn from MedMNIST v2, where it achieved state-of –the-art results on two of the datasets and the best results on another two datasets which suggests its generalization and effectiveness in medical image diagnosis [113].

Shaheen et Al.’s coupling mechanism proposes to model the liver cirrhosis disease condition using a hybrid deep learning approach namely the Hybrid Convolutional Neural Network-Capsule Network (HCNN-CN) which was optimized using Adaptive Emperor Penguin Optimization (AEPO) Optimization. The first step is to de-noise the acquired 3D images by applying an Extended Guided Filter followed by binomial thresholding segmentation in order to focus on the ROI. The methods used are the Grey Level Co-occurrence Matrix (GLCM) and gray level Run-length Matrix (GRLM) for feature extraction. After the federated learning of Graph Neural Networks (GNNs) model, HCNN-CN was majorly focused on the classification of features and were enhanced through varying hyperparameters optimally using AEPO. The data set under study contains MRI images of several patients who were diagnosed with liver cirrhosis. The model was able to get good results, with 99.33% accuracy and 98.63% sensitivity which confirms that the model can diagnose the cirrhosis condition appropriately [114].

In their work, Manohar et al., talk about a technique named InceptionCaps which targets the problem of automatic classification of glaucoma through use of CapsNet based deep learning models. In their work, the authors consider an InceptionV3 based architecture as its convolutional subnetwork for better feature representation of retinal fundus images. The model overcomes the problems of limited data by transfer learning, which helps it generalize well when there are not many annotated images. The main benchmark dataset used for evaluation is taken from the RIM-ONE v2 database which consists of a total of 455 fundus images that are sectioned into glaucomatous and normal groups. In this method, astounding performance measures were obtained specifically a: accuracy value of 0.956, specificity of 0.96 and AUC access of 0.9556. Such measures indicate stability and general superiority against other many modern deep learning models for glaucoma classification tasks [115].

Though heel disease diagnosis and treatment has been addressed in multiple medical publications, to the best of our knowledge, no research exists which employs AI to heel disease diagnostics. Furthermore, we did not find any public online database of heel radiographs. This research aims to fill out these voids by adding to the medical field the problems concerning the first contact with patients suffering from foot diseases, namely Heel spur and Sever's disease. The aims of the research include the development of a new set of data to mitigate the data sparsity issue that affects classification efforts and development of more efficient capsule neural network models that provide better performance in detection and classification of congenital disorders of the heel bone.

PART 4

METHODOLOGY

4.1. DATASET CREATION AND ANALYSIS

In this part of the study, we present a previously non-existing dataset that specifically addresses the issue of heel bone disease classification, including Sever's disease and heel spurs. The dataset is supposed to fill in the crucial gap in the publicly available data in this particular medical field so that AI models can be developed to make accurate diagnoses and classifications of diseases. We followed strict collection, annotation, and preprocessing procedures to put together such a dataset that contains multiple images of different conditions and injuries, thus ensuring the availability of a wide range of deep learning models for training and testing purposes. The following section is focused on the methods of dataset creation, annotation, and preprocessing, as well as on the methods of dataset balancing to improve the performance of a trained model.

4.1.1. Collection of Lateral Foot X-ray Images

The X-ray images used in this dataset were taken from the Kirkuk General Hospital, utilizing the hospital's resources in order to gather high-quality medical imaging data. The images were first obtained in a Digital Imaging and Communications in Medicine (DICOM) format, which is the preferred format for medical imaging. This format allows for important metadata and a particular level of imaging quality to be maintained. In order to ease their applicability in machine learning application pipelines, the images were exported into the JPG format of a licensed software known as MicroDicom, which is highly regarded in working with DICOM images.

MicroDicom also allows batch conversion, thus speeding up the rate at which many images would be processed. The images that were collected mostly had a size of $[2021 \times 2021]$ pixels, which was quite useful in providing great detail of the anatomical features of interest. The high resolution, however, posed difficulties regarding computational ability. Thus, to simplify the analysis and concentrate on the region of interest (ROI), the images were resized during the preprocessing stage.

4.1.2. Dataset Characteristics

The dataset is composed of 3,956 images categorized into three classes:

- Normal Heel: 1,842 images
- Heel Spur: 1,316 images
- Sever's Disease: 798 images

Examples of X-ray images are shown in Figure 4.1, Normal Heel, Figure 4.2, Heel Spur, and Figure 4.3, Heel with Sever disease.



Figure 4.1. Normal heel.



Figure 4.2. Heel spur.



Figure 4.3. Heel with sever disease.

While the dataset provides a comprehensive representation of the conditions studied, it exhibits an imbalance across the three classes. The "normal" class contains the most images, followed by the "heel spur" class, with the "Sever" class having the least. This imbalance reflects the natural prevalence of these conditions in the population but poses challenges for training AI models, as imbalanced datasets often lead to biased predictions favoring the majority class.

To correct this imbalance, some oversampling of the minority classes was also conducted during model training to guarantee even knowledge and accuracy across all the classes.

4.1.3. Annotation Process

Annotation of the dataset included locating the area of interest in each X-ray image. The ROI included the heel bone, which is required to diagnose ailments like heel spurs and severe disease. This was done using specialized software, for example, Votte v.2.2.0 and labelImg-1.8.6, which are both known to be precise as well as user-friendly when it comes to creating bounding boxes around pertinent regions of the medical images.

The region pertaining to the heel bone was marked using a bounding box, making the rest of the image unnecessary and ensuring the model used the parts of the image that had useful diagnostic information. This approach has immensely minimized the processing load and also enhanced the performance of the machine learning pipeline. The labeling process was done by a qualified orthopedic and traumatology surgeon with a radiology background to counter-check the validity of the labeling. Illustrated examples of X-ray photographs demonstrate specific classes which include (A) none affected heel (B) heel spur, (C) Sever's disease demonstrated in Figure 4.4. These annotations made it possible to form a dataset with features belonging to well defined and labeled classes which are key in the development of an effective AI model.



Figure 4.4. (a) Normal heel, (b) Heel spur, (c) Sever disease.

4.1.4. Data Preprocessing Techniques

Data preprocessing is an important phase in making the dataset ready for machine learning applications. This research involved implementing various native techniques to facilitate the optimization of the dataset:

- **Resizing:** Resizing all X-ray images to pixel images of [64 x 64] enabled uniform standardization of the input dimensions into the AI model. This is resizing lowered computational needs without affecting the fundamental features that are necessary for diagnosis.
- **Cropping the ROI:** A rectangular box was used to crop the area of interest in the image, thus removing uninteresting parts of the picture and concentrating the model on the heel bone. Figure 4.5. Show the samples of cropped ROI images.
- **Normalization:** The pixel values were rescaled to fall between the range of 0 and 1, so there is image intensity homogeneity across the dataset. This step enhanced the capacity of the model to learn features.

These series of preprocessing steps made it possible for the dataset to adapt to a neat and suitable form, which was ready to go for training and testing of deep learning models, thus producing ideal and precise outcomes during the classification of the diseases of the heel bone.

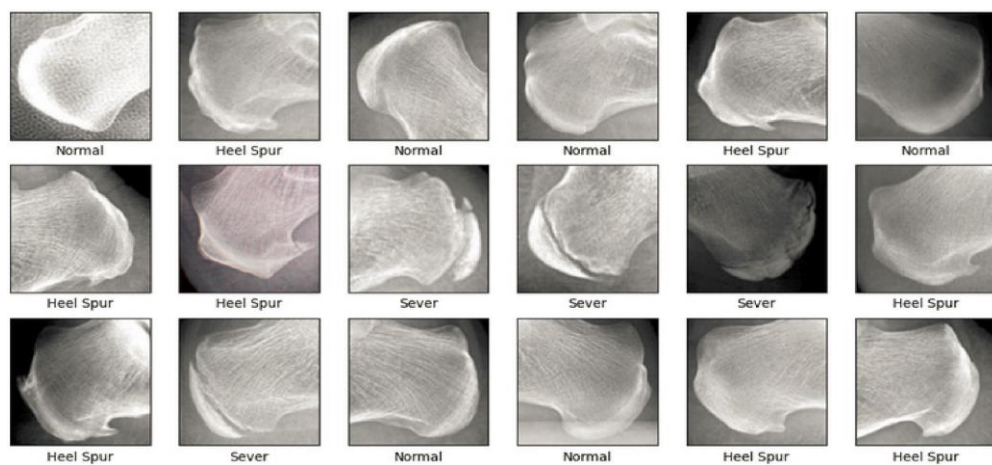


Figure 4.5. Samples of cropped ROI images.

4.2. BASELINE MODEL - ORIGINAL CAPSNET

As depicted in part 3, the 2017 invention of Sabour et al., referred to as CapsNet, offers a solution to scaling problems faced in standard convolution neural networks [12]. Explaining this feature in further detail, we see that these networks assist in patterning data interactions as they take a vector form instead of a scalar, which makes them effective in hierarchal spatial applications. Targeting our heel disease dataset with the help of CapsNet presented us with some specific issues with regard to suitability based on the design aspects that follow suit in this section.

4.2.1. Application of CapsNet to Heel Disease Dataset

The initial version of any CapsNet model was mostly developed and tested on the MNIST dataset, which is a standard benchmark and a set of binary images of handwritten digits with a size of $[28 \times 28]$ pixels. Although the MNIST dataset is simple and has a consistent data structure, the heel disease dataset is not. The X-ray images in our dataset show fine bone structures, different pixel levels, and images with different resolutions, which are extremely different from the binary images of the MNIST dataset.

In order to transfer the original CapsNet model to the heel disease data set, the images were first resized to $[64 \times 64]$ pixels so that they were compatible with the model's input layer; Figure 4.6 shows the applying the original capsule architecture on the heel dataset. Nevertheless, the results were still disappointing as the proposed model did not perform well on the heel disease dataset. This could be due to two or more of the data having different characteristics, and thus:

- **Complex Pixel Information:** Unlike MNIST's binary representation, the heel disease dataset contains grayscale X-ray images with high pixel density, which requires a more robust feature extraction mechanism.
- **Dataset Size:** The relatively small size of the heel disease dataset posed challenges for the original CapsNet, which was designed for datasets with a larger volume of samples.

- **Feature Variability:** The dataset comprises three distinct classes—normal, heel spur, and Sever’s disease—each exhibiting unique patterns that the original CapsNet struggled to distinguish effectively.

Although CapsNet inherently provides better performance in capturing spatial hierarchies and part-to-whole relationships compared to CNNs, the limitations of its original design became evident when applied to medical imaging tasks.

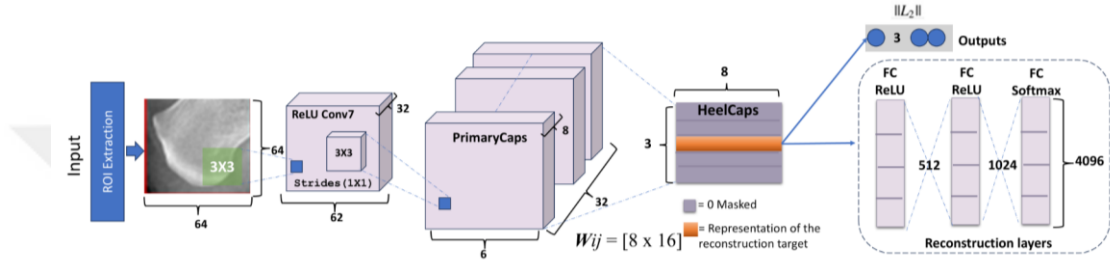


Figure 4.6. Original capsule architecture with heel dataset.

4.2.2. Performance Analysis (Accuracy, AUC, Cross-Validation)

The performance of the original CapsNet model on the heel disease dataset was assessed using key metrics, including accuracy, AUC, and a confusion matrix analysis. These metrics provide insights into the model's effectiveness in detecting and classifying heel disease.

4.2.2.1. Accuracy

Accuracy evaluates the proportion of correctly classified samples out of the total dataset. The CapsNet model achieved an accuracy of 73.99%, highlighting its general performance in classifying heel disease conditions. Figure 4.7 shows the curve of training accuracy. Figure 4.8 shows the curve for the training loss of the model. Additionally, the class-wise performance metrics, including precision, recall, and F1-score for each class, are detailed in Table 4.1, highlighting the performance of CapsNet in distinguishing between the categories of the heel disease dataset.

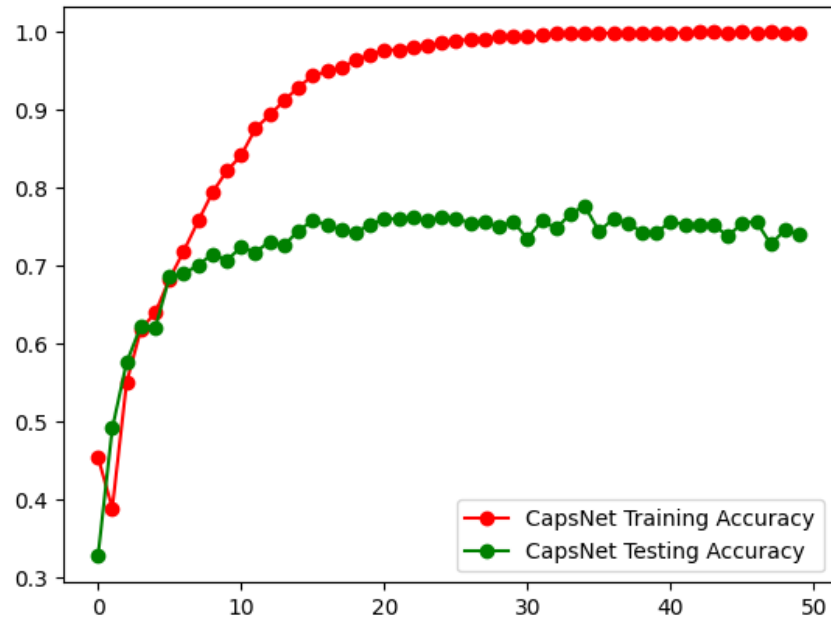


Figure 4.7. Training and testing accuracy of CapsNet applied on heel dataset.

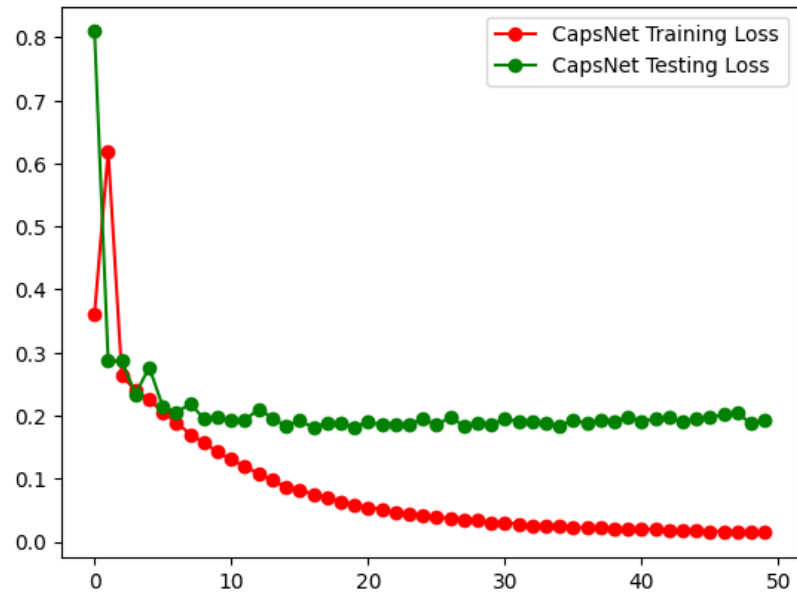


Figure 4.8. Training and testing loss of CapsNet applied on heel dataset.

4.2.2.2. Area Under the Curve

The AUC measures the model's ability to differentiate between classes, with a value closer to 1.0 indicating better performance. For the CapsNet model, the AUC was recorded as 66.95%, demonstrating its potential in binary and multi-class classification tasks. Figure 4.9 shows the curve of AUC for original CapsNet applied on Heel dataset.

4.2.2.3. Confusion Matrix Analysis

A confusion matrix was generated to further understand the model's classification performance, as shown in Figure 4.10. The confusion matrix provides a detailed breakdown of the true positive (TP), false positive (FP), true negative (TN), and false negative (FN) values for each class in Table 4.1. The confusion matrix shows the capabilities and limitations of the model against each class, seeking the class where some error or misclassification occurred. Although these metrics are useful to some extent, the performance indicates that a better model is required that would be more capable of dealing with the problems of perplexing X-ray images along with an enhanced level of classification performance.

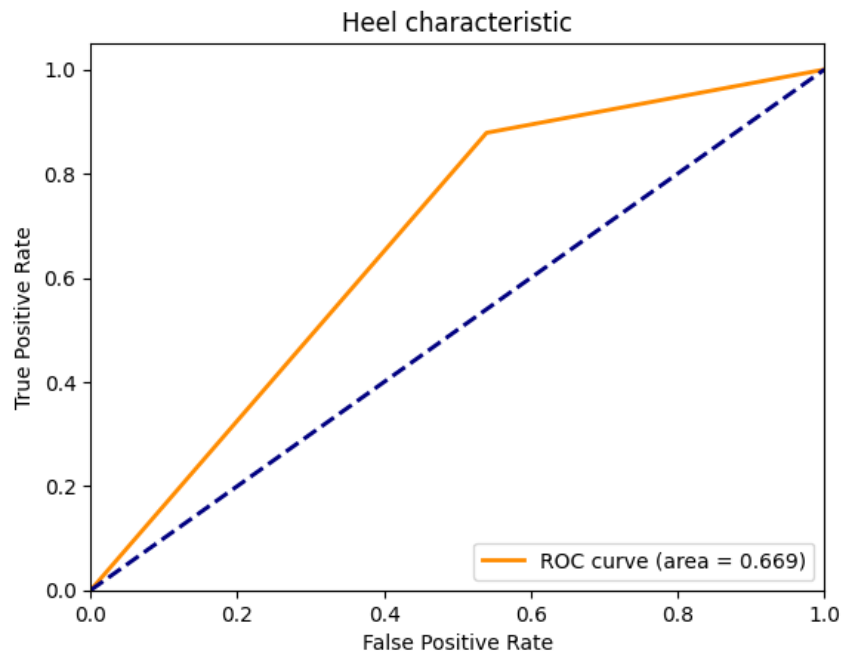


Figure 4.9. Curve of AUC for original CapsNet applied on heel dataset.

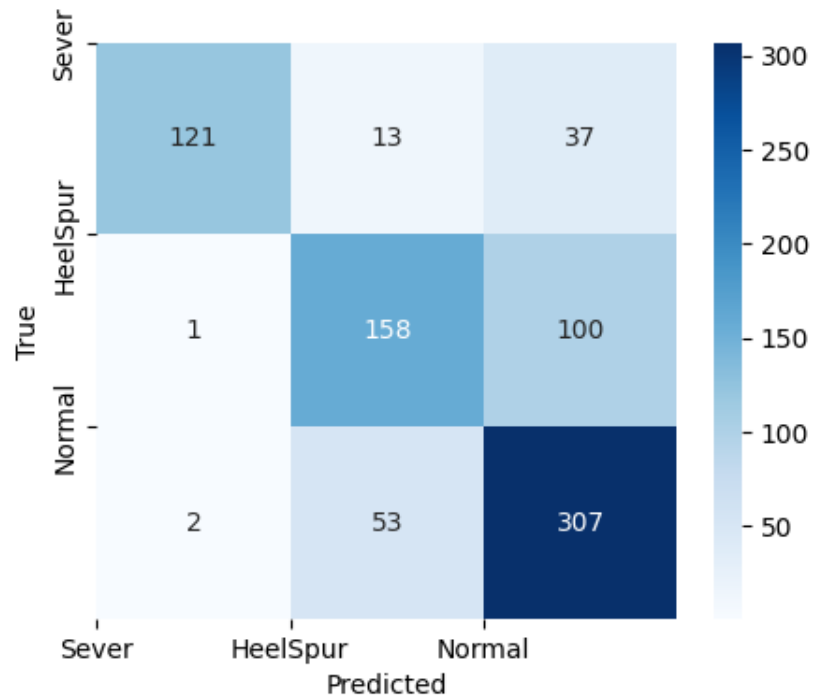


Figure 4.10. Confusion matrix of original CapsNet applied on heel dataset.

Table 4.1. Class-Wise Performance Metrics of CapsNet on heel disease dataset.

Classes	Precision	Recall	F1-score
Sever	0.98	0.71	0.82
Heel Spur	0.71	0.61	0.65
Normal	0.69	0.85	0.76

4.2.3. Limitations of CapsNet for Heel Disease Detection

In spite of the groundbreaking architecture of CapsNet, the application in heel disease detection brought out certain shortcomings:

- **Incompatibility with Medical Data:** The original CapsNet was tailored for data sets like MNIST that are rather straightforward in terms of pixel representation and the number of features. Medical X-ray images, for example those in our dataset, are richer in features with greater pixel variability. This variation means that considerable changes must be made to the CapsNet model architecture to be able to handle such data effectively.

- **Dependence on Dataset Size:** Optimizing the parameters and vice versa of CapsNet requires a lot of data. In this regard, the moderate amount of heel disease dataset that included three classes and 3,956 images was not enough to explore the entire potential of the original model, and it only caused overfitting and poor generalization issues.
- **Suboptimal Hyperparameter Settings:** First, the intrinsic parameters or default settings, such as the kernel size, stride, and the number of capsules of CapsNet, did not meet the requirements of our dataset. Because of the limited size and complexity of the images, a configuration where some particular hyperparameters were altered was appropriate, which was not the case for the basic configuration of the CapsNet.
- **Lack of Multi-scale Feature Representation:** The original formulation of CapsNet evidently has difficulty scaling its feature extraction capabilities, which is crucial for identifying fine differences between X-ray images. Its inability to adjust for variable feature size and shape considerably reduced its usefulness in many tasks, especially in the correct diagnosis of heel diseases.
- **Computational Complexity:** Though the architecture of CapsNet is unique, it is quite demanding in terms of computation requirements, which makes it less appealing for specific tasks in the real world, such as medical applications, which emphasize efficiency. The high computational cost also added to the complications of training the model on a relatively small dataset.

In summary, to conclude, the primary research gaps that need to be filled by future research are also addressed. And so, as these seeming problems began to emerge when the first CapsNet was employed for heel disease detection purposes, it became clear that the remaining tasks were simpler than they appeared. Such limitations prompted the search for a better-suited, selectively focused deep learning model capable of coping with medical images.

4.3. AN ENHANCED CAPSULE NETWORK (HECAPSNET)

4.3.1. Introduction to HeCapsNet

The HeCapsNet is a model produced by applying a set of improvements to the original CapsNet architecture. One of the major improvements that we have introduced is the use of the “he_normal” kernel initializer, as proposed by Kaiming He in 2015 [116]. This kernel enables the model to work effectively, which is the reason why we opted to call this “HeCapsNet” because of its applicability. The enhanced model is designed for automatic detection and classification of the Heel spurs and Sever disease using lateral foot X-ray images. These heel diseases are common ailments that cause pain and discomfort and are usually diagnosed through X-ray images viewed by orthopedic doctors. However, diagnostics and therapies, in a way, depend on manual diagnosis, which is a problem in areas with limited access to healthcare workers. This is an example of how AI-based systems are useful to remedy the wide diagnostic gap.

Even though CapsNet, such as the original CapsNet, has shown to be quite resourceful, they have only been used on standard datasets, such as the MNIST dataset, which only consists of binary pixels. This dataset, however, being relatively simple, is in stark contrast to the complexity of medical imaging, which does tend to have high resolution features along with well-defined anatomical structures. We trained CapsNet on the heel disease dataset and found it had a rather large performance gap, and the accuracy was unsatisfactory. We were able to identify the issue as being a performance gap, and we may need to permanently restructure the architecture of the system to apply it more effectively for medical imaging applications.

HeCapsNet implements the above discussed adjustments so as to enhance the feature extraction and representation. Such adjustments consist of extra convolutional layers, optimized kernel initializers and dropout techniques for fighting overfitting.

Consequently, all these improvements allow HeCapsNet to capture the intricate pixel arrangements of medical images which result in such performance evaluation metrics as accuracy, AUC and cross validation results to increase significantly. This model not

only improves diagnostic capabilities in low-resource settings but also prepares the stage for AI-driven developments in the medical imaging domain. The novel dataset and better architecture are the two features that HeCapsNet embodies, which is an interjection in the AI revolution in the scope of medicine and day scale and accurate solution to bone disorders.

4.3.2. Architectural Modifications

HeCapsNet's architecture includes several architectural modifications and improvements that were devised as responses to issues encountered with the original model CapsNet. These changes were made in order to enhance the model's applicability to intricate medical images and performance on the heel disease dataset. These architectural modifications are as follows:

- **Additional Convolutional Layers:** For the sake of improving the extraction of features, HeCapsNet inserts additional convolutional layers prior to the primary capsule layer. These layers are designed to extract the high-dimensional aspects of the X-ray images of heel bone, thus making it much easier for the network to accurately code the features associated with heel disease indicators. With these layers stacked, the model is capable of extracting both lower and higher-level features that aid in proper classification.
- **Kernel Initializers:** Achieving an accuracy of 91.68% on the modified CapsNet, "*he_normal*" kernel initializer was employed to better the performance. The "*he_normal*" initializer was invented in 2015 by Kaiming He, and it is mainly used to increase the gradient flow in a deeper network that uses ReLU activation [116]. This strategy helps prevent problems such as very small or large outputs during initialization and thus allows faster and more reliable training. The application of this kernel initialization technique enables the model to achieve an appropriate weight distribution, which is advantageous in datasets with high variability in pixel intensities.
- **Dropout Layers:** HeCapsNet employs dropout layers to solve overfitting, which gains significance when the number of samples in a dataset is low. Regularization is implemented within the network through dropout, ensuring

the model is able to make accurate predictions on new data. **Modified Capsule Layers:** The capsule layers have been improved to incorporate spatial relations and hierarchies between features. The primary capsule layer has been modified to use a higher-dimensional capsule to encode more intricate feature representations, while the classification capsule layer's purpose is to develop accurate predictions by employing the input obtained from the other layers.

The structure of the proposed HeCapsNet architecture model can be explained as follows: The convolutional layers are described in blocks. Block one has an inception convolutional layer made up of 64 filters of size $[3 \times 3]$ and a stride of one. The mentioned layer employs ReLU as its activation function, is initialized with "he_normal" and maintains the same padding scheme. After this, a dropout layer is employed with a ratio of 0.2 in order to decrease overfitting. With 64 filters and the previous kernel again, another Conv2D layer was introduced. This layer frames a $[2 \times 2]$ maximum pooling layer, which is aimed at reducing irrelevant spatial features. The second block consists of the Conv2D layer with parameters of 128 filters, respective kernel size of $[3 \times 3]$, and activation of ReLU with the same padding. With a three times rate, the dropout was used again. Another Conv2D layer with the same parameters as above, for example, 128 filters and $[3 \times 3]$ kernel, is added, which is then followed by another $[2 \times 2]$ max pooling layer. The third block of feature extraction augmented to enhance efficiency is comprised of 256 filters Conv2D's, two in total, with a $[3 \times 3]$ kernel, and a dropout section is placed between them with a 0.5 rate. Later, a max-pooling layer with $[2 \times 2]$ size is applied after convolutional layers. The data is then forwarded to the primary capsule layer, which can be seen as a convolutional layer with $8 \times 16 = 128$ capsules, where each capsule has eight parameters implemented in a $[3 \times 3]$ kernel with a stride of one. To represent the output of the capsules, the output is reshaped to yield eight-dimensional vectors, and then the squash function is applied to scale the output in the range of $[0, 1]$. All these architectural modifications together improve the overall processing capabilities of HeCapsNet in terms of dealing with intricate medical images, feature extraction, and prediction. The new set of convolutional and capsule layers along with dropout and advanced initializers, augment the robustness of the network's overbalanced and

imbalanced datasets. The architecture of enhanced HeCapsNet is illustrated in Figure 4.11, which mentions these modifications that were integrated into the model.

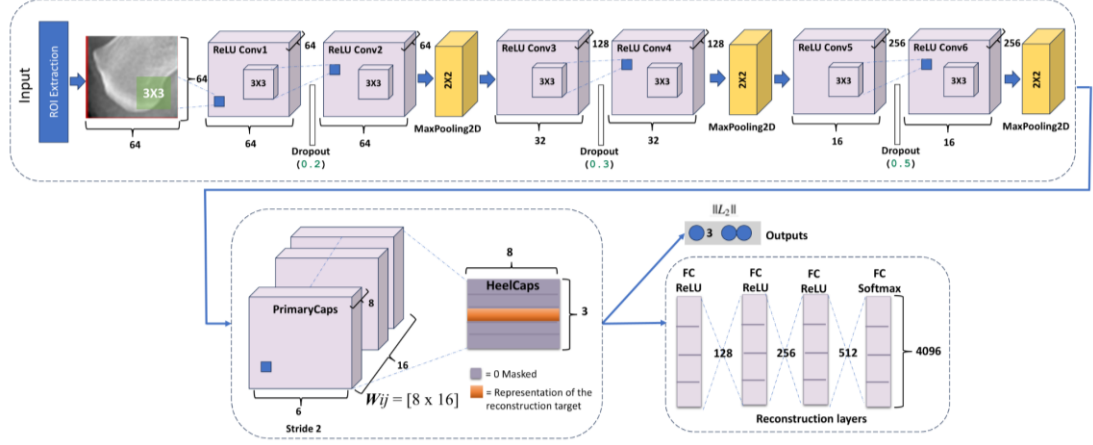


Figure 4.11. Proposed enhanced CapsNet (HeCapsNet).

4.3.3. Incorporating Yolo With HeCapsNet

To improve the preprocessing pipeline of HeCapsNet, the YOLO model was implemented to automatically recognize and extract the region of interest (ROI) in heel X-ray images. YOLO is one of the best algorithms for object detection owing to its speed and accuracy, which makes it suitable for use in medicine where the accurate position of features is necessary [117].

When the whole foot X-ray is presented to the model, it assists the user in automatically detecting the heel area without requiring the user to select the region of interest, which enhances the diagnosis of the given condition. YOLO solves this challenge by fully automating the detection process, thus guaranteeing the reliable and accurate classification of the heel region. This step considerably minimizes the preprocessing workload and enables the HeCapsNet model to emphasize the key features in the images while ignoring regions that are not relevant.

4.3.3.1. Implementation of YOLO Steps

- **YOLO Model Architecture:** In this study, we made use of YOLOv8 for object detection concerning the ROI. YOLO splits the input image into grid cells and predicts each grid cell's bounding boxes and class probabilities. For the heel disease dataset, the model was trained using a part of x-ray images in order to detect and locate the heel area. As shown in Figure 4.12, the YOLO was applied to the heel input image to produce the cropped ROI image.
- **Training Process:** The model of YOLO underwent fine-tuning using the part of the heel disease dataset, which has been annotated with bounding boxes on the heel parts.
- **Output Integration:** The bounding boxes found by YOLO allowed for automatic cropping of the heel areas. These followed cropped ROIs, which were resized to $[64 \times 64]$ pixels and sent to HeCapsNet for classification tasks.

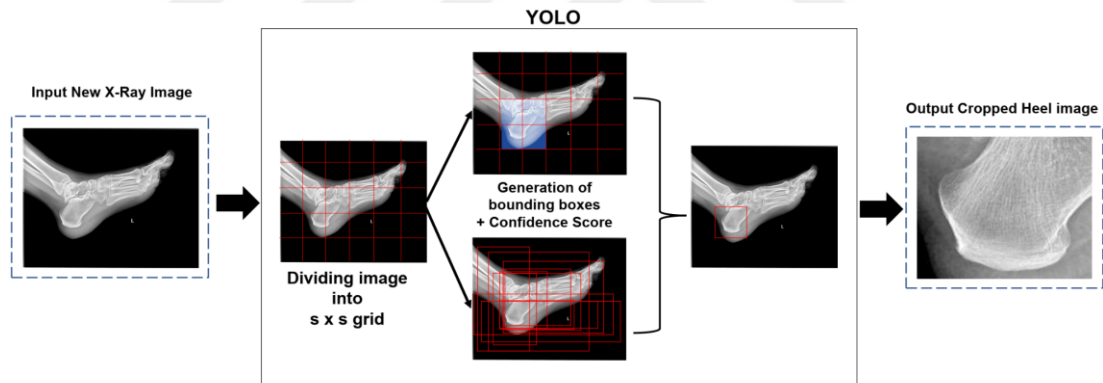


Figure 4.12. YOLO is applied to a heel X-ray image to automatically crop the ROI.

4.4. HYBRID CAPSULE NETWORK (MEDCAPSNET)

4.4.1. Introduction to MedCapsNet

The field of automated medical diagnosis and analysis, which uses medical image processing, has always relied on complex neural network structures to accomplish accurate and dependable diagnoses. CapsNet has been revolutionary in covering the drawbacks of convolution neural networks, like the failure to preserve spatial

hierarchies. However, the original model of CapsNet, which was developed for simpler data sets like MNIST, has not been successful in meeting the requirements for more complex medical imaging data sets. Such datasets generally include high-resolution images with complex structural anatomy and extreme differences in pixel structure and intensity. To cover this shortcoming, we designed a MedCapsNet, which represents a combination of CapsNet architecture and a modified DenseNet201 model. This approach combines DenseNet's hierarchical feature extraction capability with CapsNet's spatial extent, thus allowing flexibility in its application to medical image classification tasks. Therefore, we decided to name the model "MedCapsNet" since it is an improved version of the Capsule network, which has been optimized for medical images.

CapsNet's first implementation tests were performed on the MNIST dataset, which consists of black-and-white images with pixel values of either 0 or 1. Although MNIST serves as a standard and simple platform for neural network evaluation, it is completely different from medical imaging datasets. Medical images have more intricate characteristics, higher resolutions, and wider distributions of pixel intensities. In addition, medical datasets are usually smaller and more skewed than datasets like MNIST. Such discrepancies made the original model of CapsNet unsuitable for medical datasets, and the architecture needed considerable changes to handle such intricate tasks.

One of the issues associated with medical image datasets, e.g., X-ray and MRI images, is their high intra-class variability, smaller quantities, and intricate details within pixel level. These factors make it difficult for one to apply the traditional neural networks without heavily augmenting the training data provided. To alleviate these problems, MedCapsNet uses a modified DenseNet201 structure, which is better suited to the requirements of medical imaging. The structure of DenseNet201 allows for the improvement of the flow of information across its layers by forming dense inter-layer connections that allow every layer to receive input from every preceding layer. It achieves this by promoting better propagation of the gradients, and very deep networks can be built to learn and model complex features essential for the interpretation of medical images.

4.4.2. Introduction to DenseNet201 Model

The DenseNet201 model, being an enhanced variation of the Dense Convolutional Networks, is perhaps one of the most distant developments in the sequence of deep learning architectures. As noted by Huang et al. in 2017, DenseNets were formulated with the expectation of enhancing the interlayer connections, improving gradient flow, and optimizing computing costs without compromising parameter efficiency. In particular, DenseNet201 is more than an implementation of such logic: it is a deep neural network with 201 layers designed for complex image recognition and classification tasks [118].

DenseNet201 is based on the idea of dense connectivity, and it completely contradicts the structural paradigm where information is passed from one layer to another in a one-way sequential manner. In contrast, DenseNet proposes to interconnect directly every layer with all the layers that come before it in order to allow for a complete and unidirectional supply of information to the layers. This integration is accomplished by feature concatenation such that each layer not only takes inputs from the closest layer but also collects feature maps from all the layers before it.

Thus, this connectivity pattern enables DenseNet201 to possess $L(L+1)/2$ connections in a network with L layers. For example, in five layers, the second layer **H2** obtains input from **x0** and **x1** within the dense block connectivity pattern which ensures that the first layer or **H1** is an input while the other layers **H3**, **H4** obtain input from all the lower layers as shown in Figure 6.1.

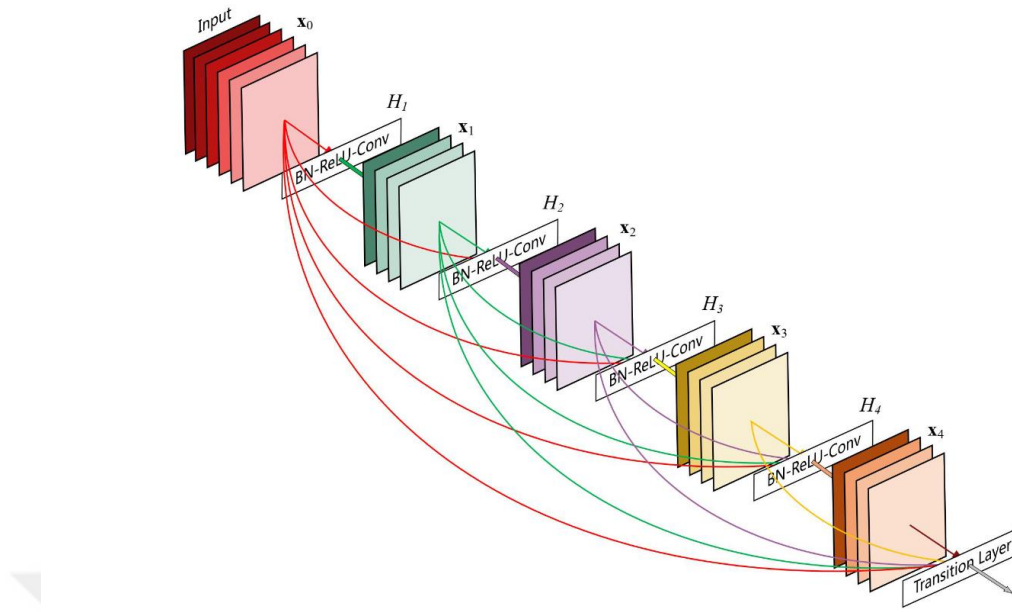


Figure 4.13. Dense connectivity pattern adapted from [118].

DenseNet201 has several distinctive properties that make it an exceptional architecture for deep learning tasks wherein the most distinguishing characteristics are [118]:

- **Enhanced Gradient Flow:** A dense connectivity pattern helps in getting around the vanishing gradient problem through better gradient flow within the network. All layers of a deep network architecture have direct access to the output gradient and the original input signals, which are helpful in learning deep networks.
- **Parameter Efficiency:** Using concatenation instead of summation for feature reuse, DenseNet201 greatly decreases the number of parameters in comparison with classic networks. This saving of resources not only lessens memory consumption but also enhances the generalization power of the network.
- **Feature Reuse:** Each layer in DenseNet201 adds its feature maps to all the subsequent layers so that features can be reused. This increases the number of available features for each layer, which assists each layer in representation learning, further improving performance in more complex tasks.
- **Reduced Computational Burden:** Despite its numerous layers, DenseNet201 utilizes a reduced number of parameters, thereby attaining enhanced computational efficiency. This enhancement in efficiency can be attributed to

the incorporation of dense blocks, a structural element that serves to reduce redundant computations, thereby focusing on the iterative refinement of feature maps.

- **Growth Rate and Dense Blocks:** The concept of a growth rate (k), which defines the number of feature maps to be added by each layer, is introduced in DenseNet201. The setting of a larger growth rate aids in having more dense connections since each layer is able to utilize more features. For instance, in a dense block with $k = 4$, the fourth layer and all subsequent layers use denser feature maps than the third layer, among others.

DenseNet201 consists of several dense blocks separated by transition layers; these combine to allow the network to learn hierarchical features and become adaptable to several tasks:

4.4.2.1. Dense Blocks

The majority of the structure is constituted of dense blocks, where each layer is connected to all layers of the preceding blocks. Such an arrangement enhances the efficiency of feature reuse alongside making representation learning more robust.

4.4.2.2. Transition Layers

The dense blocks have transition layers placed in between them to down-sample the spatial resolution of the feature maps and control the number of feature maps being sent to the next blocks. These layers make use of convolution, batch normalization, and pooling, among others, to reduce the complexity of the network.

4.4.2.3. Growth Rate Optimization

The growth rate in denseNet201 has been designed to enable a reasonable tradeoff between efficiency in computation and richness of representation. Each layer is influenced by this parameter since it specifies the count of new feature maps that are produced in each layer, making the network suitable for both small and large datasets.

4.4.2.4. Output Layer

Densenet201 ends with a global average pooling layer and FC layer, which merges all the features learned in the previous layers and makes the last predictions for classification tasks.

DenseNet201 is currently one of the most widely used deep learning models across different fields. This is largely because it is able to handle big datasets and multiple complex features. For instance, in regard to medical imaging, DenseNet201 impressively searches for very complex parts of the human body and any slight changes in scans, thus making diagnosis better. DenseNet201 is among the most advanced architectures for deep learning, and it is useful in complex image recognition problems. It uses many dense connections, is parameter efficient, and extracts features efficiently, making it ideal for cases that need high precision and less computational consumption. Finally, DenseNet201, in essence, tackles the loss of information and redundancy in computations; therefore, it is convincingly a building block to deep neural networks and all the potential inventions that will emerge in the future of deep learning.

4.4.3. Integration Of DenseNet201 with Capsule Network

Before entering the integration of DenseNet201 with CapsNet, it's worth noting that we first trained the DenseNet201 model with our heel bone dataset. The training achieved an accuracy of 92.17% and an AUC of 92.46%; this gave an indication of powerful model characteristics to learn deep features such as complex medical images. Thus, we decided to use it as a model for integration with CapsNet to improve performance.

MedCapsNet makes use of the underlying framework of DenseNet201 but incorporates specific changes in order to increase its usability for the processing of medical images. These modifications improve the model performance when confronted with medical data complexities such as high dimensional X-ray images by

integrating both the dense connectivity patterns of DenseNet201 and the hierarchical and rich feature representations of CapsNet. The MedCapsNet architecture, with the help of its key features, enhances the original idea of DenseNet201 and modifies it for better performance on complex image classification, as shown in Figure 4.14. Below are discussed in detail the differences in the original structures and the modified version of the structure of the DenseNet201, which were incorporated in MedCapsNet:

- Dense Block
- Original DenseNet201: Employs densely connected layers in a block wherein each layer receives input from all previous layers within the block. This connectivity guarantees strong gradient flow, solves the problem of vanishing gradients, and features reusing.
- Modified DenseNet201: The fundamental idea concerning the issue of dense connectivity remains, but hyperparameters have been adapted to better manage the complexity of the problem associated with medical image datasets. Additionally, an increase in the severity of overfitting can be prevented, and computational requirements can be reduced by limiting growth rates and the number of filters.

Dense Layer

- Original DenseNet201: Applies a basic convolutional layer with batch normalization and the ReLU function. This configuration works well with standard datasets but is somehow limited in terms of grasping the fine details in complex medical images.
- Modified DenseNet201: Integrates two convolutional layers of different filter sizes in order to extract more complex features. The dropout layer is included in the model to enhance generalization capability and avoid the issue of overfitting. In addition, the “*he_normal*” initializer is used to secure better weight initialization for complex pixel distributions, which in turn assists in better convergence and improved gradient for the model’s multiple layers. The “*he_normal*” initializer takes samples from a truncated normal distribution, which has a mean of 0 and a variance as described in equation (6.1), with n being the number of input units in the weight tensor [119].

$$\sigma = \sqrt{\frac{2}{n}} \quad (6.1)$$

Transition Layer

- Original DenseNet201: In order to enhance performance while minimizing the number of feature maps, batch normalization, ReLU activation, and global average pooling are used.
- Modified DenseNet201: To avoid overfitting, dropout is employed instead of batch normalization, and global average pooling is replaced by MaxPooling2D. The latter better retains details of the local features, which is important in medical imaging since little details could be important in diagnosis.

Convolutional Layer

- Original DenseNet201: Does not entail the application of extra convolutional networks prior to the entrance into DenseNet structure. It is obvious that the model performs directly on the provided input data, which may constrain its use in more complex datasets.
- Modified DenseNet201: Presents two convolutional blocks prior to DenseNet. Each block is composed of two convolution layers: Conv1 and Conv2. The first convolutional layer, Conv1, includes a kernel size of $[3 \times 3]$, the same padding technique, and a normal kernel initializer, generating 64 channels of output for the first block and 128 for the second one. After conv1, a dropout layer is used with a value of 0.2 and 0.3 for the first and second blocks, respectively. Subsequently, Conv2 is performed with the same settings as Conv1, with the addition of a 128-channel output layer. Finally, a Kernel MaxPooling2D layer with a Kernel Size of $[2 \times 2]$ is used to form the completed blocks. These blocks act on the raw input by normalizing and enhancing feature maps, which, in turn, facilitates the other processes within DenseNet.

Number of Blocks and Growth Rate

- Original DenseNet201: Consists of 4 dense blocks with a growth rate of 32 and 6, 12, 32, and 48 filters, respectively. This architecture works well for a high number of data sets but is costly and tends to overfit for low datasets.
- Modified DenseNet201: Decreases the architecture to one dense layer with a growth rate of 4 instead of 32 and 2 filters only instead of a group of

(6,12,32,48) filters, respectively. Such specialization minimizes the expenses for modeling, allowing the core function of the model to be relevant without overtraining with minor datasets.

Table 4.2 highlights the differences between the original DenseNet201 architecture and the modified DenseNet201 used in MedCapsNet. Such a combination of the modified Dense Net 201 and CapsNet makes it possible for MedCapsNet to transcend the challenges that the original architecture had. These problems included problems in interpreting datasets containing complex spatial patterns and increased chances of overfitting in smaller datasets. Addressing the said problems, MedCapsNet shows notable advancement in terms of diagnostic accuracy, robustness, and computational efficiency. These transformed MedCapsNet will allow it to capture a rich hierarchy amongst the input data while being able to process medical data sets efficiently, making it more appropriate for heel disease detection in the larger context of complicated imaging tasks.

Experimental results have proved the heel disease detection and classification superiority of MedCapsNet that achieves good accuracy and robustness by taking advantage of DenseNet201's dense connectivity with CapsNet spatial awareness.

Table 4.2. Comparison of original and modified DenseNet201.

Aspect	Original DenseNet201	Modified DenseNet201
Dense Block	Use multiple densely connected layers with large growth rates.	Retains dense connectivity but uses reduced growth rates and fewer filters to minimize computational complexity.
Dense Layer	Single Conv2D layer with batch normalization and ReLU activation.	Two Conv2D layers with varied filter sizes, dropout for regularization, and “he_normal” initializer for weight optimization.
Transition Layer	Includes batch normalization, ReLU, and global average pooling.	Replaces batch normalization with dropout and global average pooling with MaxPooling2D to retain critical local features.
Convolutional Layer	No additional convolutional layers; directly processes raw input.	Adds two convolutional blocks before DenseNet, including dropout and MaxPooling2D for enhanced feature extraction.
Number of Blocks	Four dense blocks with filters (6, 12, 32, 48).	One dense block with two filters for simplicity and reduced computational overhead.
Growth Rate	Fixed at 32, suitable for large datasets but resource intensive.	Reduced to 4, balancing feature reuse and computational efficiency.
Feature Map Pooling	Global average pooling used in transition layers.	MaxPooling2D for better retention of small but diagnostically significant local features.
Weight Initialization	Standard initialization methods, prone to gradient vanishing/exploding issues.	“he_normal” initializer ensures stable training and convergence for complex medical images.

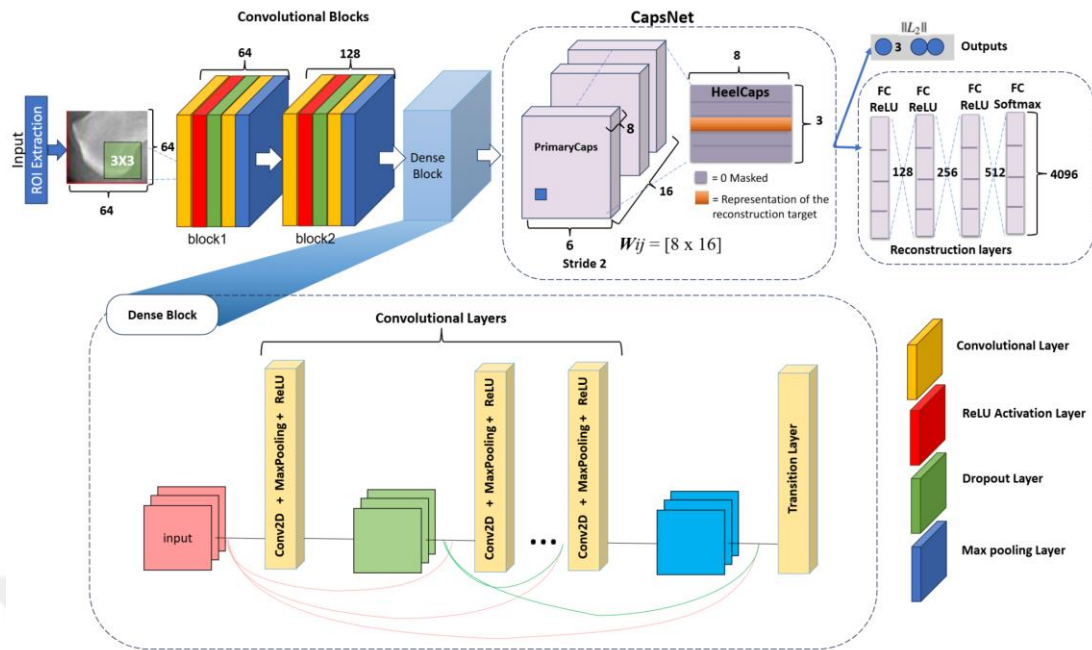


Figure 4.14. Architecture of MedCapsNet model.

PART 5

EXPREMENTS AND RESULTS

5.1. APPLICATION OF HECAPSNET ON HEEL DISEASE DATASET

The dataset on heel disease comprises 3956 X-ray images, grouped into three classes: Normal, Heel spur, and Severs disease. The images contain a lot of anatomical detail, and these were annotated by specialist orthopedics and radiology. Prior to submission of information to HeCapsNet, The X-ray images were normalized to a size of $[64 \times 64]$ pixels. Such resizing allows for the graphical information to be appropriate to the model's input layer and does not eliminate features relevant to diagnosis. The data set is then divided into balance and imbalance data sets to test the performance of the model under different conditions.

Using powerful feature extraction layers techniques, “he_normal” initializer, and modified decoder parameters, HeCapsNet was able to classify better and was trained on this dataset. Its ability to cope with variations in the quality and size of an image was subjected to rigorous tests that demonstrated its robustness and adaptability in actual medical situations.

HeCapsNet exhibited excellent performance in heel bone disease detection and classification since the accuracy achieved was 97.29% on balanced data and 94.19% on imbalanced data. More specifically, HeCapsNet had managed to yield results of an accuracy of 91.68% on the new dataset, this time though there was no “he_normal” initializer; nonetheless, once the initializer was applied, the result drastically changed to 98.37%. The model is effectively capable of shielding a high AUC of 98.69, contributing to an excellent power of discriminating between various classes of objects, namely Normal, Heel spur, and Sever's disease. Along with this, the fivefold

cross-validation achieved an accuracy of 95.77%, thereby reiterating the strength of the HeCapsNet model across different portions of the dataset.

Several important figures graphically demonstrate the performance of HeCapsNet. The training accuracy recorded in Figure 5.1 and the loss curve in Figure 5.2 indicates the overall epoch-wise performance of the model since the accuracy has been reported to improve gradually while the loss keeps on reducing through all the phases of the training. The confusion matrix gives details of how the model performed with regard to the classification by presenting the overall proportion of true and false positives and true and false negatives for all three classes: normal, heel spur, and Sever's disease, as illustrated in Figure 5.3. Additionally, the class-wise performance metrics, including precision, recall, and F1-score for each class, are detailed in Table 5.1, highlighting the effectiveness of HeCapsNet in distinguishing between these categories. Furthermore, the AUC ROC curve, which is presented in Figure 5.4, emphasizes the capacity of the model to separate the classes since AUC, which is large, provides an indication of good predictive power. In this collection of figures, the outstanding performance of HeCapsNet in the context of automated diagnosis of heel diseases is confirmed.

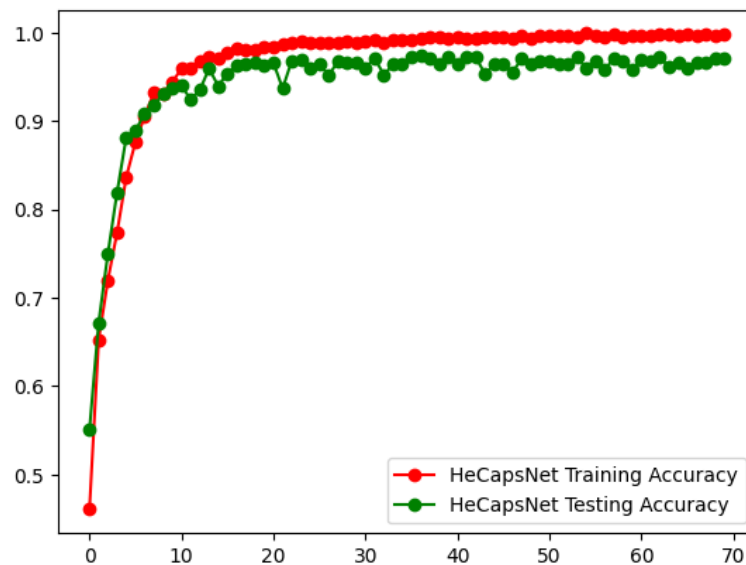


Figure 5.1. HeCapsNet training - testing Accuracy over epochs on heel dataset.

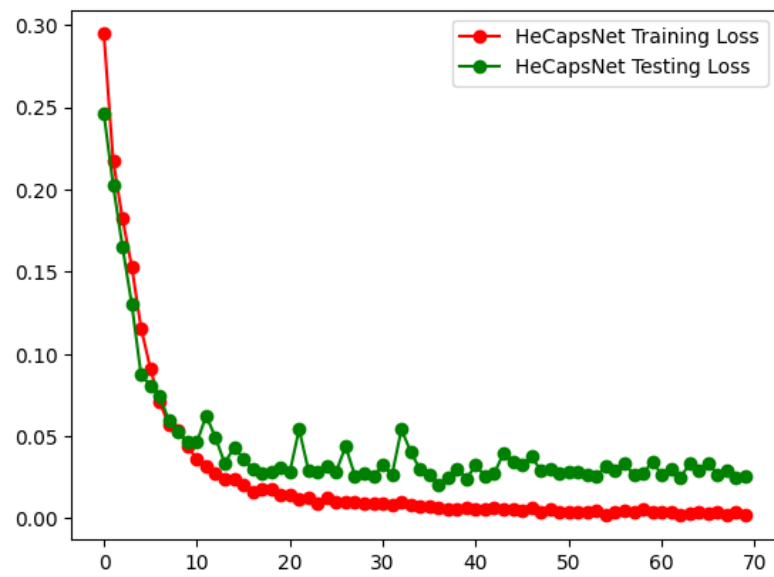


Figure 5.2. HeCapsNet training - testing loss over epochs on heel dataset.

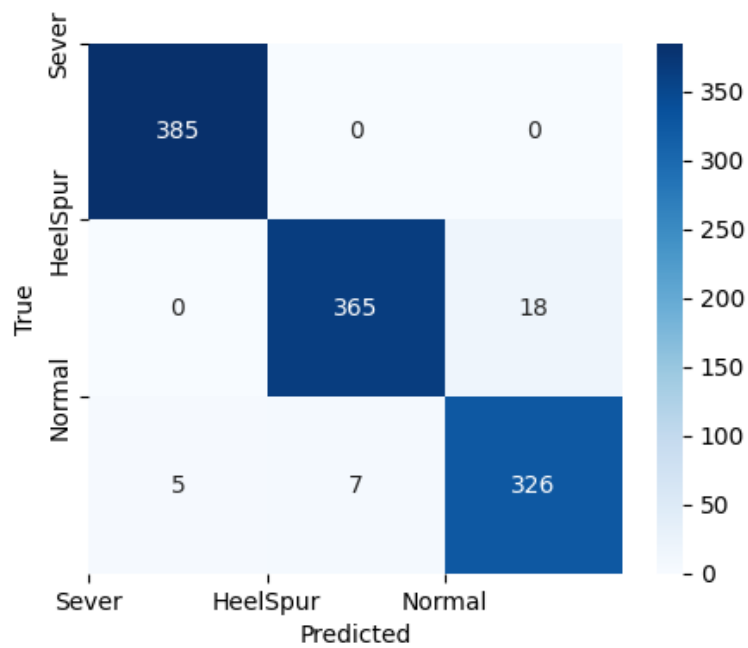


Figure 5.3. HeCapsNet AUC ROC curve on heel dataset.

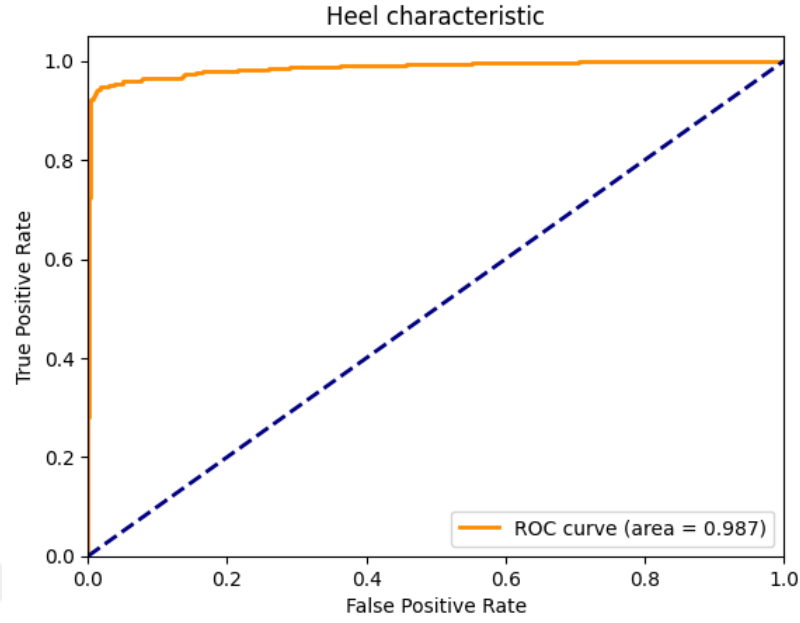


Figure 5.4. HeCapsNet AUC ROC curve on heel dataset

Table 5.1. Class-Wise Performance Metrics of HeCapsNet on Heel Disease Dataset.

Classes	Precision	Recall	F1-score
Sever	0.99	1.00	0.99
Heel Spur	0.98	0.95	0.97
Normal	0.95	0.96	0.96

5.2. APPLICATION OF HECAPSNET ON DIFFERENT DATASETS

A range of different datasets were used to systematically evaluate the performance of HeCapsNet for image classification. For this assessment, the HeCapsNet model was fully trained on the F-MNIST, MNIST, CIFAR-10, OA knee, and Brain MRI datasets. Such datasets were incorporated in order to further analyze the strengths and weaknesses of HeCapsNet in simple and complex imaging tasks.

The MNIST dataset is renowned for composing 60000 training samples and 10000 testing samples and formatting grayscale images ranging from digits 0 to 9 into $[28 \times 28]$ pixel resolution. Even though it is basic in its architecture, it is still reliable in assessing a given model's performance, and due to its nature of being easy with the features, the MNIST continues to be a core pillar for developing deep learning models. In parallel, F-MNIST can be considered as a substitute for MNIST, which consists of

10 categories of various clothing items. The set also provides 60,000 training and 10,000 testing images, and its structure is also $[28 \times 28]$ resolution, each image framed as a grayscale. These data sets set the threshold for the assessment of HeCapsNet's capability in handling grayscale images of lower complexity.

In continuation of proceeding towards advanced domains, the next challenging dataset is the CIFAR-10 dataset. This dataset, which was built by Alex Krizhevsky and Ilya Sutskever, focuses on image classification in 10 different categories that include vehicles, animals, and household items. To note our differences, as opposed to the CIFAR-10 dataset, where its images are in RGB format, both the MNIST and F-MNIST datasets do not require a model to operate or manipulate three color channels at once. Each image in the CIFAR-10 dataset is over $[32 \times 32]$ pixels, and there are a total of 50,000 training images and 10,000 testing images within it. This dataset gives an additional complexity of color features, which would make it hard for models initially built for the grayscale ones, such as the MNIST models.

In order to assess the efficacy of HeCapsNet on medical imaging tasks, two complex datasets were included in our studies, which are Knee osteoarthritis (OA) and brain tumor MRI; the Knee OA severity grading dataset hosted on Mendeley has comprehensive X-ray images that are crucial for OA grading. We picked this dataset due to its nearness to our heel disease dataset since all have complex anatomical features that can be best viewed in X-ray imaging. The Kaggle brain tumor MRI database has a total of 3,064 MRI scans and is used as a baseline to classify the diagnosis of patients with brain tumors. The complexity of composing these medical datasets stems from but is not limited to, the information on the pixels and the different types of anatomy present, as well as the requirement for precise feature construction. Training on these datasets helped to confirm the strength that HeCapsNet possesses in dealing with the problems associated with medical imaging.

The Heel Disease dataset contains 3956 images in three classes: normal, heel spur, and Sever's disease, and was created in order to test HeCapsNet for heel disease recognition and classification. All images were resized to the resolution of $[64 \times 64]$ pixels, retaining underlying important anatomical structures and compatibility with input

dimensions. Such a preprocessing step made it easier for the model to concentrate on functionality only relevant to the region of interest while disregarding computation over nonconcentrated areas.

HeCapsNet was trained on all the datasets, setting 70 epochs of training iteration. In terms of medical images, where high-resolution features and complex pixel variations are more difficult challenges, the model's performance clearly outperformed the initial version of the CapsNet. HeCapsNet faced datasets in the form of Brain MRI, Knee OA, and Heel Disease, which already had rich anatomical structures and varied quality of images, which made it suitable to test its advanced feature extraction and classification capabilities. In contrast, more straightforward tasks were posed by easier datasets such as MNIST, F-MNIST, and CIFAR-10, which all included 60,000 images.

According to Table 5.2, HeCapsNet obtained sufficient results for all datasets. For the Knee OA dataset, the model managed to achieve 83.42% accuracy, which is even better than the original 78.82% performance of the CapsNet. For the Heel Disease dataset, HeCapsNet obtained 97.29% accuracy for the balanced dataset, 94.19% for the imbalanced dataset, and 95.77% with five cross-validations. With the use of the "he_normal" initializer, HeCapsNet achieved a knee dataset accuracy of 83.42%, while the knee dataset accuracy without the technique was 77.35%. CapsNet achieved an accuracy of 91.68 and 97.29 percent, respectively, for our new dataset with and without the use of a "he_normal" initializer. Results indicate how the model can manage to be trained on datasets of different complexities and balancing.

Comparative performance analysis with other works, as shown in Table 5. 3, indicates that HeCapsNet is superior in achieving accuracy on different HeCapsNet datasets. For example, for the datasets considered, HeCapsNet was able to accomplish the following results: 99.51% on MNIST, 93.33% on F-MNIST, 81.94% on CIFAR-10, and 98.37% (balanced) on Brain MRI datasets. These figures are well above what has been achieved using the original CapsNet and many enhanced capsule-based approaches, which in turn demonstrate the better generalization and performance consistency of HeCapsNet across diverse areas.

Table 5.2. Accuracy comparison between HeCapsNet and original capsule models for knee and heel disease datasets.

Model	Knee (OA)	Heel Dataset
Original CapsNet	78.82%	73.99%
Proposed enhanced CapsNet (Without he_normal initializer)	77.35%	91.68% (balanced dataset)
Proposed CapsNet (HeCapsNet)	83.42%	97.29% (balanced dataset) 94.19% (imbalanced dataset) 95.77% (fivefold cross-validation)

HeCapsNet has shown its versatility by including both medical and non-medical datasets. Its capability to manage complicated imaging tasks, especially in medical applications, indicates its potential use in actual diagnostic systems. HeCapsNet's advanced approach combines architectural innovations and rigorous training and clearly illustrates the benefits of HeCapsNet over conventional architectures in medical images, opening up great opportunities for researchers and practitioners in medical imaging.

In this part, we described HeCapsNet, a modified fully convolutional CapsNet that is aimed at automatically identifying and classifying heel ailments from the lateral foot X-ray images. To this end, we carried out a series of experiments employing our recently proposed model in addition to various datasets, which were MNIST, Fashion-MNIST, CIFAR-10, brain tumor MRI, and knee osteoarthritis (OA) X-rays. The results confirmed that HeCapsNet exceeded the original CapsNet, especially on the more complicated medical datasets.

When HeCapsNet is compared with other modified CapsNet found in literature, we experimentally observed similar effectiveness for it on the simpler datasets such as MNIST and F-MNIST. On the other hand, HeCapsNet achieved better performance when dealing with more complex medical datasets like brain MRIs, knee X-rays, muscle diseases, and heel disease datasets. In particular, for the heel disease dataset, the proposed model recorded accuracy rates of 97.29% on balanced data, 94.19% on imbalanced data, an AUC of 98.69%, and 95.77% accuracy regarding fivefold cross-validation.

Table 5.3. Performance (accuracy) comparison between proposed and state of the art models.

Models	MNIST	F-MNIST	CIFAR-10	Brain MRI
[62] [12] Original CapsNet	99.23% [12]	92.49% [62]	62.98% [62]	78.69% [62]
[61] Improved capsule network	99.65%	-	82.31%	-
[101] Optimized capsule network	99.87%	-	82.45%	94.85%
[63] Modified capsule network	-	-	-	90.89%
[64] Improved capsule network	-	-	-	86.56%
Proposed enhanced CapsNet (without he_normal initializer)	99.46%	93.21%	82.84%	96.98%
Proposed enhanced CapsNet (HeCapsNet)	99.51%	93.33%	81.94%	98.37% (balanced) 93.59% (imbalanced)

The initial design of the CapsNet had issues with medical images for two reasons: first, it had too few convolutional layers, which rendered it incapable of properly feature extracting from intricate images, and second, the raw kernel initializer that was deployed in conjunction with the ReLU activator was ineffective for medical images that contained heavy pixel detail. To address these issues, we implemented a “he_normal” kernel initializer on the convolutional layer. These modifications greatly improved the results that HeCapsNet delivered when deployed on complicated datasets.

Furthermore, adding more convolutional blocks before the capsule layers enabled the model to extract better features. These changes increased accuracy by allowing the network to encode hierarchical structures in the data and improve the encoding of spatial features. The additional convolutional blocks were also useful for regularization and lessening overfitting in order to enhance generalization on new data.

Our 3,956 X-ray images of the lateral foot have been annotated from a reliable dataset that seeks to bridge the existing gap in heel diseases such as Sever’s disease and heel

spurs. Maintaining results that seem to be well substantiated does come with its share of challenges. As noted, the offered set of comprehensively analyzed images can facilitate reliable diagnoses, although expanding the set may assist in enhancing the capacity of the system or the model offered. Furthermore, the model is capable of only diagnosing two conditions, Sever's disease and heel spurs, and thus is unable to assist in identifying other conditions affecting the heel or even the fractures.

In practical terms, this system could be integrated into X-ray machines or used on its own for the first step in disease detection. It could also be enhanced into a web application where users can upload X-ray images, and the application gives a diagnosis. This would enable the system to be available to medics and help orthopedists give faster and more accurate first-contact diagnoses of heel illnesses.

5.3. APPLICATION OF MEDCAPSNET AND OTHER MODELS TO HEEL DATASET

The X-ray heel dataset forms a special category described in 4.1.2 and 4.1.3 with what can be termed three distinct classes: 1842 images, which can be termed as 'normal,' 1316 images, which can be dubbed the brusque, heel spur, and 798 images that captured the severe backlash in the region. The dataset then contained a total of 3,956 images comprising all classes. To facilitate disease identification in the heel bone images and to optimize network efficiency by excluding irrelevant areas, a bounding box method was used to create the region of interest (ROI).

For the annotation and bounding box drawing of the ROI, the labelImg v1.8.6 software was applied. It is important to note that all the images went through basic preprocessing procedures, such as cropping to the ROI and then scaling the image sizes to $[64 \times 64]$ pixels, prior to being incorporated into the experimental arrangement. Considering the challenge posed by the unbalanced distribution of the images across the dataset classes, techniques for data balancing, especially oversampling methods, were employed to achieve better representation. Oversampling was carried out using the RandomOverSampler from the imblearn.over_sampling library to equalize the distribution of the three classes: Sever, Heel spur, and Normal. Some more usage

instances were added to the dataset. Originally, there were only 3,956 images available. The method introduces random samples from classes with a lower number of images until all classes have the same number as the class of interest. In this case, the aim was to have 1,842 images for each individual class. Consequently, the total number of images in the dataset was raised to 5,526, with an even distribution among the classes. The results of performing the mentioned method render the training of machine learning models free from class imbalance as all the classes are better represented, and no single class has a greater chance of skewing the bias of the algorithms. The balancing of the heel diseases dataset is shown in Figure 5.5.

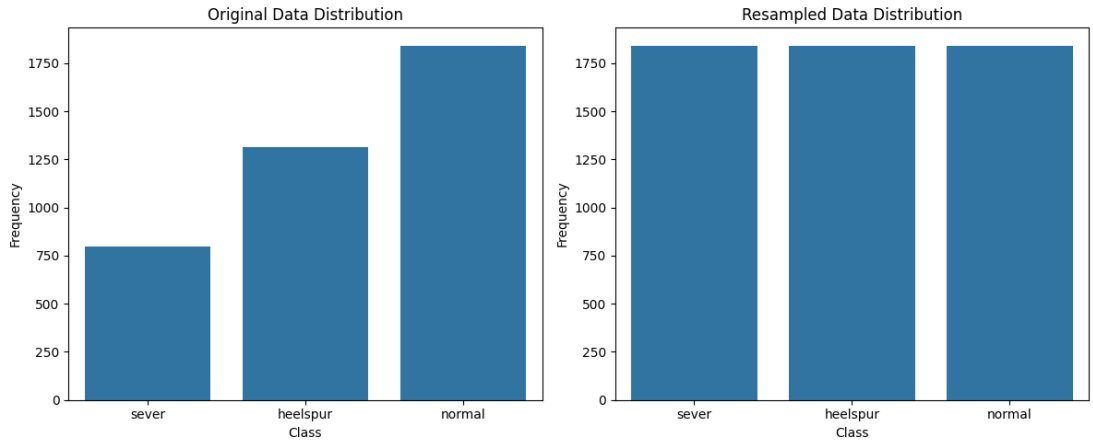


Figure 5.5. Distribution of balancing heel diseases dataset.

We have implemented the original CapsNet [12] Res2Net+Caps [78], FixCaps [76], and BoostCaps [77] models according to the descriptions in their respective papers. To ensure the correct implementation of the existing models, we first trained the models using the datasets that were employed in their studies. After achieving results consistent with those reported in the papers, we proceeded to train the models with the heel dataset.

Table 5.4 shows the model accuracy values of the original CapsNet and the other CapsNet compared with MedCapsNet on the heel dataset. In our implementation, while original CapsNet showed an accuracy of 73.99%, Res2Net+Caps showed an accuracy of 34.81%, and FixCapsNet showed an accuracy of 73.33% on the heel dataset. Regarding BoostCaps in [77], we encountered difficulties in method implementation due to some of the architecture components and parameters not being

clearly specified. MedCapsNet demonstrated 96.38% accuracy, as the epochs shown in Figure 5.6 and the loss epochs shown in Figure 5.7. Furthermore, the AUC score is 98.27%, as demonstrated in Figure 5.8, along with the corresponding confusion matrix in Figure 5.9. Additionally, the class-based performance metrics, including precision, recall, and F1-score for each class, are detailed in Table 5.5, thereby highlighting the effectiveness of MedCapsNet in distinguishing between heel dataset classes.

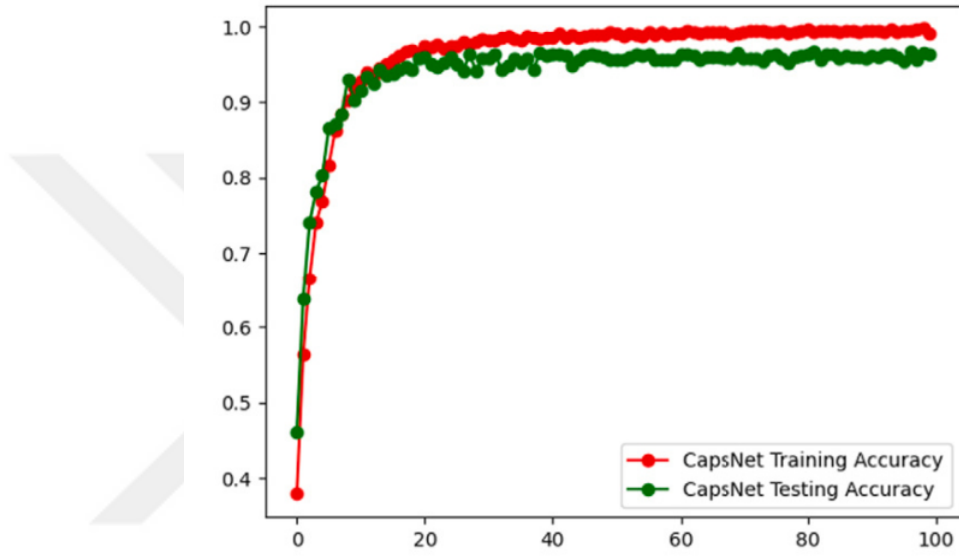


Figure 5.6. Training and testing accuracy of heel diseases dataset.

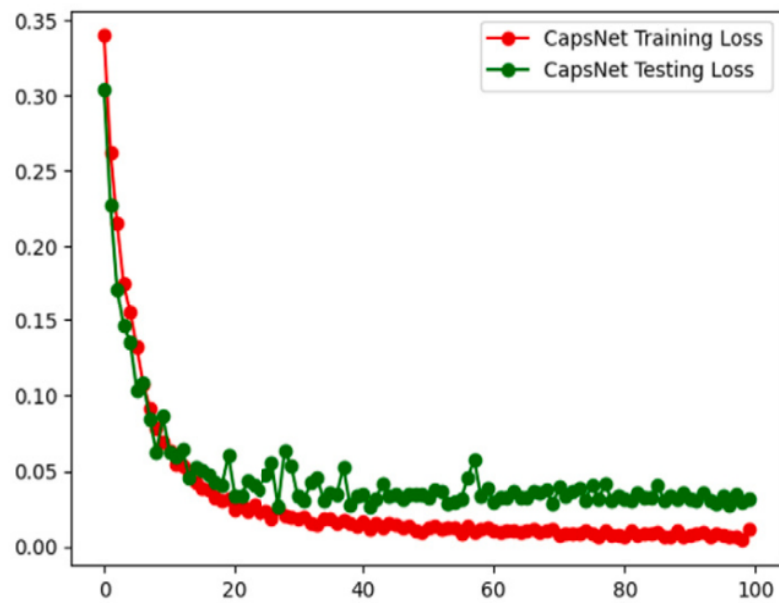


Figure 5.7. Training and testing loss of heel diseases dataset.

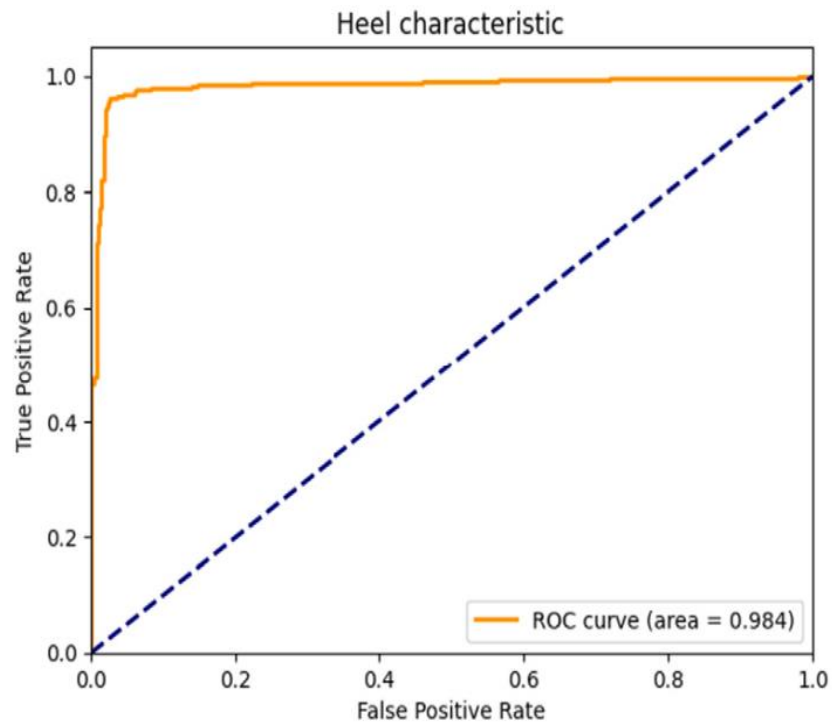


Figure 5.8. AUC curve for heel diseases dataset.

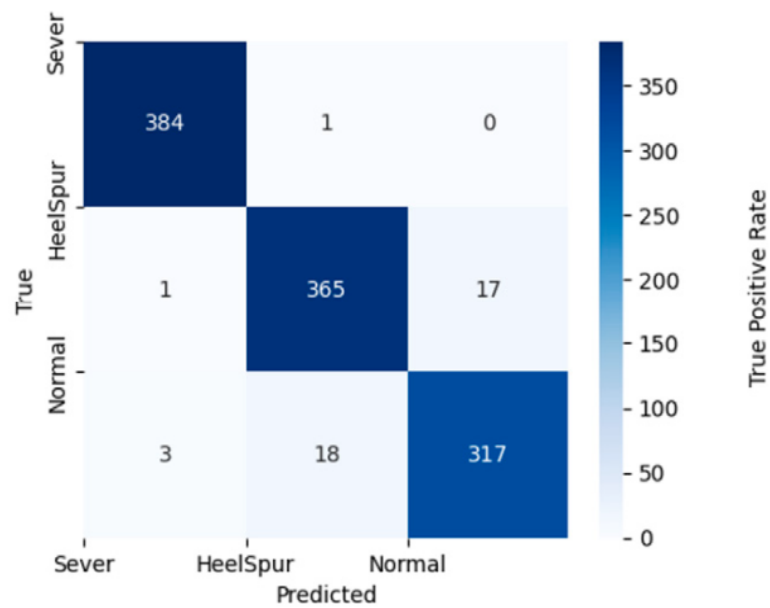


Figure 5.9. Confusion matrix of heel diseases dataset.

Table 5.4. Results in comparison with base and recent capsule network models for heel dataset.

Architecture	Accuracy
Original CapsNet [12]	73.99%
Res2Net+Caps [78]	34.81%
FixCaps [76]	73.33%
BoostCaps [77]	-
MedCapsNet	96.38%

Table 5.5. Class-wise performance metrics of MedCapsNet on heel disease dataset.

Classes	Precision	Recall	F1-score
Sever	0.99	1.00	0.99
Heel Spur	0.95	0.95	0.95
Normal	0.95	0.94	0.94

Furthermore, we implemented a 5-fold cross-validation method on MedCapsNet. The full heel dataset of 5526 images has been split into 70% for training and 30% for testing. For training the model we used the part of 70%, while for the evaluation of model performance we used the part of 30% which split for the test and never exposed to the model training, also not the part that k-fold that splitting it for internal evaluation in iterations steps. The result shows that the model performance evaluation was done with a cross-validation accuracy of 95.69%, as its epoch is shown in Figure 5.10, and the loss epochs are shown in Figure 5.11. Additionally, the ROC AUC is 98.87%, as demonstrated in Figure 5.12, along with the corresponding confusion matrix in Figure 5.13. Additionally, the class-wise 5-fold cross-validation performance metrics, including precision, recall, and F1-score for each class, are detailed in Table 5.6, highlighting the effectiveness of MedCapsNet in distinguishing between these categories.

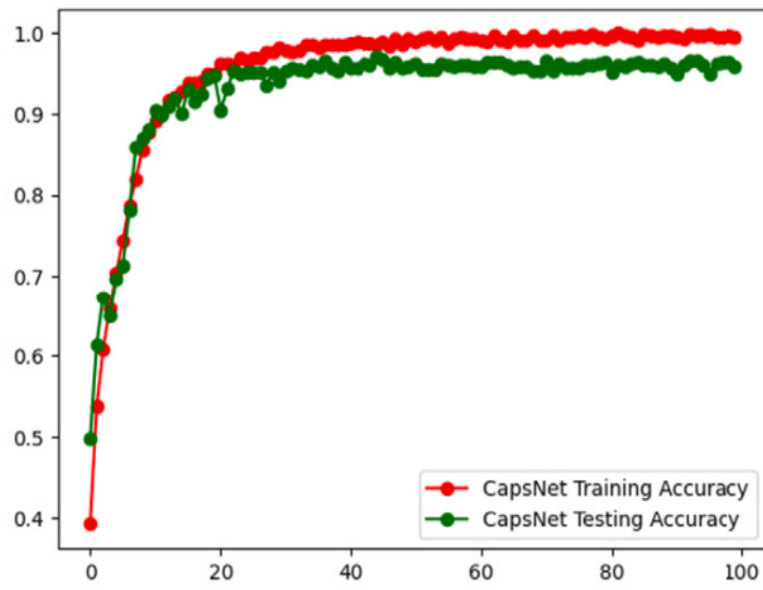


Figure 5.10. Training and testing accuracy of cross validation.

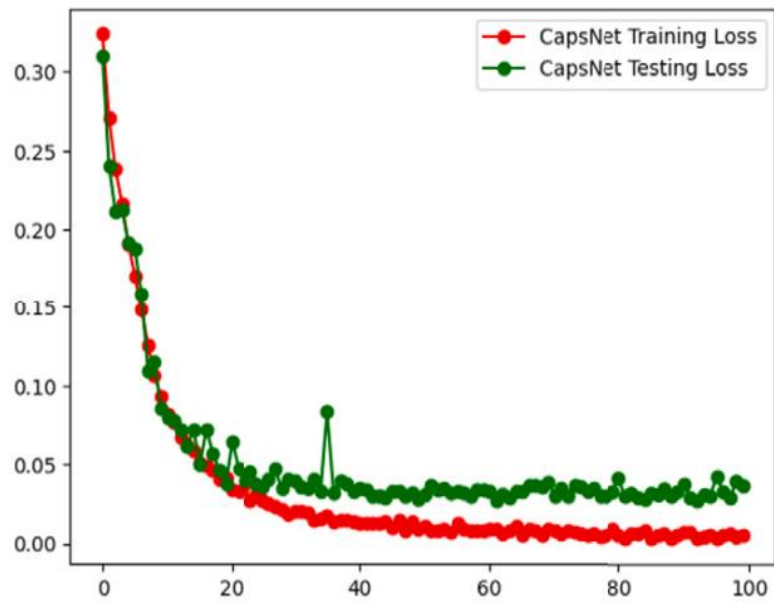


Figure 5.11. Training and testing loss of cross validation.

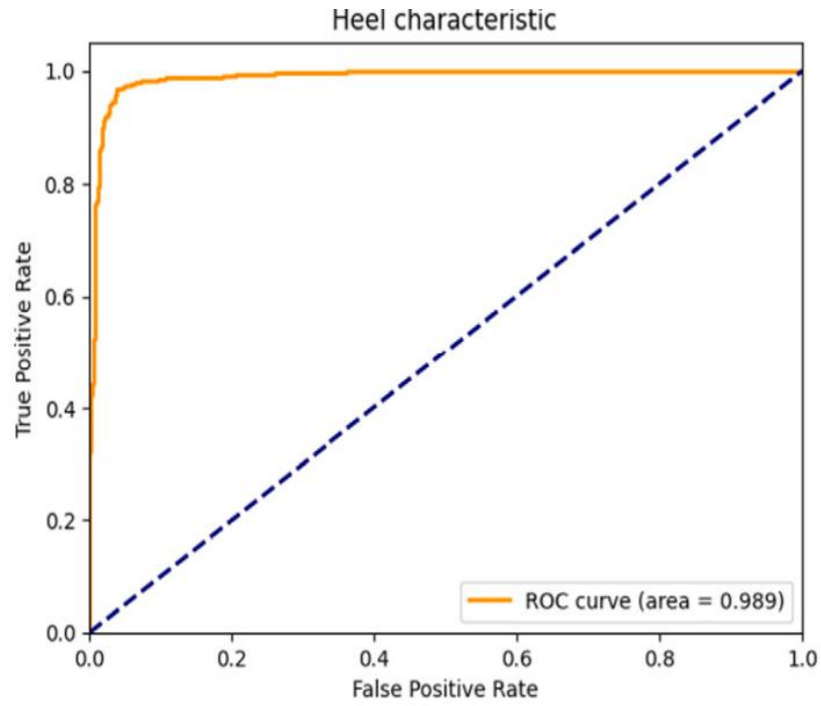


Figure 5.12. AUC curve of 5-fold cross validation of heel dataset.

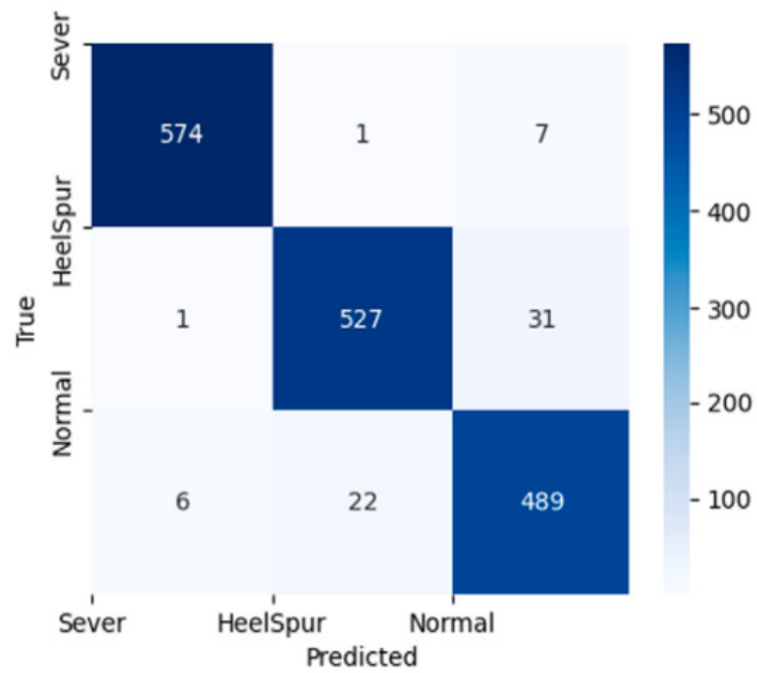


Figure 5.13. Confusion matrix of 5-fold cross validation of heel dataset.

Table 5.6. Class-Wise of cross validation performance metrics of MedCapsNet on heel disease dataset.

Classes	Precision	Recall	F1-score
Sever	0.99	0.99	0.99
Heel Spur	0.96	0.94	0.95
Normal	0.93	0.95	0.94

5.4. EXPERIMENTS OF MEDCAPSNET ON ADDITIONAL MEDICAL DATASETS

We evaluated the performance of MedCapsNet for various medical datasets, including Breast_BreaKHis_v1, HAM10000 skin cancer dataset, and Jun Cheng's Brain Imaging MRI Dataset, to demonstrate its superiority compared to existing studies. For this, we tested the model on the datasets used by the studies in the recent enhancement of the CapsNet and compared the results. In these experiments, the 5-fold cross-validation method was implemented using the `kf.split` algorithm to evaluate the performance, and the full data was split into 80% for training and 20% for testing. The performance results show that the model can achieve high accuracy with small and large medical image datasets. Table 5.10 presents the cross-validation accuracy results of MedCapsNet across various datasets, comparing its performance with the accuracy of the other models. Additionally, we used 100 epochs in each experiment to monitor the training and loss values in each iteration. All experiments were conducted using Google Colab environments, which included Python v3 as the runtime type, high RAM, and T4 GPU as the acceleration hardware.

5.4.1. Breast BreakHis v1 Dataset

The BreakHis dataset, known as the Breast Cancer Histopathological Image Classification, is a collection of 9,109 microscopic images of breast tumor tissue. These images were obtained from 82 patients and were captured at various magnification scales, namely 40X, 100X, 200X, and 400X. The dataset consists of 2,480 benign samples and 5,429 malignant samples. Each image has dimensions of 700 x 460 pixels and is represented in RGB color space with 3-channel information. The pixel depth in each channel is 8 bits, and the images are stored in PNG format.

[120]. This experiment used a small subset of images from a breast cancer dataset, consisting of 1000 images rescaled to $[65 \times 65]$ pixels and converted to 1 color channel in grayscale mode. The total images were used for training 800, while the remaining 200 images were used to test the performance with a limited sample size. The experiment yielded a 5-fold cross-validation accuracy of 98.40% for training and testing as its epochs demonstrated in Figure 5.14 and loss epochs of training and testing demonstrated in Figure 5.15. Along with the corresponding confusion matrix, which is demonstrated in Figure 5.16, and the AUC of 98.96%, as demonstrated in Figure 5.17. Our proposed MedCapsNet model demonstrates superior performance with a small sample of the dataset without using data augmentation compared to the model presented in [78], which achieved an accuracy of 95.6% using data augmentation. Additionally, the class-wise performance metrics, including precision, recall, and F1-score for each class, are detailed in Table 5.7, highlighting the effectiveness of MedCapsNet in distinguishing between these categories of breast cancer.

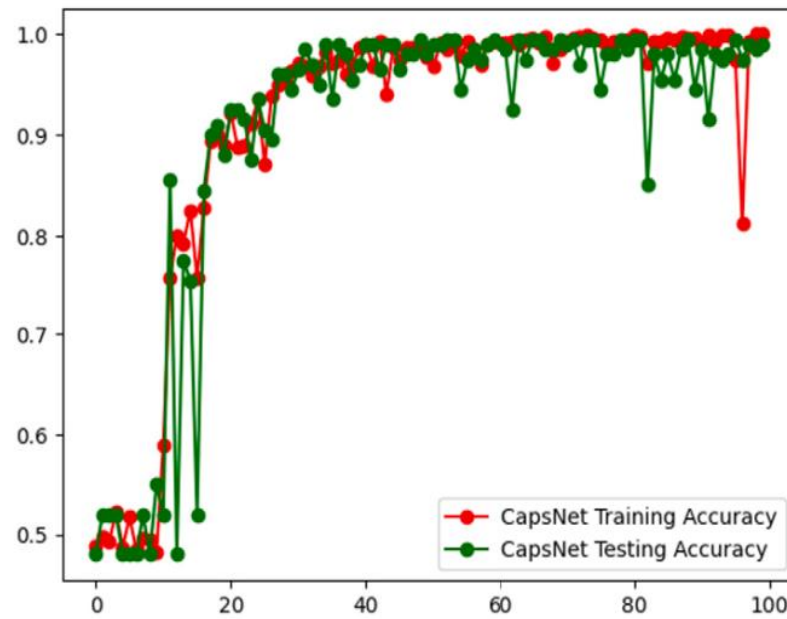


Figure 5.14. Training and testing accuracy of breast BreakHis v1.

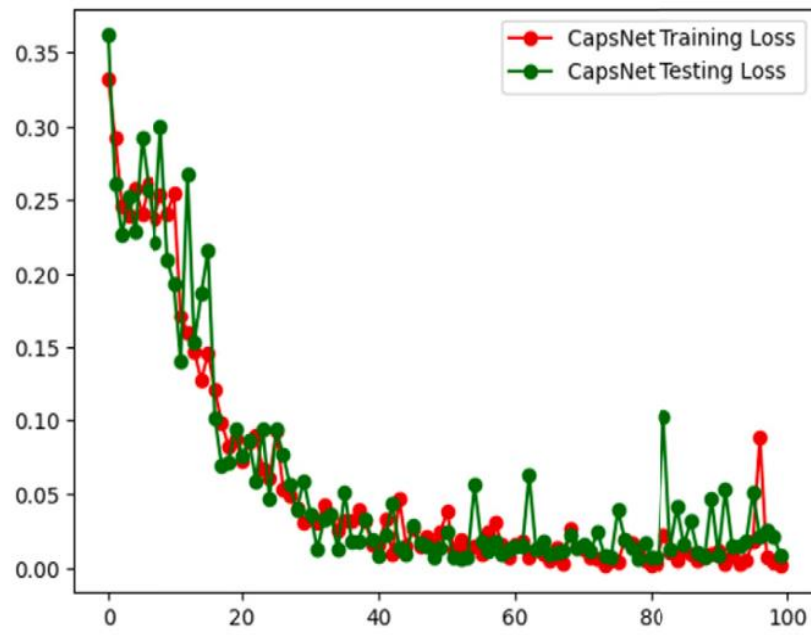


Figure 5.15. Training and testing loss of breast BreakHis v1.

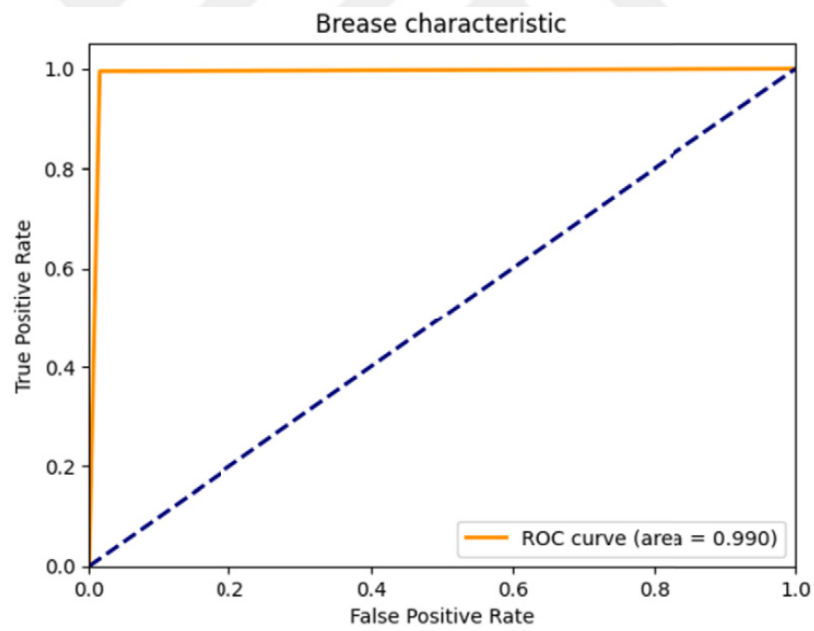


Figure 5.16. AUC curve of breast BreakHis v1.

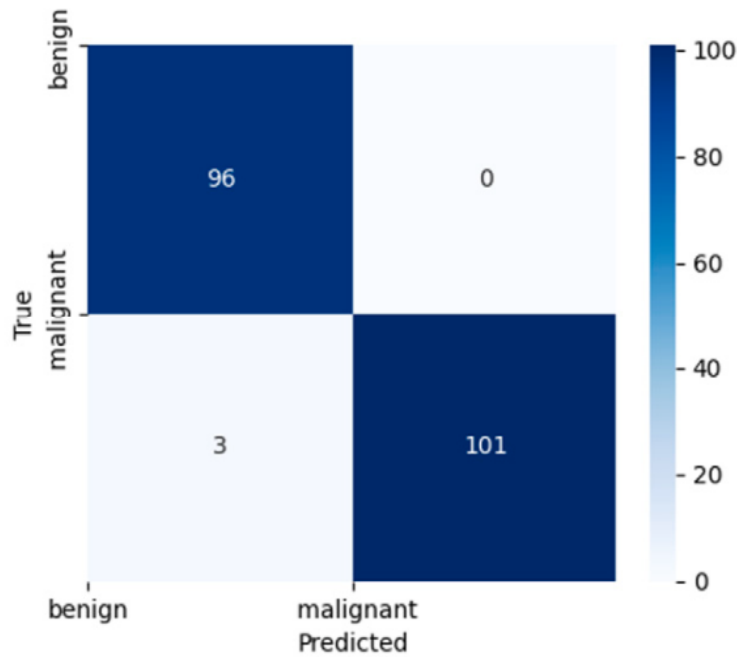


Figure 5.17. Confusion matrix of mrcast BreakHis v1.

Table 5.7. Class-Wise performance metrics of MedCapsNet on breast BreakHis v1.

Classes	Precision	Recall	F1-score
Benign	0.99	0.99	0.99
Malignant	0.99	0.99	0.99

5.4.2. HAM10000 Skin Cancer Dataset

The HAM10000 dataset, provided by the Harvard Dataverse Organization, addresses the limited size and diversity of existing dermatoscopic image datasets for training neural networks in the diagnosis of pigmented skin lesions. It consists of 10,015 dermatoscopic images collected from different populations and acquisition modalities. The dataset covers important diagnostic categories such as melanoma, basal cell carcinoma, and benign keratosis-like lesions [121]. The dataset is provided in 3 color channels on which the proposed model was trained in RGB mode. In addition, since the provided dataset is imbalanced, the oversampling method has been used to balance the dataset and also make the training number of images in large numbers, including 37548 images. The result shows the high performance of the model, achieving cross-validation accuracy of 5-fold of 98.29%, as its epochs are given in Figure 5.18, and

loss for training and testing is given in Figure 5.19. AUC of 99.00%, as demonstrated in Figure 5.20, along with the corresponding confusion matrix given in Figure 5.21. Our proposed MedCapsNet model shows superior performance compared to the model presented in [76], which achieved an accuracy of 96.49%. Table 5.5 demonstrates the precision, recall, and F1-score for each class, thereby indicating the effectiveness of MedCapsNet in recognizing the distinct characteristics of the heel dataset classes.

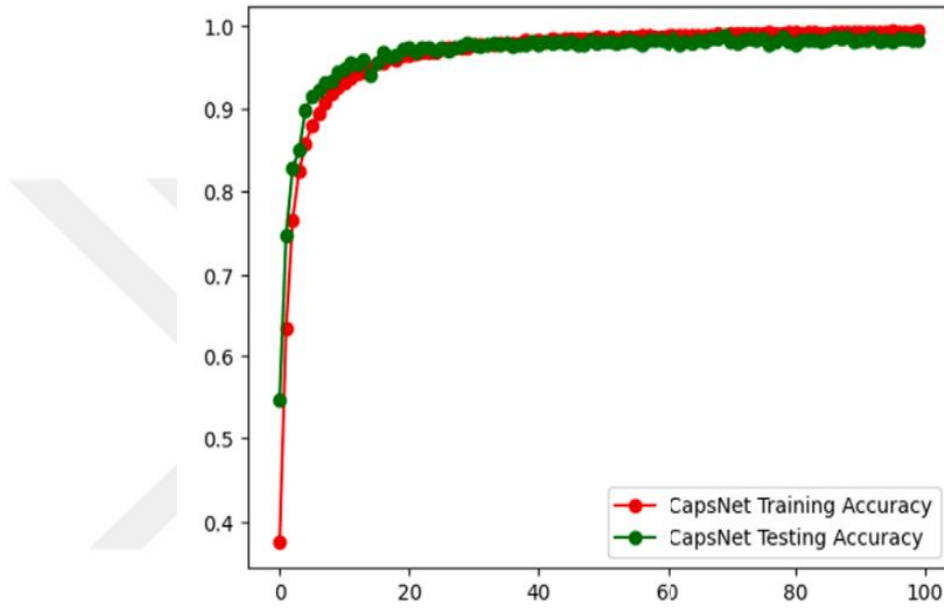


Figure 5.18. Accuracy of HAM10000 skin cancer dataset.

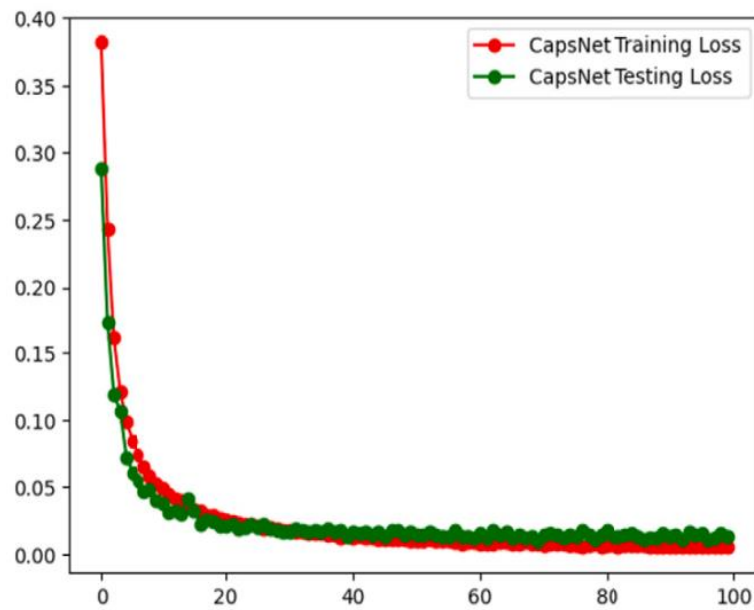


Figure 5.19. Loss of HAM10000 skin cancer dataset.

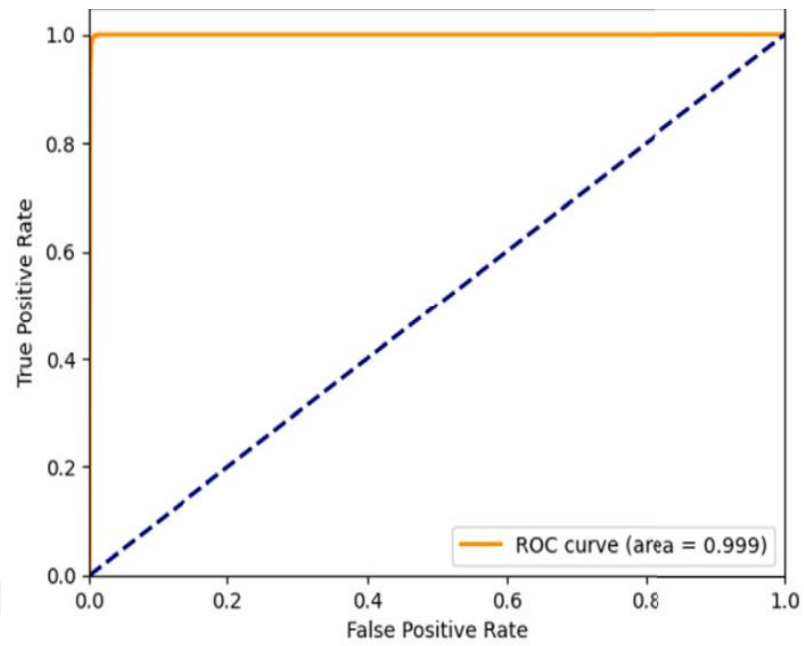


Figure 5.20. AUC curve of HAM10000 skin cancer dataset.

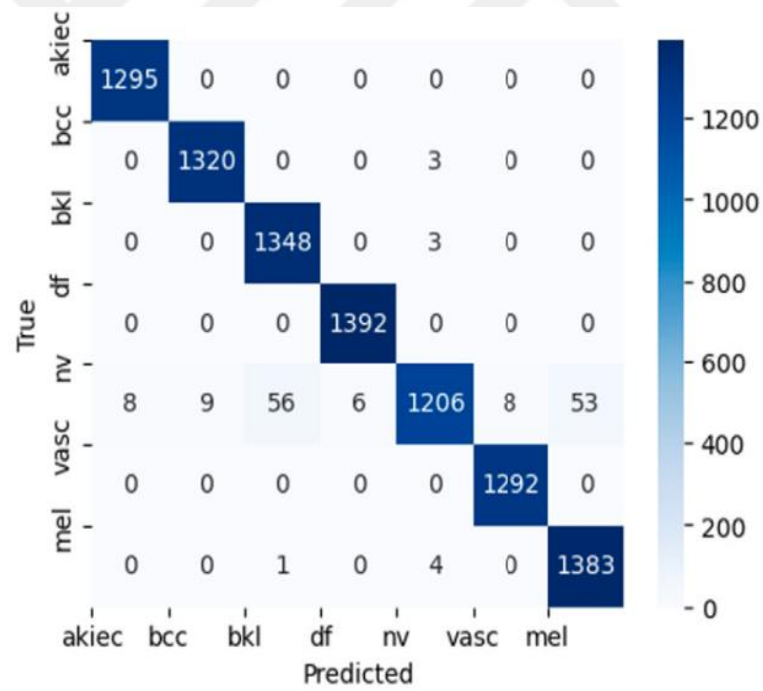


Figure 5.21. Confusion matrix of HAM10000 skin cancer dataset.

Table 5.8. Class-Wise performance metrics of MedCapsNet on HAM10000 skin cancer.

Classes	Precision	Recall	F1-score
Actinic	0.99	1.00	1.00
Malinga	0.99	1.00	1.00
Benign	0.96	1.00	0.98
Dermatofibroma	1.00	1.00	1.00
Melanocytic	0.99	0.90	0.94
Pyogenic	0.99	1.00	1.00
Melanoma	0.96	1.00	0.98

5.4.3. Jun Cheng’s Brain Imaging MRI Dataset

The brain tumor dataset has 3064 images representing three different categories of brain tumors: meningioma, glioma, and pituitary tumor, published by Jun Cheng [122]. The provided dataset was imbalanced; therefore, oversampling was used to balance three classes. The total number of images after balancing is 4246. The proposed MedCapsNet model was trained on the dataset with $[64 \times 64]$ image size in 3-channel RGB mode. The result shows the high performance of the model, achieving cross-validation accuracy 5-fold of 97.67%, as its epochs curve given in Figure 5.22 and loss epochs shown in Figure 5.23. The AUC of 99.80% is demonstrated in Figure 5.24, along with the corresponding confusion matrix in Figure 5.25. Our proposed MedCapsNet model achieved a higher accuracy than the boosted CapsNet presented by [77], which reported an accuracy of 92.45%. Additionally, the class-based performance metrics, including precision, recall, and F1-score for each class, are detailed in Table 5.9, highlighting the effectiveness of MedCapsNet in distinguishing between these categories of brain tumor.

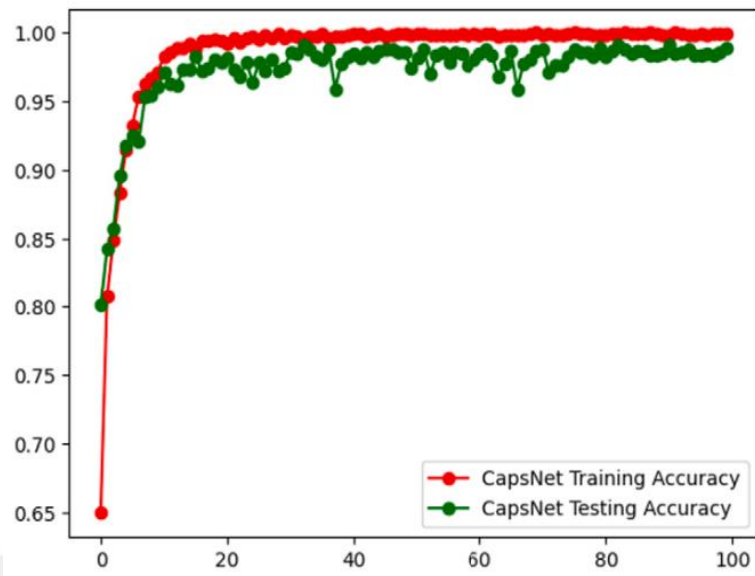


Figure 5.22. Accuracy of Jun Cheng's brain imaging MRI dataset.

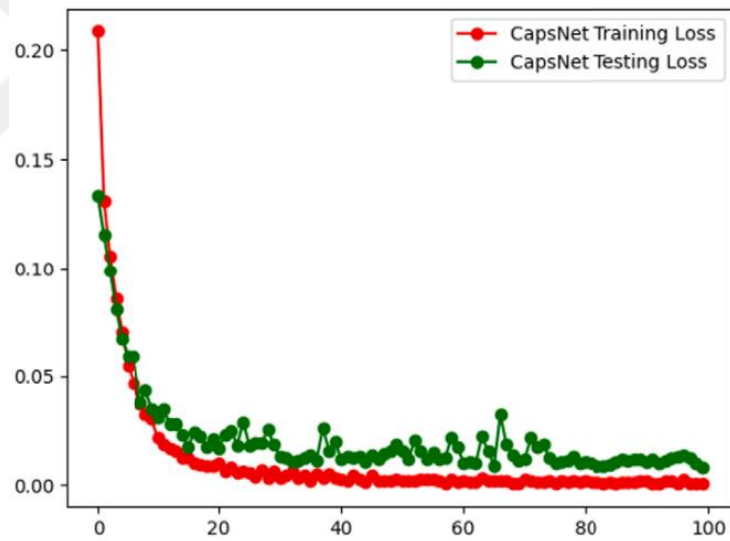


Figure 5.23. Loss of Jun Cheng's brain MRI dataset.

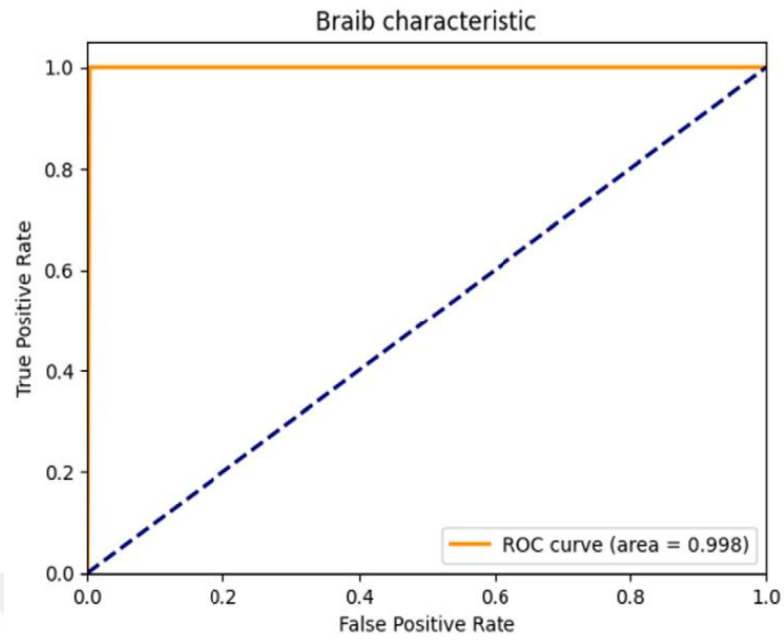


Figure 5.24. AUC curve of Jun Cheng’s brain imaging MRI dataset.

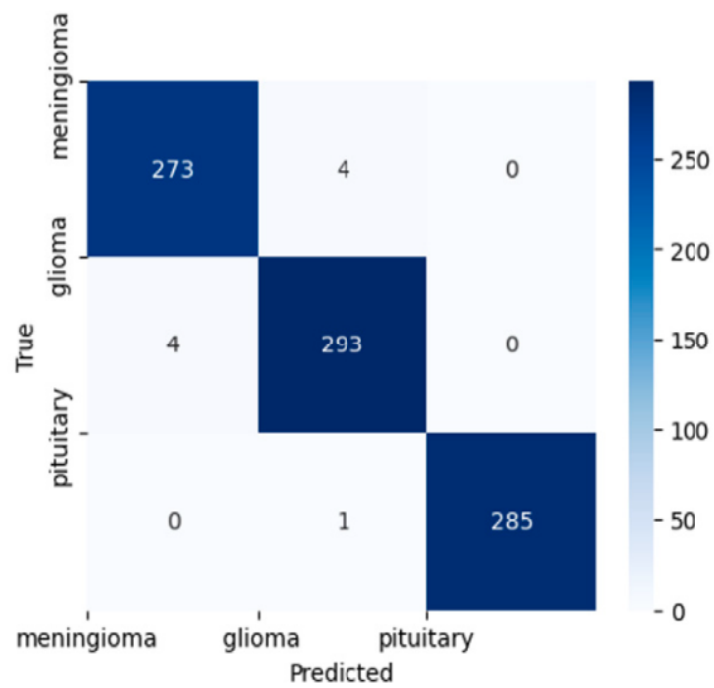


Figure 5.25. Confusion matrix of Jun Cheng’s brain imaging MRI dataset.

Table 5.9. Class-Wise performance metrics of MedCapsNet on Jun Cheng brain tumor dataset.

Classes	Precision	Recall	F1-score
Meningioma	0.96	0.98	0.97
Glioma	0.99	0.95	0.97
Pituitary	0.98	1.00	0.99

In addition to the results Table 5.10, present comparison for various medical datasets with MedCapsNet against other models showing sufficient performance of the model compared to other models.

Table 5.10. Results in comparison for various medical datasets with MedCapsNet vs. other models.

Dataset	MedcapsNet (5-fold CV Accuracy)	Other Models (Accuracy)
BreakHis_v1 [120]	98.40% (without augmentation)	Res2Net+Caps [78]: 95.6% (with augmentation)
HAM10000 [121]	98.29%	FixCaps [76]: 96.49%
Jun Cheng [122]	97.67%	BoostCaps [77]: 92.45%

In this part, MedCapsNet was introduced as a hybrid model that combines the strengths of DenseNet201 and CapsNet to address the limitations of traditional CapsNets in processing complex medical images. By incorporating a modified DenseNet201 architecture with advanced feature extraction capabilities prior to the capsule layers, the model demonstrated significant improvements in classification accuracy for the heel disease dataset as well as other challenging medical datasets.

The main advancements encompassed the inclusion of convolutional layers in the feature extraction, addressing the vanishing/exploding gradient problem using the “he_normal” kernel initializer and utilizing dropout and MaxPooling2D for regularization. These improvements enabled MedCapsNet to obtain hierarchical relations, spatial relations, and complex structural details in medical images. Experimental results affirmed its better yield than so far existing approaches without modification or specific tuning of the architecture and massive data augmentation for other datasets.

The model MedCapsNet has only been tested on a limited number of data. Therefore, considerations for further revisions could entail texts using more extensive datasets that contain a mix of various heel diseases and even more advanced medical imaging tasks. Increasing the data available would improve the model's accuracy and general reliability, allowing it to be used in a wider array of scenarios.

One further promising direction is to adapt MedCapsNet to the analysis of multi-modality datasets. There is a possibility of creating a more universal diagnostic tool by merging X-rays and MRIs. In this manner, the individual advantages of various modalities would be exploited so as to enhance disease description and the diagnosis process.

Optimizing the existing lightweight models of MedCapsNet towards real-time implementation in MedCapsnet is the other area one can look forward to examining. These optimized versions of MedCapsNet can be used embedded in medical imaging systems or in cell phones, thus improving the practical use and effectiveness of this technology. Also, as clinics may be assisted by the model's decision within the process, explainability via saliency maps and feature visualization can be integrated as part of the therapy. Such techniques would increase the trust level towards MedCapsNet by making its operations more transparent and facilitating its use within clinical workflows.

In addition, future studies might also evaluate MedCapsNet against other cutting-edge techniques, including ViTs, GNNs, etc. Such comparisons would explain unique characteristics that can be complementary and suggest opportunities for integration, which may broaden the horizons of medical image analysis.

Finally, adding patient age, gender, and medical history data would be a step toward transforming MedCapsNet into a specific diagnostic agent. Such evolution would allow better projection and further customization of treatment, which already signifies relevance to the precision medicine concept in relation to medical images.

5.5. ASPECTS OF COMPARATIVE MODELS OVER A SHARED DATASET

Concerning medical image classification, the transition from CapsNet to HeCapsNet and MedCapsNet was critical, especially with intricate datasets such as the heel disease dataset. Every one of these models performed better than the previous version by introducing architectural improvements that overcame the shortcomings of the earlier model. For example, while CapsNet’s dynamic routing mechanism allowed for spatial relationship representation, its minimalist design and lack of multi-scale feature extraction made it unsuitable to high-resolution medical images with a performance of only 73.99% accuracy and 66.95% AUC on the heel disease dataset.

Various modifications were made to HeCapsNet, among them being the addition of convolutional layers before capsule layers, the use of “he_normal” kernel initializer, and dropout layers to handle overfitting and class imbalance problems, respectively. These led to a significant improvement in performance whereby 97.29% accuracy and 98.69% AUC on the heel disease dataset and 98.37% accuracy on the Brain MRI dataset were achieved, respectively. This thus warranted architectural refinement for capsule networks used in complex medical imaging.

MedCapsNet took advantage of these enhancements by fusing CapsNet with DenseNet201 through the introduction of convolutional layers before DenseNet blocks, MaxPooling2D in transition layers, and replacing batch normalization with dropout for better generalization. It demonstrated superior adaptability by obtaining an accuracy rate of 96.38%, with an AUC value equal to 98.27% on the heel disease data set; for the brain MRI data set, it obtained an accuracy rate equal to 97.67%, an AUC of 99.80%; however, it managed an accuracy rate worth mentioning which is above average on HAM10000 data set, i.e., equal to 98.29%.

Comparatively, CapsNet had a hard time with difficult data types. HeCapsNet overcame this problem with improved feature extraction and regularization. MedCapsNet integrated DenseNet201 and produced a generalizable and reliable model applicable to any dataset yet still outperformed other models on specialized

tasks and showed adaptability across multiple datasets. These models are compared in Table 5.11 on common datasets.

Table 5.11. Performance Comparison of Capsule Network Models.

Dataset	Metric	CapsNet	HeCapsNet	MedCapsNet
Heel Disease	Accuracy	73.99%	97.29%	96.38%
Heel Disease	AUC	66.95%	98.70%	99.80%
Brain MRI (Jun Cheng)	Accuracy	N/A	98.37%	97.67%
Brain MRI (Jun Cheng)	AUC	N/A	99.74%	99.80%

PART 6

DISCUSSION

6.1. ANALYSING OF PROPOSED MODELS RESULTS

The success of HeCapsNet and MedCapsNet in classifying medical images is also demonstrated within the scope of heel ailments. HeCapsNet, which incorporates improved convolutional blocks with a normal kernel initializer, showed a good advancement over basic CapsNet. It registered an accuracy of 97.29% on balanced heel disease datasets, 94.19% on imbalanced ones, and an AUC of 98.69%, with the fivefold cross-validation accuracy being 95.77%. These metrics indicate its capacity to function in complicated medical imaging environments, characterized by fine pixel combinations and various anatomical structures.

Equally, MedCapsNet, which implements a modified DenseNet201 architecture, attained broader applicability across multiple datasets. When Deep Networks' feature extraction architecture was fused with the spatial hierarchical comprehension of CapsNet, MedCapsNet surpassed other models on the Heel Disease Dataset, Breast_BreaKHis_v1, HAM10000, Jun Cheng's Brain Imaging MRI dataset, among others. With its hybrid strategy, MedCapsNet overcame conventional models' weaknesses, which primarily performed very well even without having separate architectures constructed or heavy data augmentation applied for every set. Its application of automated region-of-interest (ROI) detection employing YOLO V 8 also reduced the amount of preprocessing needed and increased the accuracy of results.

Both models demonstrated stable potential to surpass the existing methods across different settings. Although HeCapsNet was particularly good at heel disease and Brain MRI classification, the fact that MedCapsNet could work with diverse medical datasets speaks volumes for its versatility and strength. These results highlight the need

for new architectural designs and suitable preprocessing to enhance medical image classification issues.

6.2. ADVANTAGES AND INNOVATIONS OF PROPOSED MODELS

The main advantages of HeCapsNet and MedCapsNet are found in their advanced structural-layer characteristics that provide solutions to the problems associated with the deep convolution neural networks and the previous variants of CapsNet. HeCapsNet aims to develop a more advanced CapsNet that modifies feature learning by employing more extra convolutional layers and deploying the kernel initializer "he_normal", increasing gradient flow during training. This invention improves not only the accuracy of the model but also the extent to which the model can generalize on diverse data sets.

The MedCapsNet model has several advances and innovations. It combines a CapsNet with a densenet201 architecture, thus improving data feature extraction capability and representation of spatial hierarchies, even without data augmenting. This two-component type of improvement delivers better performance with high accuracy and AUC values over several medical datasets like X-ray, MRI, and dermatoscopic images. The model scope of application is relatively broad due to its fixed architecture and hyperparameters, which give reproducible results for different types of images and diseases. The MedCapsNet model provides faster training, good generalization, and lower overfitting risks, making it a practical and useful model for analyzing medical images.

These models constitute a considerable advancement in the use of deep learning in medical imaging as they provide a platform that could be modified for different diagnostic tasks. They are more useful in circumstances where specialized medical personnel are hard to come by because of their potential to enhance the accuracy of diagnosis and, at the same time, being economical.

6.3. CHALLENGES AND LIMITATIONS

Despite the strengths of HeCapsNet and MedCapsNet, many issues and limitations still exist. One of the issues is the lack of large sets of labeled data for model training, which is a must-have. While HeCapsNet presented a new dataset of 3956 labeled lateral foot X-ray images, the low availability of such datasets for other types of medical conditions reduces the scope of usage of these models. Such dependency on massive data collection efforts will likely discourage wide use in different clinical environments.

Additionally, the complexity of the models raises issues regarding interpretability. Since HeCapsNet and MedCapsNet are deep learning models, they can be considered black boxes, challenging clinicians to understand the rationale behind the diagnosis. Such a degree of opacity may hinder their uptake in centralized clinical settings where a high level of faith is placed on devices.

Moreover, these systems rely on effective region of interest (ROI) identification systems (for example, YOLO), which means any error in the identification ROI will decrease the classification accuracy.

Additionally, the computational resources required for training and inference, while optimized, can still be significant, limiting deployment in resource-constrained environments.

Also, the optimization of the used infrastructure has not resolved the problem of the high computational power necessary for the training and inference, which limits the deployment of the solution in low-resource environments.

It's worth noting that incorporating ethical and practical elements such as extensive testing, legislative approvals, and embedding into the current clinical workflows makes it even more challenging to implement in practice.

Lastly, the non-interpretability feature of some deep learning models may also make clinicians resistant to using them when making crucial diagnosis decisions. It also shows potential areas for further development and better collaboration with clinicians.

To conclude, although HeCapsNet and MedCapsNet show potential in the analysis of medical images, the broader application is hampered by other factors such as the small size of the datasets, the interpretability of the models, high computation requirements, and the need for precise ROI identification. Furthermore, factors such as ethical compliance, regulatory approval procedures, and the incorporation of the technology into clinical practice present additional challenges to its implementation, which would necessitate further research and coordination with medical practitioners to eliminate these constraints.

PART 7

CONCLUSION

In this study, we proposed two advanced CapsNet structures, HeCapsNet and MedCapsNet, to detect and classify heel diseases automatically using lateral foot X-ray images. In terms of these contributions, both models completely overcame the insufficiencies of the original CapsNet, which would have posed significant challenges in medical imaging applications, by incorporating some sophisticated modifications.

HeCapsNet was developed to enhance feature map extraction and classification accuracy by incorporating a few strategies like additional convolutional layers, he_normal kernel initializer, and other advanced regularization methods. This claim was validated through extensive experiments on a number of datasets such as MNIST, F-MNIST, CIFAR-10, Brain MRI, and our new heel disease dataset. On the heel disease dataset, HeCapsNet delivered impressive results of 97.29% and 94.19% accuracy on the balanced and imbalanced data, respectively, with an AUC of 98.69% and a 5-fold cross-validation accuracy of 95.77%. However, certain limitations restrict the applicability of HeCapsNet as it targets only heel spur and severe diseases; the generalizability is limited due to the relatively small sample size of the dataset. A larger dataset and more inclusion of other heel conditions could address this limitation.

MedCapsNet, an enhanced CapsNet with modified DenseNet201, further improved the detection and classification capabilities of the methodology proposed. MedCapsNet compensated for the limits of the original CapsNet by improving feature extraction and implementing DenseNet201, thus reducing the total amount of data augmentations needed. The model achieved remarkable results on various medical databases, including Breast_BreKHis_v1, the HAM10000 skin cancer dataset, and Jun Cheng's Brain Imaging MRI Dataset. It also automatically detected Regions of Interest (ROI) in foot X-ray images with the help of the built-in YOLO V8 and,

therefore, simplified the data preprocessing. MedCapsNet can achieve high performance without requiring distinct architectural configurations or hyperparameter adjustments for each dataset, underscoring its robustness and adaptability.

Both HeCapsNet and MedCapsNet are promising ways of improving the domain of medical image analysis. HeCapsNet provides a workable method for diagnosing heel diseases, while MedCapsNet shows effectiveness on different medical datasets. These models can be incorporated into X-ray machines, served as web-based diagnostic centers, or used in clinical settings. The proposed models can offer a reasonable and rapid preliminary diagnosis, which could benefit the patients, lessen the workload of the diagnosticians, and make the healthcare delivery systems more effective. Subsequent engineering actions will target increasing the datasets, adding more medical diseases, and improving the models to clinical utility.

REFERENCES

1. Kaul, Vivek, Sarah Enslin, and Seth A. Gross. "History of artificial intelligence in medicine." *Gastrointestinal endoscopy* 92.4: 807-812 (2020).
2. Zhu, Zhi, et al. "Knowledge-based and data-driven behavioral modeling techniques in engagement simulation." *Simulation* 99.10: 1069-1089 (2023).
3. Briganti, Giovanni, and Olivier Le Moine. "Artificial intelligence in medicine: today and tomorrow." *Frontiers in medicine* 7: 509744 (2020).
4. Mich, Luisa. "Artificial intelligence and machine learning." *Handbook of e-Tourism*: 1-21 (2020).
5. Park, Chan-Woo, et al. "Artificial intelligence in health care: current applications and issues." *Journal of Korean medical science* 35.42 (2020).
6. Pinto-Coelho, Luís. "How artificial intelligence is shaping medical imaging technology: A survey of innovations and applications." *Bioengineering* 10.12: 1435 (2023).
7. Alhasan, Mustafa, and Mohamed Hasaneen. "Digital imaging, technologies and artificial intelligence applications during COVID-19 pandemic." *Computerized Medical Imaging and Graphics* 91: 101933 (2021).
8. Olveres, Jimena, et al. "What is new in computer vision and artificial intelligence in medical image analysis applications." *Quantitative imaging in medicine and surgery* 11.8: 3830 (2021).
9. Swarnkar, Suman Kumar, et al. "Deep learning techniques for medical image segmentation & classification." *International journal of health sciences* 6.S10: 408-421(2022).
10. Taye, Mohammad Mustafa. "Understanding of machine learning with deep learning: architectures, workflow, applications and future directions." *Computers* 12.5: 91 (2023).
11. Abdou, Mohamed A. "Literature review: Efficient deep neural networks techniques for medical image analysis." *Neural Computing and Applications* 34.8: 5791-5812 (2022).
12. Sabour, Sara, Nicholas Frosst, and Geoffrey E. Hinton. "Dynamic routing between capsules." *Advances in neural information processing systems* 30 (2017).

13. Taher, Osamah, and Kasım Özacar. "HeCapsNet: An enhanced capsule network for automated heel disease diagnosis using lateral foot X-Ray images." *International Journal of Imaging Systems and Technology* 34.3: e23084 (2024)
14. Taher, Osamah, and Kasım Özacar. "MedCapsNet: A modified Densenet201 model integrated with capsule network for heel disease detection and classification." *Heliyon* 10.14 (2024).
15. Altuntas, Emrah, and Ahmet Uzun. "The Relationship Between Age-Related Incidences of Heel Spur With Sex and Side." *International Journal of Morphology* 40.2 (2022).
16. Hosny, G. A., et al. "Current concept review of Sever's disease in paediatric age group." *Benha Journal of Applied Sciences* 6.4: 297-303 (2021).
17. Li, Xin, et al. "Role of Artificial Intelligence in Medical Image Analysis: A Review of Current Trends and Future Directions." *Journal of Medical and Biological Engineering* (2024): 1-13.
18. Bangert, Patrick, et al. "Active learning performance in labeling radiology images is 90% effective." *Frontiers in radiology* 1 :748968 (2021).
19. Kumar, Yogesh, et al. "Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda." *Journal of ambient intelligence and humanized computing* 14.7: 8459-8486 (2023).
20. Rasool, Saad, et al. "Harnessing predictive power: exploring the crucial role of machine learning in early disease detection." *JURIHUM: Jurnal Inovasi dan Humaniora* 1.2: 302-315 (2023).
21. Arefin, Sabira. "Leveraging AI for Healthcare Advancement in Africa." *Academic Journal of Science and Technology* 7.1: 1-11 (2024).
22. Sharifani, Koosha, and Mahyar Amini. "Machine learning and deep learning: A review of methods and applications." *World Information Technology and Engineering Journal* 10.07: 3897-3904 (2023).
23. Radak, Mehran, Haider Yabr Lafta, and Hossein Fallahi. "Machine learning and deep learning techniques for breast cancer diagnosis and classification: a comprehensive review of medical imaging studies." *Journal of Cancer Research and Clinical Oncology* 149.12: 10473-10491 (2023).
24. Sarker, Iqbal H. "Machine learning: Algorithms, real-world applications and research directions." *SN computer science* 2.3: 160 (2021).
25. Mirzaei, Golrokh, and Hojjat Adeli. "Machine learning techniques for diagnosis of alzheimer disease, mild cognitive disorder, and other types of dementia." *Biomedical Signal Processing and Control* 72: 103293 (2022).

26. Mahesh, Batta. "Machine learning algorithms-a review." *International Journal of Science and Research (IJSR)*. [Internet] 9.1 : 381-386 (2020).
27. Sarker, Iqbal H. "Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions." *SN computer science* 2.6: 420 (2021).
28. Tsuneki, Masayuki. "Deep learning models in medical image analysis." *Journal of Oral Biosciences* 64.3 : 312-320 (2022).
29. Guo, Yujuan, Jingjuan Liao, and Guozhuang Shen. "A deep learning model with capsules embedded for high-resolution image classification." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14: 214-223 (2020).
30. Guenther, Frank H. "Neural networks: Biological models and applications." Oxford: *International Encyclopedia of the Social & Behavioral Sciences*: 10534-7 (2001).
31. Madsen, Birger. "Neural Networks. H. Abdi, D. Valentin and B. Edelman, Sage University Paper Series on Quantitative Applications in the Social Sciences, Series no. 07-124, 1999. No. of pages: vi+ 90. ISBN 0-7619-1440-4." : 2199-2200 (2000).
32. Mehrotra, Kishan, Chilukuri K. Mohan, and Sanjay Ranka. Elements of artificial neural networks. MIT press, 1997.
33. Karsoliya, Saurabh. "Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture." *International journal of engineering trends and technology* 3.6: 714-717 (2012).
34. Uzair, Muhammad, and Noreen Jamil. "Effects of hidden layers on the efficiency of neural networks." *2020 IEEE 23rd international multitopic conference (INMIC)*. IEEE, 2020.
35. Agatonovic-Kustrin, S., and Rosemary Beresford. "Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research." *Journal of pharmaceutical and biomedical analysis* 22.5: 717-727 (2000).
36. "Necessity of Bias in Neural Networks." *Turing.com*, Turing, Accessed Date [23/10/2024], <https://www.turing.com/kb/necessity-of-bias-in-neural-networks>.
37. **Deeplizard**. "Neural Network Bias." *YouTube*, 21 Mar. 2018, Accessed Date [23/10/2024] <https://www.deeplizard.com/learn/video/HetFihsXSys>.
38. Nwankpa, Chigozie, et al. "Activation functions: Comparison of trends in practice and research for deep learning." *arXiv preprint* arXiv:1811.03378 (2018).

39. Ramachandran, Prajit, Barret Zoph, and Quoc V. Le. "Searching for activation functions." *arXiv preprint* arXiv:1710.05941 (2017).
40. Wang, Qi, et al. "A comprehensive survey of loss functions in machine learning." *Annals of Data Science*: 1-26 (2020).
41. Alzubaidi, Laith, et al. "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions." *Journal of big Data* 8: 1-74. (2021)
42. Anil, Anil Kumar Prajapati, and Umesh Kumar Singh. "An Optimal Solution to the Overfitting and Underfitting Problem of Healthcare Machine Learning Models." *Journal of Systems Engineering and Information Technology (JOSEIT)* 2.2: 77-84 (2023).
43. Zhang, Haotian, Lin Zhang, and Yuan Jiang. "Overfitting and underfitting analysis for deep learning based end-to-end communication systems." 2019 11th *international conference on wireless communications and signal processing (WCSP)*. IEEE, (2019).
44. Du, Xuedan, et al. "Overview of deep learning." 2016 31st *Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. IEEE, 2016.
45. Liang, Hong, et al. "Text feature extraction based on deep learning: a review." *EURASIP journal on wireless communications and networking* 2017: 1-12 (2017).
46. Xu, Yan, et al. "Deep learning of feature representation with multiple instance learning for medical image analysis." 2014 *IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, (2014).
47. LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521.7553: 436-444 (2015).
48. Rao, T. Venkat Narayana, et al. "Reliance on artificial intelligence, machine learning and deep learning in the era of industry 4.0." *Smart healthcare system design: security and privacy aspects*: 281-299 (2022).
49. Yamashita, Rikiya, et al. "Convolutional neural networks: an overview and application in radiology." *Insights into imaging* 9: 611-629 (2018).
50. Dhillon, Anamika, and Gyanendra K. Verma. "Convolutional neural network: a review of models, methodologies and applications to object detection." *Progress in Artificial Intelligence* 9.2: 85-112 (2020).
51. Krichen, Moez. "Convolutional neural networks: A survey." *Computers* 12.8: 151 (2023).

52. Li, Zewen, et al. "A survey of convolutional neural networks: analysis, applications, and prospects." *IEEE transactions on neural networks and learning systems* 33.12: 6999-7019 (2021).
53. Cong, Shuang, and Yang Zhou. "A review of convolutional neural network architectures and their optimizations." *Artificial Intelligence Review* 56.3: 1905-1969 (2023).
54. O'Shea, K. "An introduction to convolutional neural networks." *arXiv preprint arXiv:1511.08458* (2015).
55. Saxena, Aarush. "An introduction to convolutional neural networks." *Int. J. Res. Appl. Sci. Eng. Technol* 10.12: 943-947 (2022).
56. Lillo, Giuseppe. Capsule Networks as Alternative to DCNN. Rel. Elio Piccolo, Politecnico di Torino, *Corso di laurea magistrale in Ingegneria Informatica (Computer Engineering)*, 2018.
57. Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting." *The journal of machine learning research* 15.1 : 1929-1958 (2014).
58. Haq, Mahmood Ul, Muhammad Athar Javed Sethi, and Atiq Ur Rehman. "Capsule network with its limitation, modification, and applications—A survey." *Machine Learning and Knowledge Extraction* 5.3: 891-921 (2023).
59. A. Jiménez-Sánchez, S. Albarqouni, D. Mateus, Capsule Networks against Medical Imaging Data Challenges, CVII-STENT/LABELS@MICCAI, 2018.
60. Toraman, Suat, Talha Burak Alakus, and Ibrahim Turkoglu. "Convolutional capsnet: A novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks." *Chaos, Solitons & Fractals* 140 :110122. (2020)
61. Zhang, Zhe, et al. "Enhanced capsule network for medical image classification." *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, (2020).
62. Afriyie Y, Weyori BA, Opoku AA. Comparative evaluation performances of capsule networks for complex image classification. *J Data Inform Manage*. 4(3–4):267-276 (2022).
63. Afshar, Parnian, Konstantinos N. Plataniotis, and Arash Mohammadi. "Capsule networks for brain tumor classification based on MRI images and coarse tumor boundaries." *ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, (2019).

64. Afshar, Parnian, Arash Mohammadi, and Konstantinos N. Plataniotis. "Brain tumor type classification via capsule networks." **2018 25th IEEE international conference on image processing (ICIP)**. IEEE, (2018).
65. Quan, Hao, et al. "DenseCapsNet: Detection of COVID-19 from X-ray images using a capsule neural network." **Computers in biology and medicine** 133: 104399 (2021).
66. Sezer, Aysun, and Hasan Basri Sezer. "Capsule network-based classification of rotator cuff pathologies from MRI." **Computers & Electrical Engineering** 80: 106480 (2019).
67. Koppar, Anant, et al. "Pediatric Bone Age Detection Using Capsule Network." **Inventive Computation and Information Technologies: Proceedings of ICICIT 2020**. Springer Singapore, (2021).
68. Sun, Guangcong, et al. "A novel dense capsule network based on dense capsule layers." **Applied Intelligence** 52.3: 3066-3076 (2022).
69. Ai, Xiongzhi, et al. "ResCaps: an improved capsule network and its application in ultrasonic image classification of thyroid papillary carcinoma." **Complex & Intelligent Systems** 8.3: 1865-1873. (2022)
70. Bushara, A. R., RS Vinod Kumar, and S. S. Kumar. "An ensemble method for the detection and classification of lung cancer using Computed Tomography images utilizing a capsule network with Visual Geometry Group." **Biomedical Signal Processing and Control** 85: 104930 (2023).
71. Qiu, Shuo, Jinlin Ma, and Ziping Ma. "IRCM-Caps: An X-ray image detection method for COVID-19." **The Clinical Respiratory Journal** 17.5: 364-373 (2023).
72. Saravagi, Deepika, et al. "Predicting lumbar spondylolisthesis: a hybrid deep learning approach." **Intelligent Automation & Soft Computing** 37.2 (2023): 2133-2151(2023).
73. Imran, Azhar, et al. "Skin cancer detection using combined decision of deep learners." **IEEE Access** 10: 118198-118212 (2022).
74. Wang, Hao, Kun Shao, and Xing Huo. "An improved CapsNet applied to recognition of 3D vertebral images." **Applied Intelligence** 50.10: 3276-3290 (2020).
75. Gour, Mahesh, Sweta Jain, and Sushant Kaushal. "XCapsNet: a deep neural network for automated detection of diabetic retinopathy." **International Journal of Imaging Systems and Technology** 33.3: 1014-1027 (2023).
76. Lan, Zhangli, et al. "FixCaps: An improved capsules network for diagnosis of skin cancer." **IEEE Access** 10: 76261-76267 (2022).

77. Afshar, Parnian, Konstantinos N. Plataniotis, and Arash Mohammadi. "BoostCaps: a boosted capsule network for brain tumor classification." **2020 42nd annual international conference of the IEEE engineering in medicine & biology society (EMBC)**. IEEE, (2020).
78. Khikani, Hayder A., et al. "Breast cancer classification based on histopathological images using a deep learning capsule network." **arXiv preprint arXiv:2208.00594** (2022).
79. Kumari, Pammi, and Priyank Saxena. "Automated diabetic retinopathy grading based on the modified capsule network architecture." **IETE Journal of Research** 70.3: 2917-2928 (2024).
80. Sengul, Sumeyra Busra, and Ilker Ali Ozkan. "HMedCaps: a new hybrid capsule network architecture for complex medical images." **Neural Computing and Applications**: 1-18 (2024).
81. Kalyani, G., Janakiramaiah, B., Karuna, A. et al. Diabetic retinopathy detection and classification using capsule networks. **Complex Intell. Syst.** 9, 2651–2664 (2023).
82. Oulhadj, Mohammed, et al. "Diabetic retinopathy prediction based on vision transformer and modified capsule network." **Computers in Biology and Medicine** 175: 108523 (2024).
83. Kumar, S.A., Muchahari, M.K., Poonkuntran, S. et al. Application of hybrid capsule network model for malaria parasite detection on microscopic blood smear images. **Multimed Tools Appl** (2024)
84. Akoto-Adjepong, Vivian, et al. "TTDCapsNet: Tri Texton-Dense Capsule Network for complex and medical image recognition." *Plos one* 19.3: e0300133 (2024).
85. Veni, Paluri Krishna, and Ashish Gupta. "Revolutionizing Acne Diagnosis with Hybrid Deep Learning Model Integrating CBAM, and Capsule Network." **IEEE Access** (2024).
86. Saif, A. F. M., et al. "CapsCovNet: A modified capsule network to diagnose Covid-19 from multimodal medical imaging." **IEEE Transactions on Artificial Intelligence** 2.6: 608-617 (2021).
87. Koresh, H. James Deva, Shanty Chacko, and M. Periyannayagi. "A modified capsule network algorithm for oct corneal image segmentation." **Pattern Recognition Letters** 143: 104-112 (2021).
88. Gupta, P. K., et al. "COVID-WideNet—A capsule network for COVID-19 detection." **Applied Soft Computing** 122: 108780 (2022).

89. Tiwari, Shamik, and Anurag Jain. "A lightweight capsule network architecture for detection of COVID-19 from lung CT scans." *International Journal of Imaging Systems and Technology* 32.2: 419-434 (2022).
90. Madhu, G., et al. "Imperative dynamic routing between capsules network for malaria classification." *CMC-Computers Materials & Continua* 68.1: 903-919 (2021).
91. Gocer, Evgin. "Classification of skin cancer using adjustable and fully convolutional capsule layers." *Biomedical Signal Processing and Control* 85: 104949 (2023).
92. Behara, Kavita, Ernest Bhero, and John Terhile Agee. "An Improved Skin Lesion Classification Using a Hybrid Approach with Active Contour Snake Model and Lightweight Attention-Guided Capsule Networks." *Diagnostics* 14.6 : 636 (2024).
93. Sadeghnezhad, Elham, and Sajjad Salem. "InceptionCapsule: Inception-Resnet and CapsuleNet with self-attention for medical image Classification." *arXiv preprint arXiv:2402.02274* (2024).
94. Aydin Atasoy, Nesrin, and Amina Faris Abdulla Al Rahhawi. "Examining the classification performance of pre-trained capsule networks on imbalanced bone marrow cell dataset." *International Journal of Imaging Systems and Technology* 34.3: e23067 (2024).
95. POKUAA, Henrietta Adjei, et al. "DPafy-GCaps: Denoising patch-and-amplify Gabor capsule network for the recognition of gastrointestinal diseases." *Turkish Journal of Electrical Engineering and Computer Sciences* 32.3: 452-464 (2024).
96. Remya, R., and K. Parimala Geetha. "Health-Care Monitoring for the Brain Tumor Disorder Patients' by Estimating its Thickness by an Enhanced Capsule Network." *Wireless Personal Communications* 130.3: 1743-1757 (2023).
97. Ye, Zhiwei, et al. "Hyperparameter Optimization for Capsule Network Based Modified Hybrid Rice Optimization Algorithm." *Intelligent Automation & Soft Computing* 37.2 (2023).
98. Chaudhari, Jitendra P., et al. "Automated bacteria genera classification using histogram-oriented optimized capsule network." *Engineering Science and Technology, an International Journal* 46: 101500 (2023).
99. Mursalim, Muhammad KN, et al. "BiCaps-DBP: Predicting DNA-binding proteins from protein sequences using Bi-LSTM and a 1D-capsule network." *Computers in Biology and Medicine* 163: 107241 (2023).
100. Chen, Hua, et al. "MSCCNet: Multi-Scale Convolution-Capsule Network for Cervical Cell Classification." 2023 IEEE *International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, (2023).

101. Adu, Kwabena, et al. "SqueezeCapsNet: enhancing capsule networks with squeezenet for holistic medical and complex images." *Multimedia Tools and Applications* 83.1: 2823-2852 (2024).
102. Pallapu, Srinivasa Rao, and Khasim Syed. "ACNGCNN: Improving Efficiency of Breast Cancer Detection and Progression using Adversarial Capsule Network with Graph Convolutional Neural Networks." *International Journal of Advanced Computer Science & Applications* 15.5 (2024).
103. Dubey, Vineet Kumar, and Vandana Dixit Kaushik. "Epidermis lesion detection via optimized distributed capsule neural network." *Computers in Biology and Medicine* 168: 107833 (2024).
104. Boruah, Minakshi, and Ranjita Das. "CaDenseNet: a novel deep learning approach using capsule network with attention for the identification of HIV-1 integration site." *Neural Computing and Applications* 35.23: 17113-17128 (2023).
105. Manikandan, J., and R. Yogeshkumar. "Glaucoma Disease Detection Using Hybrid Deep Learning Model." 2023 *Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*. IEEE, (2023).
106. Zhang, Zequn, et al. "CapsNet-LDA: predicting lncRNA-disease associations using attention mechanism and capsule network based on multi-view data." *Briefings in Bioinformatics* 24.1: bbac531 (2023).
107. Ravindran, U., and C. Gunavathi. "Cancer Disease Prediction Using Integrated Smart Data Augmentation and Capsule Neural Network." *IEEE Access* (2024).
108. Zhang, Hui, et al. "Cn2a-capsnet: a capsule network and CNN-attention based method for COVID-19 chest X-ray image diagnosis." *Discover Applied Sciences* 6.4: 190 (2024).
109. Boruah, Minakshi, and Ranjita Das. "MLCapsNet+: A multi-capsule network for the identification of the HIV ISs along important sequence positions." *Image and Vision Computing* 145: 104990 (2024).
110. Han, Hyunkyung, Jihyeon Seong, and Jaesik Choi. "CardioCaps: Attention-Based Capsule Network for Class-Imbalanced Echocardiogram Classification." 2024 *IEEE International Conference on Big Data and Smart Computing (BigComp)*. IEEE, (2024).
111. Pant, Himanshu, Manoj Chandra Lohani, and Ashutosh Kumar Bhatt. "X-rays imaging analysis for early diagnosis of the thoracic disorders using Capsule Neural Network with Transfer learning." 2023 *9th International Conference on Advanced Computing and Communication Systems (ICACCS)*. Vol. 1. IEEE, (2023).

112. Almarshad, Fahdah A., et al. "An Efficient Optimal CapsNet Model-based Computer-Aided Diagnosis for Gastrointestinal Cancer Classification." *IEEE Access* (2024).
113. Lei, Yongjia, et al. "BP-CapsNet: An image-based Deep Learning method for medical diagnosis." *Applied Soft Computing* 146: 110683 (2023).
114. Shaheen, H., et al. "An efficient classification of cirrhosis liver disease using hybrid convolutional neural network-capsule network." *Biomedical Signal Processing and Control* 80: 104152 (2023).
115. Manohar, Gyanendar, and Ruairi O'Reilly. "InceptionCaps: A Performant Glaucoma Classification Model for Data-Scarce Environment." 2023 *31st Irish Conference on Artificial Intelligence and Cognitive Science (AICS)*. IEEE, (2023).
116. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. Proceedings of the IEEE International Conference on Computer Vision. (2015).
117. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016).
118. Huang, Gao, et al. "Densely connected convolutional networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
119. Kumar, Amish, et al. "CSNet: A new DeepNet framework for ischemic stroke lesion segmentation." *Computer methods and programs in biomedicine* 193: 105524 (2020).
120. Spanhol, Fabio A., et al. "A dataset for breast cancer histopathological image classification." *Ieee transactions on biomedical engineering* 63.7 (2015): 1455-1462 (2015).
121. Tschandl, Philipp, Cliff Rosendahl, and Harald Kittler. "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions." *Scientific data* 5.1 (2018): 1-9 (2018).
122. Cheng, Jun, et al. "Retrieval of brain tumors by adaptive spatial pooling and fisher vector representation." *PloS one* 11.6: e0157112 (2016).

RESUME

Osamah Fadhil Taher TAHER began his primary and elementary education in Diyala, Iraq. He completed his high school studies at Al-Saadiyah Secondary School in Diyala. In 2008, he earned his bachelor's degree in Computer Science from the College of Education for Pure Sciences at the University of Diyala. In 2009, he moved to Ankara, Türkiye, to pursue a Master of Science (M.Sc.) degree. He enrolled in the Department of Electronic and Computer Education at Gazi University, Institute of Informatics. In 2020, he relocated to Karabük, Türkiye, to further his academic journey and pursue a Doctor of Philosophy (Ph.D.) degree. He joined the Department of Computer Engineering at Karabük University.