An Efficient Deep Learning Approach to Detect Bone Fractures from X-ray Images

by

MD Fahmidur Rahman
24241275
MD Ashaduzaman Ridoy
20301021
SK Fahema Mahajabin
20341013
Sadman Rahman Tasawar
21201680

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
June 2025

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

MD Ashaduzaman Ridoy
20301021

MD Fahmidur Rahman
24241275

SK Fahema Mahajabin
20341013

Sadman Rahman Tasawar
21201680

Approval

The thesis/project titled An Efficient Deep Learning Approach to Detect Bone Fractures from X-ray Images” was submitted by

1. MD Ashaduzaman Ridoy (20301021)
2. MD Fahmidur Rahman (24241275)
3. SK Fahema Mahajabin (20341013)
4. Sadman Rahman Tasawar (21201680)

Of Summer, 2024 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on Oct 19, 2024.

Examining Committee:

Supervisor:
(Member)

Md. Ashraful Alam, PhD
Associate Professor
Department of Computer Science and Engineering
BRAC University

Program Coordinator:
(Member)

Md. Golam Rabiul Alam, Phd
Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)

Sadia Hamid Kazi, Phd
Chairperson and Associate Professor
Department of Computer Science and Engineering
BRAC University

Abstract

We recommend an effective,reliable and efficient deep-learning approach to classify bone fractures using X-ray images. The proposed system comprises several steps as dataset collection, preprocessing, and categorization of fractures. The dataset contains different types of X-ray images for various fractures and annotated as fractured and non-fractured cases. Dataset was collected from Kaggle. The images were processed through resizing,augmentation to be standardized for deep learning models. The classification of fractures utilizes neural network paged deep learning models as the last process. The research uses Inception,VGG-19 and a custom CNN model for fracture detection while using Grad-CAM to demonstrate how each model makes its decisions. These models utilize CNN architectures to overcome current models accuracy limitations when detecting fractures. The study also shows how CNN-based approaches elevate the precision of bone fracture detection which creates potential benefits for medical use and improved patient healthcare.

Keywords: Bone fractures, X-ray images, Medical diagnostics, Radiologists, Deep learning, Automation, Convolutional Neural Networks, Optimization Techniques, Image preprocessing, Accuracy, Explainable AI.

Dedication

We truly appreciate the aid and support provided to us in accomplishing our thesis by the individuals and organizations named below.

First and foremost,we want to express our gratitude to our thesis supervisor Md. Ashraful Alam,PhD for his unflagging sup- port,astute criticism and unending patience throughout this research trip.Your knowledge and guidance have been very helpful in determining the course of this study.

We would like to show our appreciation to the teachers and staff at Brac University for providing the tools,spaces and academic setting that made my research activities possible.

Without the collaboration and assistance of these people and organizations,this thesis would not have been possible.We appreciate your participation in this scholarly project.

MD Fahmidur Rahman

MD Ashaduzaman Ridoy

SK Fahema Mahajabin

Sadman Rahman Tasawar

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

AI Artificial Intelligence

ANN Artificial Neural Network

AUC Area Under the Curve (used to evaluate model performance)

CNN Convolutional Neural Network

DenseNet Densely Connected Convolutional Network

DL Deep Learning

DNN Deep Neural Network

FN False Negative

FP False Positive

GPU Graphics Processing Unit

Grad-CAM Gradient-weighted Class Activation Mapping

ML Machine Learning

ReLU Rectified Linear Unit (activation function)

ResNet Residual Network

ROC Receiver Operating Characteristic (curve used to evaluate classifiers)

ROC AUC Receiver Operating Characteristic Area Under the Curve

TN True Negative

TP True Positive

X-ray X-radiation used in medical imaging

XAI Explainable Artificial Intelligence

YOLO You Only Look Once

Chapter 1

Introduction

1.1 Background

The human body depends on bones being one of the most important structural and functional factors. It plays the part of the body structure framework and undertakes important biological functions. Bones can only take certain amounts of force before breaking, and they're strong and dense. Fractures represent one of the most common global injuries due to these many issues. These fractures are caused by many reasons such as concussion, stress, or a medical condition that affects the strength in the bone.Acute fractures occur at an instant and happens due to a fall, a twist, or a sports injury stress fracture develops over time as a result of constant stress put on a bone, commonly in athletes. The second group is those that take place as a result of other bone diseases such as osteoporosis, cancer or infection which affects the bone's strength and makes it susceptible to breaking. Furthermore, timely accurate identification of bone fractures is crucial to avoid such problems as poorly healed, chronic pain or carrying lifelong disability. Fracture imaging used to diagnose fractures still relies on imaging in the form of X-rays. While using traditional methods of detecting fractures in X-ray images can be accomplished by radiologists, several issues arise from this methodology including the complexity of the task, the inability to detect specific types of fractures and the load placed on radiologists in large hospitals (Tanzi, Bonetto, & Zavanone, 2020). The pressure of an emergency room makes manual interpretation of X-ray images laborious and computationally expensive. Hairline, comminuted, and compound fractures add another level of complexity to the diagnostic process: Even trained radiologists may need help to tell the difference between minor or partially overlapping fractures.

1.2 Problem Statement

Although the goal of automating medical imaging using deep learning has been effective, important challenges still exist. The first problem that is also the largest is that the algorithms require fully labeled data to train. The current supervised learning models can be successful when measured data is large and there are a lot of labeled examples (Cheng, Smith, Lee, Park, 2024). In medical settings though, the process of tagging X-ray images is time-consuming, resource intensive, and needs specialized expertise, which poses difficulties to acquire large high-quality datasets. This limitation in data availability brings the problem of model generalization; most of the time such models do not perform well in other patient populations or with different imaging techniques or X-rays from different sources thus compromising their clinical reliability. (Guan, Zhang, Li, 2018). X-ray imaging is still considered as one of the simplest diagnostic methods although its sensitivity and accuracy depend on factors such as radiologist efforts, experience, as well as on the subjectivity of the images obtained. Such microscopic cracks as hairline cracks are not easily discovered and can be diagnosed late because of the serious complications it can cause due to the bleeding it causes internally (Adigun et al., 2020). Moreover, there is a demand for rising imaging for diagnosing healthcare problems which puts a lot of pressure on radiologists and lead to a delay in diagnosis and human errors (Choi and al in 2020). One more problem is that the quality of the X-ray image varies during the AE process. This raises some inconsistencies due to differences in resolution of equipment, positioning of the patient, light and contrast. This variability is somewhat problematic for all models based on these data sets, and hence results in low effectiveness whenever it is employed on new or other data types (Sharma, 2023). So we need to create a model system that can function optimally irrespective of the setting and across distinct hospitals with different equipment. Previous approaches to machine learning fracture detection required modifications to available data and could not perform well during testing and were weak on generalizing to new data(Cheng et al., 2024). The existing CNN-based algorithms, albeit with a high level of accuracy regarding the issue of image classification, fail to take into consideration the higher levels of complexity of real-life medical imagery. The lesions can be quite indistinct and subtle or easily obscured by the overlying anatomy structures in certain cases. Most of the CNNs extract common characteristics of the images, but do not have techniques to selectively attend to areas of interest such as the region where the fracture is most likely to occur in clinical settings. This general observation reduces their chances of identifying finer cracks specific to a region and which often takes a more regionalized approach, including the ones that are thinner than a piece of hair or the stress fractures which might be missed at the first instance by overemphasizing on a general view. Besides, the instability of the X-ray image quality complicates the study. Change in instrumentation such as imaging equipment, the orientation of the patient, intensity of light, and resolution causes changes that cause variability in the model which makes it less accurate. Such differences can mean missed diagnoses which can occur in cases where the fracture is minor or otherwise obscure. CNNs which might be appropriate in extracting features are not always ready to compute these discrepancies in real life. . Firstly, high accuracy is as important as high interpretability in order to be used clinically. Clinicians need models that not only are accurate but also explain the reasoning behind them, so the radiologists can rely on it. This interpretability is considered as helpful for increasing diagnostic certainty and for matching them with the clinician's expectations on what the model should do. These drawbacks serve as a basis for this research, generating and training an CNN model that would recognize fractured & non fractured bones with high accuracy. It is opportune and crucial to develop a complex and functional deep-learning model to implement this goal. The suggested model will incorporate methods of enhancing regional focus and concentration of important regions in the X-ray where most fractures may be suspected. Furthermore, the model will implement explainable AI features, which are Grad-CAM in this case, to help the radiologists to understand the model's decision-making. This model must be capable of successfully working on different data and different clinical settings to maintain a good efficiency regardless of the choice of the hospital or the type of imaging equipment. Accomplishing these challenges will greatly improve the recognition of automated fracture detection giving positive impact to patients' experience and improving the workload of health care professionals.

1.3 Research Objectives

The proposed research is based on creating the optimum deep learning model that will enhance the performance, the quality, and meaning of bone fractures identified through X-ray. In this respect, the following research objectives should be pursued:

The proposed study will produce the efficient deep learning model to classify bone fractures in a binary way i.e. as fractured and non-fractured bones on X-ray images. The paper is aimed to develop a hybrid architecture that integrates the Convolutional Neural Network (CNN) with transformer-attentive focus for increasing the accuracy in fractures' detection. Through the use of multi-scale feature extraction methods, the model will be useful in discerning all types of fracture detection, whether large or small.

Semi-supervised learning is also introduced in this study to improve generalization of the model on various imaging scenarios as well as datasets.Also to improve standardization of the X-ray inputs, techniques like denoising, pixel-level transformation (PLT) and image enhancement will be adopted in this study to enhance accuracy and reliability of results. Besides, Grad-CAM++ interpretability techniques will be applied and used for developing the clinical trust through supporting the radiologists to analyze and verify the model predictions and datasets.Comparison will be made between Inception,ResNet, DenseNet,EfficientNet and some other architectures in order to determine which is the most accurate in binary fracture detection. The ultimate goal is to build a large-scale near real-time solution that can be easily integrated into the healthcare domain thereby decreasing radiologists' burden and enhancing the assessment precision. Through these objectives, this study is aimed at developing a high performance, reliable deep learning model that would use the automated bone fracture detection to surpass previous similar models both in accuracy and clinical practice relevance.

1.4 Thesis Outline

- In chapter 2 , the paper discusses literature on the detection of bone fractures through deep learning models. It discusses the basics of Convolutional Neural Networks (CNNs) and the literature of the past works on fracture-detection, and the key problems of the research domain. Architectures of models like ResNet50, EfficientNet, InceptionV3 and Hybrid models are discussed as well in the chapter.

- In chapter 3, the data collection process is described, including those available on Kaggle, hospitals, and medical clinics. It lists the process involved in data preprocessing including resizing, augmentation, and normalization of X-ray images. Also, in this section, it is reported how data was divided into training, validation, and test parts.

- Chapter 4 thoroughly describes the process of model implementation on pre-trained CNN architectures (ResNet50, EfficientNetB3, InceptionV3) as well as the evolution of the Hybrid model based on several architectures). It covers the training process, the configuration of the optimizer, loss and optimization strategies of the model.

- Chapter 5, contains the description of the experimental set and performance. It has the performance statistics of the models on accuracy, precision, recall, F1-score as well as confusion matrices. Besides, GradCAM++ visualisations are given that demonstrate how both of the models provide their predictions, and the Hybrid model results are compared to other models.

- Chapter 6, addresses limitations on the current approach, both constraints of the dataset and the generalization of the model. It also presents suggestions on how it can be extended to future use, such as expanding the dataset, multi class classification and adding clinical metadata. The chapter closes with a conclusion of the contribution of the study and the possible improvement points

Chapter 2

Background

2.1 Background

Deep learning techniques have proven to be quite accurate in determining if there is a broken bone or not, and there have been enormous advances as compared to the traditional manual analysis that has occurred. The technology has significantly transformed the application of radiological diagnostics in most of the anatomical areas including the wrist, hand and vertebrae. This kind of deep learning approach in medical image analysis is essentially founded on CNNs. This skill can be very helpful in detecting small regularities and anomalies in medical images, because complex data: their hierarchical nature enables them to automatically generate the features, which are hard to extract even with the human eye. The competency is not limited to bone fractures, but also to thoracic illnesses and pneumonia. (Sharma, 2023).

Developed models had high accuracy in detecting fractures. For example, the YOLO models are top-of-the-line in quickly detecting wrist fractures on X-ray images. While it is still early, research has found that Residual Networks (ResNet) and DenseNet both seem to work really well at these scales when finding very small, classically difficult-to-see fractures with the human eye. This will enable radiologists to faster and better diagnose, which may have implications for patient outcomes (Cheng et al., 2024). As the hybrid approaches in vertebral fracture detection become advanced, even more researchers started going deeper. The best solution so far has been suggested using CNNs and decision tree classifiers together. This is paired with fine grained classification based on decision trees, where reported performance exceeds that of human experts by an order of magnitude over certain professional clinical environments (Prabath Medaramatla et al. — 2024). It shows that deeper learning models are fit in other locations and types of image ready. More recently, model architecture optimization has been brought to a new level in the field of computer vision with developments like hybrid YOLO NAS (Neural Architecture Search) which automate follow-up work.

Since these models have produced impressive results, they are really meant to be used as tools that complement human skill rather than machines replacing workers. With the integration of deep learning in clinical pathways, it is possible to perform the diagnostics of bone fracture quicker and more consistently and reduce the workload of radiologists.

Although progress has been made in many methodologies of automated bone frac-

ture detection, challenges still exist. As has been known for some time now, one of the many crucial stumbling blocks in this line is the non-availability of large-scale datasets with labeled ground truths.However, this can be solved using data augmentation and transfer learning techniques, at the model level, which is to say, models can be generalize over a larger patient cohort or differ from a given imaging hardware. Ending to End Model interpretability is one such accessible yet largely ignored problem. Models need to predict well, but those predictions won't be considered by clinicians unless the reason behind them is interpretable. For instance, attention mechanisms and explainable AI more than ever will start to be standard practice. In the future, numerous studies will be required to determine the portability and generalizability of deep learning models across diverse patient groups, imaging situations. Other research focused on areas of fracture care—for example, showing the benefit in applying deep learning models to predict whether fractures would go on to heal and thereby ascertaining how treatment plans could be optimized. These applications include not only the initial diagnostic but long-term patient care as well (Johnson et al., 2023).
Deep learning techniques grow in advanced imaging type availability, the potential for MRI and CT in new fracture detection options is expected to increase. This is relevant as multi-modality techniques, although not practised yet, could be used to make better decisions basing on the guidance of available imaging modalities. (Lee et al., 2024). The mentioned improvements could not only result in improving the accuracy of models, but also aid in preserving its interpretability and use in clinical practice. This demonstrates that the use of AI combined with the human expertise of radiology (X-ray), and orthopedic fields have benefits where fracture diagnosis and treatment are both more accurate and patient care remains the same.

2.2 Convolutional Neural Networks

Artificial Intelligence and deep learning have brought a new hope to improve medical diagnosis. Among classes of DNNs , it is proven that Convolutional Neural Networks(CNN) significantly perform well in analyzing medical images like X-rays (Sharma, 2023). The image processing function of Convolutional Neural Networks relies on grid-like data algorithms. For detection convolutional layers apply kernels across an input image to compute product summations which reveal spatial image characteristics including edges and textures. It takes the advantage of being able to pull and learn hierarchical feature directly off of image data. By utilizing pixel arrangement analysis of these networks, particular patterns in the X-Ray images that differentiate fractured bones from normal are learned. With large amounts of data and a wide variety of annotated images, CNNs are capable of optimizing the identification of various kinds of fractures that can help the radiologists diagnose more quickly and constantly. The recent evidence has revealed the employment of CNNs as a potential and primary method of automating the process of detecting any fracture with the help of artificial intelligence and deep learning models. It is quite beneficial in many applications like fracture localization but the CNNs perform well in detecting the spatial pyramids in images. Pretrained CNN architectures like VGG19,Inception,ResNet, DenseNet and EfficientNet have been examined in the context of medical imaging with satisfactory enhancement.

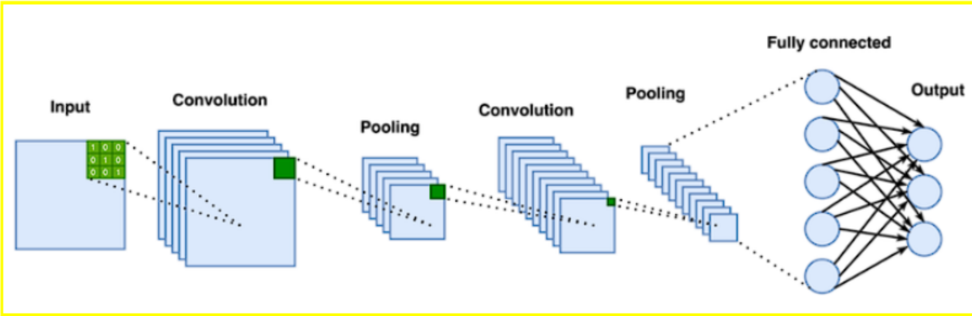

Figure 2.1: Diagram of a basic convolutional neural network

In our study, pre-trained CNN Models: ResNet50, InceptionV3, EfficientNetB3 and a Custom made Hybrid model were used. To extract latent and deep patterns, the models use hierarchical features extracted by the X-ray images. Our given model is made with Conv2D layers that optimize the data to process that specific information. Also with the aim to be more interpretable we also used Grad-CAM++ that outlines the most informative areas of the image. These architectures coupled with feature extraction, pattern recognition and classification can serve as an efficient automaton to identify a bone fracture on medical imaging using the CNNs.

2.3 Current Approaches to Fracture Detection

A variety of AI and ML methods have been studied in the recent to detect bone fracture. Traditional machine learning algorithms were first used in early approaches to CT scan segmentation, but again relied on the manual feature extraction of X-ray images (Adigun et al., 2020). Yet, deep learning models trained by using them have been mostly replaced by methods that models learn features from data automatically. CNNs, and in particular ResNet, DenseNet, VGG,Inception and YOLO – have now become the gold standard in medical image analysis (Tanzi et al. 2020), with ResNet50 in particular successful in training deep networks while fending off the vanishing gradient problem with skip connections. In particular, this architecture is very effective for classifying complex images, where detecting fragile or overlapping fractures on X-ray images (Cheng et al., 2024), for example, is challenging.

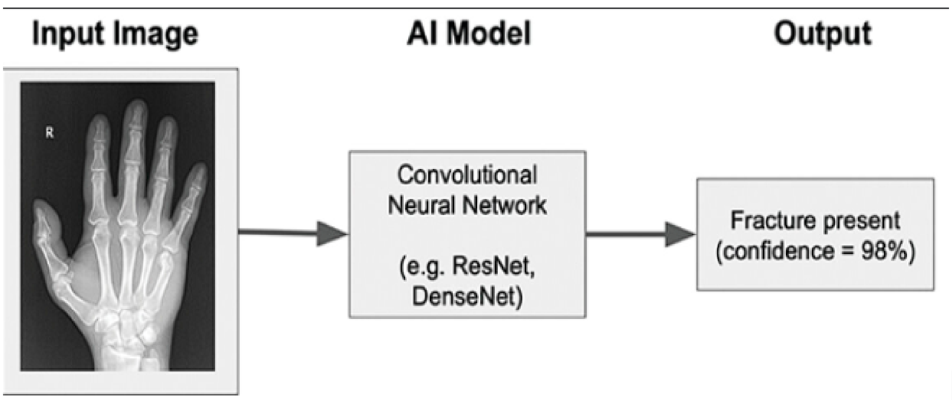

Figure 2.2: Achieving accuracy using various model of CNN

2.3.1 Efficient CNN Architectures for Fracture Detection

- **Inception:** In this architecture we have used a multi scale convolutional so that the network can extract features at different resolutions. This feature is perfect in identifying fracture as mild as hairline fracture to major fracture. This architecture is highly computationally efficient and can achieve the best accuracy since it uses different convolution sizes by incorporating inception modules. The invariance of Inception to the pattern processing capacity renders it extremely useful in automatic identification of the fractures in X-rays.(Szegedy et al., 2016)

- **VGG19:** VGG 19 is a deep CNN architecture using the 3x3 convolutional small filters and a uniform structure. The simplicity of the this allows extraction of high quality spatial features while being computationally efficient.VGG-19 excels in capturing fine details and such subtle patterns in X ray images that ordinary architectures can not capture, which is important for fracture detection.The depth of the model allows for hierarchical feature learning and makes the model ideal for picking out complicated fractures in a varied set of clinical data (Simonyan & Zisserman, 2015).

- **ResNet:** In fracture detection especially in medical image classification.ResNet controls the vanishing gradients problem in deep networks as it does not need residual connections to learn decent features on complex datasets . That ResNet50 is capable of high accuracy both in detection of subtle fractures, including hairline fractures or complicated fractures that even a human expert might not pick up .(Cheng et al., 2024).

- **DenseNet:** DenseNet connects each layer to each other layer in a feed forward manner allowing for better feature reuse. This learning-efficient training offered the possibility of far smaller models that are remarkably good at perceiving medical images, including that of bone fracture. Nonetheless, the DenseNet is especially intended at identifying the fractures in the complicated or turbulent X-ray images (Medaramatla et al., 2024).

- **YOLO (You Only Look Once):** YOLOv4 as well as derivatives like YOLO NAS are real time object detection models. It is also particularly applicable in a clinical setting, where accuracy and speed in locating and classifying fractures on X-rays are important finds, their aptitude being able to efficiently identify and classify fractures in a short duration of time. Real time and high altitude that supports large quantities of medical images, YOLO has been demonstrated to be effective (Adigun et al., 2020).

- **Transfer Learning:** The unavailability of adequate labeled medical data has given rise to transfer learning as one of the significant methods in the construction of effective bone fracture detection models. AI assisted orthopedic diagnosis has been boosted by recasting the problem with this approach, which has significantly improved the accuracy and efficiency of AI assisted orthopedic diagnosis.

According to Wang et al. (2023), fine-tuning these models using the X-ray images results in higher accuracy rates in these models compared with a cleanly trained model, even with available limited data. The domain adaptation methods have been devised to mediate the disparity between medical X-rays and natural images. With adequately constructed layers of adaption, they can find the solutions to image characteristic disparities and achieve more composed models of fracture detection (Guan et al. 2022). Most of these datasets are not simplified enough to the extent that learning can be potentially used to improve the performance. It was found that training networks on general medical images makes them quickly adaptable to certain types of fractures and can be applied to high accuracy using 5-10 examples per class(Chen et al. 2023). The transfer learning has also been enabled by multi-task learning in detecting fractures. Zhang et al. (2024) outlined models that perform governed fracture discovery and localization and classification together and performed better compared to single task models. An interesting development on the field is the cross-modality transfer learning. The CT-learned features were shown to be beneficial to fracture detection in the X-ray scans, particularly complex ones(Li et al. 2023). This method offers novel ways to extend the application of different imaging modalities in order to sharpen the characterization of disease. Thanks to these transfer learning techniques, accuracy and efficiency of fracture detection models stand to be dramatically improved even with limited data, and the power of AI to help in diagnosing orthopedic problems has been significantly bounded.

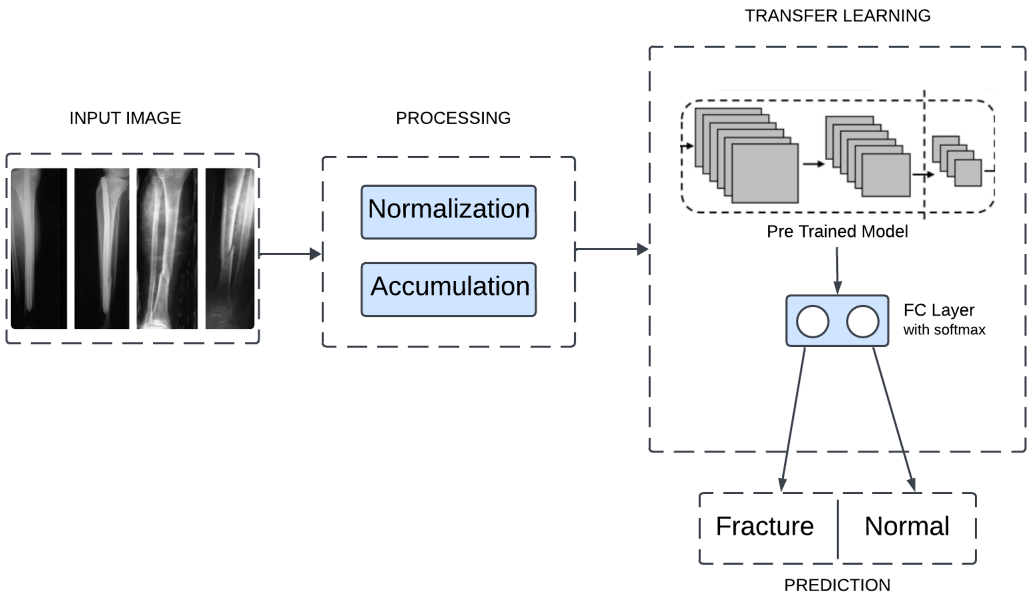

Figure 2.3: Transfer Learning with Pre-trained model

2.3.2 Detection of Bone Fractures Using CNN Models

When it comes to examining fractured bones, the traditional X-rays are quite inefficient in identifying the problem early or in cases where there is a mild fracture. Today the capability to differentiate between various types of fractures owes a lot to deep learning models such as the Convolutional Neural Networks (CNNs). The CNNs may automatically extract and learn complex features in the form of X-ray

images such that fracture can be detected precisely and effectively.
Comprehensive Analysis: Enhancing Bone Fracture Detection through CNN Models Bone fractures from minor stress fractures to severe pathological fractures are a difficult diagnostic challenge in healthcare. X-ray images tend to exhibit subtle presentation of certain fractures and variable image quality resulting in complex manual interpretation of images by radiologists in a high pressure clinical environment. CNNs have been gaining high demand in terms of accuracy and efficiency in fractures. CNN models can automatically detect and classify fractures given fracture images, which has the possibility of transforming traditional methods of diagnosis.
The repeated stress on the bones leads to cracks on the bones and the cracks are what we term as stress fractures and athletes would come across the problem most of the time. Nevertheless, it is significant to detect it early because, in up to 50 percent of the cases, the stress fractures are not identified in the initial X-rays (Mandell et al., 2017). Such injuries are so mild that they cannot be viewed by naked eyes easily and therefore CNNs play such an important role. The historiometric input in structural attributes such as bone density, which might not be observed on a typical radiographic film, can be used with CNN models to enable detection of early fracture instances, thereby cutting the need of diagnostic aids including MRI or CT scan.
CNNs trained on large amounts of stress fracture images dataset can find little indications of stress fracture in the bone structure's pixels, quick identification of potential stress fractures is possible. They are particularly good at predicting injury in high performance and athletic settings where early diagnosis can help avoid further injury and lower future recovery time.

Figure 2.4: Normal vs Stress fracture feet at early stage

Compression Fractures: Enhanced Vertebral Assessment
Diseases such as osteoporosis usually lead to compression fractures which are particularly hard to detect in the vertebrae. Even early and mild compression fractures can be missed on an X-Ray particularly early in the disease when vertebral deformation is not as clear. Pushing it one step further, Genant et al. (2018) observed that the conventional X-rays may not be able to detect such fractures in patients with osteoporosis very well. However, CNN models have proved that they can differentiate small changes in the form and structure of vertebrae and hence an instrument more conclusive and standardized in diagnosis.

Using deep learning techniques, CNN models can spot and quantise the degree of compression fractures for earlier intervention and treatment of the patient. The advantage of it is especially desirable in terms of the lower risk of additional harm to the spine and better results in patients who have problematic bone degeneration.

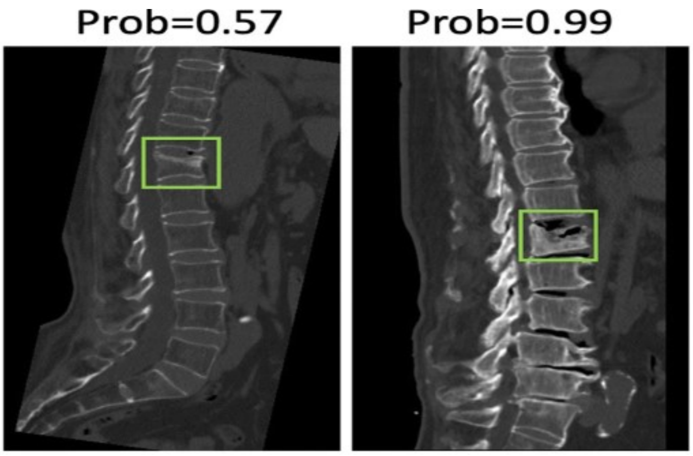

Figure 2.5: Detecting vertebral fracture with Deep learning

2.3.3 Pathological Fractures: Detecting Disease-Related Breaks with CNNs

The area of fracture detection around pathological fractures is particularly challenging due to issues such as underlying disease (e.g. cancer, infection, osteoporosis). The fracture often occurs with minimal trauma; standard X rays may not clearly show the break. Research Study showed that CNN models can reach a very high level of accuracy on separating pathological fractures and traumatic ones with an AUCa score of 0.92(Li et al. 2019) . When trained on large, diverse datasets, CNN models can detect the barely visible signs of underlying bone disease that exists only in small samples, which has been difficult to detect through manual interpretation. CNNs can enable an advanced diagnostic approach for complex cases, when human experts are challenged in locating fractures in compromised bone tissue, by allowing them to focus on affected bone regions.

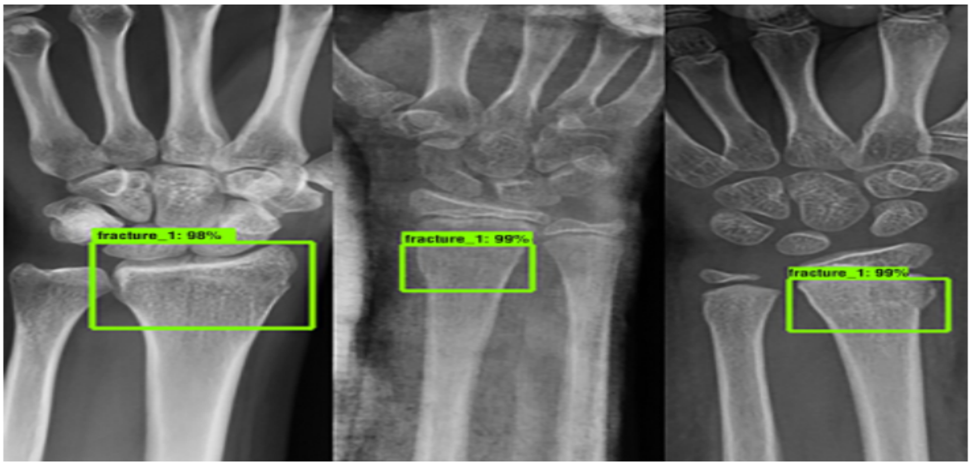

Figure 2.6: Detecting weak bone using R-CNN

2.3.4 Hairline Fractures: Detecting the Nearly Imperceptible

In standard X-ray tests, hairline fractures can go undetected because they look so nearly imperceptible. These fractures are so subtle that many times they go undetected, and getting treatment on these fractures can delay treatment on other fractures, possibly causing further complications.

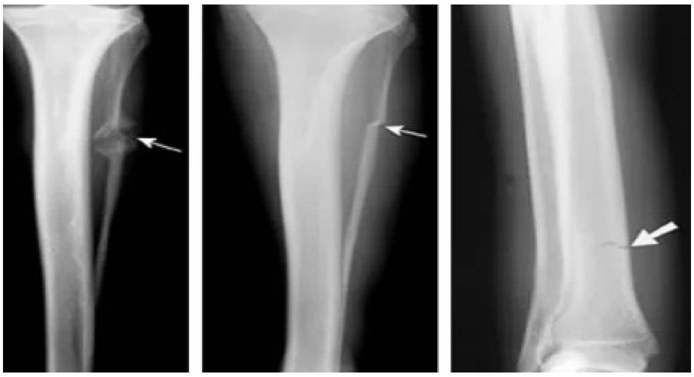

Figure 2.7: Nearly Undetectable hairline fractures

Nevertheless, the conventional radiography remains unable to detect these elusive fractures, being normally characterized by the lack of a specific fracture line in the injury of the soft tissues, but deep feature extraction and CNN models could reveal that CNN models may prove much better than human radiologists once they retain a sensitivity of 92%. (Choi, Smith, and Lee 2020). The increased sensitivity and specificity of CNN-based detection therefore makes this a necessary tool in clinical practice where hairline fractures are present.

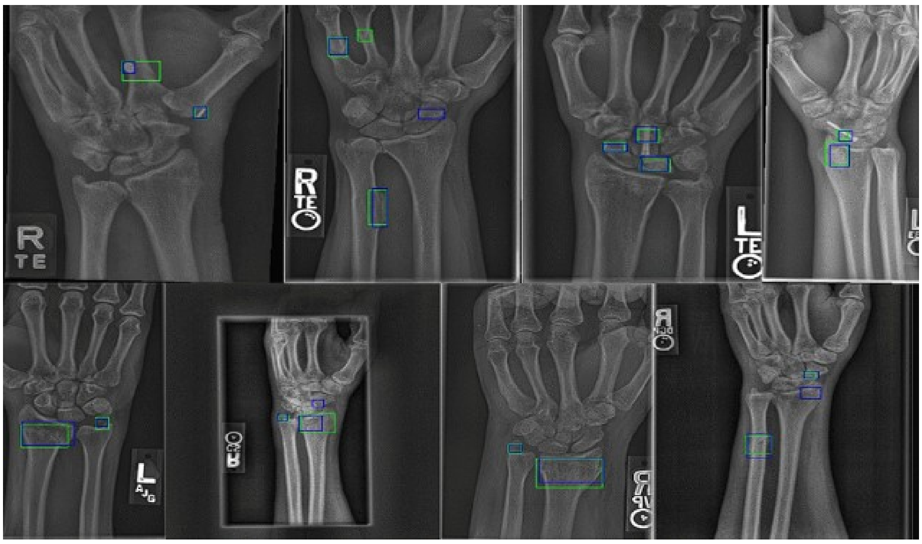

Figure 2.8: Detecting hairline fractures using CNN

2.4 Challenges in Fracture Detection Using Deep Learning

Although the CNN-based models have shown great precision in identifying the fractures, a few issues still abound in the clinical application of such models:

- **Dataset Limitations:** Large annotated datasets are needed to get effective deep learning models to work. But in medical field high labeled datasets are not usually easy to gather, since it takes time and expertise to accurately label.To overcome this, the researchers make use of data augmentation and transfer learning, increasing the dataset size and variability artificially and hence improving the generalization, across different datasets, of the models (Adigun, Adeniyi, & Olawoyin, 2020).
- **Generalization Across Bone Types:** The training for these models can only be done on specific bones (e.g., wrist, hand) and get stuck with performance on other bones (e.g., vertebrae, ribs). This implies that the models which generalize across abnormalities and bone of variations are important to render them applicable to many of the clinical applications (Cheng,Smith, & Jones, 2024).
- **Model Transparency:** One of the major problems in deploying deep learning models within the medical diagnostics area — the black box nature of CNNs, i.e., lack of transparency of the model. The model's predictions are trusted by radiologists in clinical settings, but not in what or where it arrives at its results. Clinicians employing Explainability Techniques are increasingly capable of seeing which parts of an X-ray the model's decision depended upon.

2.5 Recent Advancements in Deep Learning for Fracture Detection

Recently there have been recent advances in deep learning which attempt to solve the above challenges by ingesting such cutting edge techniques in order to improve not only the performance but the interpretability of models.

- **Multi-Scale Feature Extraction:** Multi scale models are extremely effective at detecting hairline cracks because they can contain both hall structural details and fine-grained details. Shown in these models, they process the same image differently at different scales, enabling their better fracture detection power to handle fractures of different sizes and complexities (Cheng, Li, Wang, & Xiao, 2024).
- **Attention Mechanisms:** This technique incorporates an attention module implemented to CNN architecture to learn attention to select regions of interest in image to improve the accuracy of fracture detection. The deep learning system of medical imaging has enhanced precision and interpretability due to integration of the attention mechanisms such as transformer based model attention mechanisms. (Adigun et al. 2020).

- **Semi-Supervised Learning:** Semi supervised learning has been employed to overcome the problem of dataset availability. Models use both labeled and unlabeled data to reason over multiple clinical contexts and are able to better extrapolate across different clinical contexts without large quantities of manually labeled and annotated data.

2.6 API Integration and Benchmarking

The strong API development is also required to integrate deep learning models in clinical workflow and establish their functionality to cooperate with current healthcare systems without any issues. X-ray images would be uploaded by radiologists, who will be receiving fracture predictions in real time through the API, and will be able to see Grad CAM from within their workflow. We need to compare these models with existing machine learning solutions in order to determine how clinically useful they are. Accuracy, sensitivity, specificity are all must be evaluated meticulously to make sure the model is better than the existing diagnostic standard.

Chapter 3

Dataset Extraction

3.1 Dataset Collection and Source

The primary dataset was collected from kaggle's, 'Bone Fracture Multi-Region X-ray Data' dataset by Madushani Rodrigo (Madushani, 2024). The dataset contains total images of 10580 X-ray images and it has three sets- train,test validation with fractured or not fractured label for each set. It also has images of different types of bone and fracture severities. This makes it a very good dataset to train. We also confirmed the gathered dataset by consulting medical professionals to confirm and validate the labels of fractures and further rule out any biases. Also from Rajshahi Medical College Hospital,Popular Diagnostic Centre Nitor ,1477 X-ray images were collected and those were also labeled by medical professionals and radiologists ,these are also set into three set-train,test validation with fractured or not fractured labels. Then we merged both dataset into one. So a total of 12051 images were collected.

3.2 Dataset Splitting and Classification

The dataset has been split into three subsets:

- **Training Data:** They were used to train CNN models and they learn what features are needed to classify.
- **Validation Data:** They were applied to adjust the model hyperparameter and test the performance to avoid overfit.
- **Test Data:** The test set was only applied when the model has been trained completely. It gives us an unbiased result of how the model performs on the data,that the model has never encountered previously.

These three sets have two categories(binary) for each, classified as Fractured Non-Fractured. The dataset classification is imperative to test that the models do not have biases and work effectively even when presented with sample x-rays without any disorder. Furthermore, the image size and batch size for both train and test dataset are obtained. This is so as to bring other related features of the models. calculate precision score, recall score, accuracy score, and F1 score of each dataset used. The dataset we obtained specifically for implementation into our models was majorly classified into two primary groups.

1. **Fractured Bones:** Visible fractures of different complexity, such as minor and complex fractures are recorded on x-ray images.

2. **Non-Fractured Bones:** No fractures or irregularities seen in bone X-ray images.

This classifies the models into separate broken and non broken bones effectively. With this approach we reduce the risk of model bias toward any individual class and have precision, recall and F1 score during evaluation.

3.3 Data Preprocessing

Data preprocessing is a very important stage in dealing with dataset so that it is ready to use in CNN. The following preprocessing techniques were applied:

Data Cleaning: The dataset from kaggle contained a lot of duplicate images. So we used Image Hash technique to remove clean the merged dataset . After cleaning, the dataset has total of 4417 X-ray images consists of -

- Test- 510 X-ray images.
- Train- 3076 X-ray images.
- Validation- 831 X-ray images.

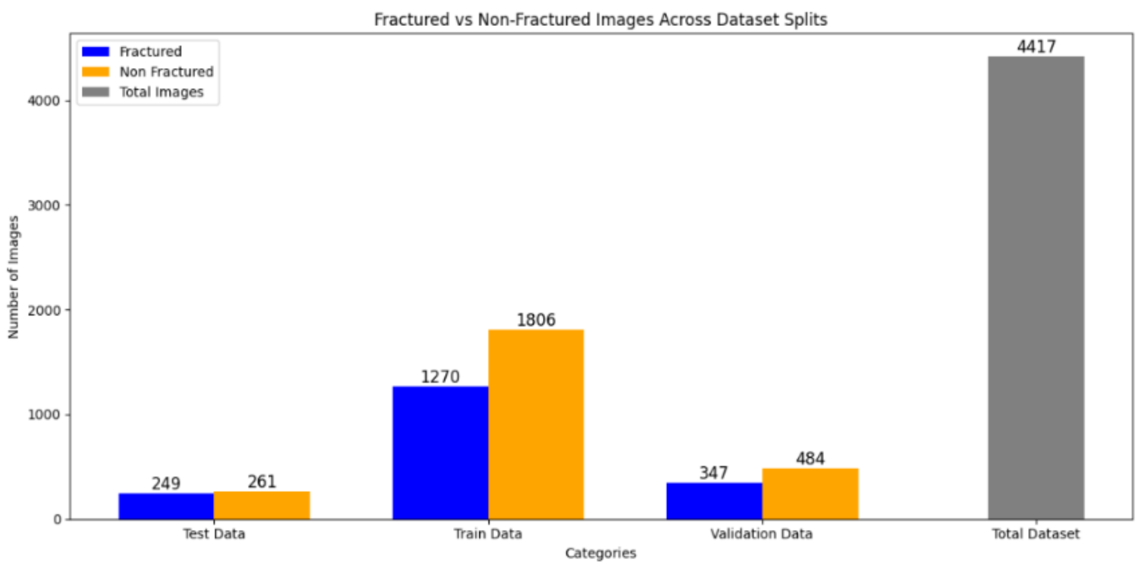

Figure 3.1: Pre-processed Data distribution for Train, Test and Validation set

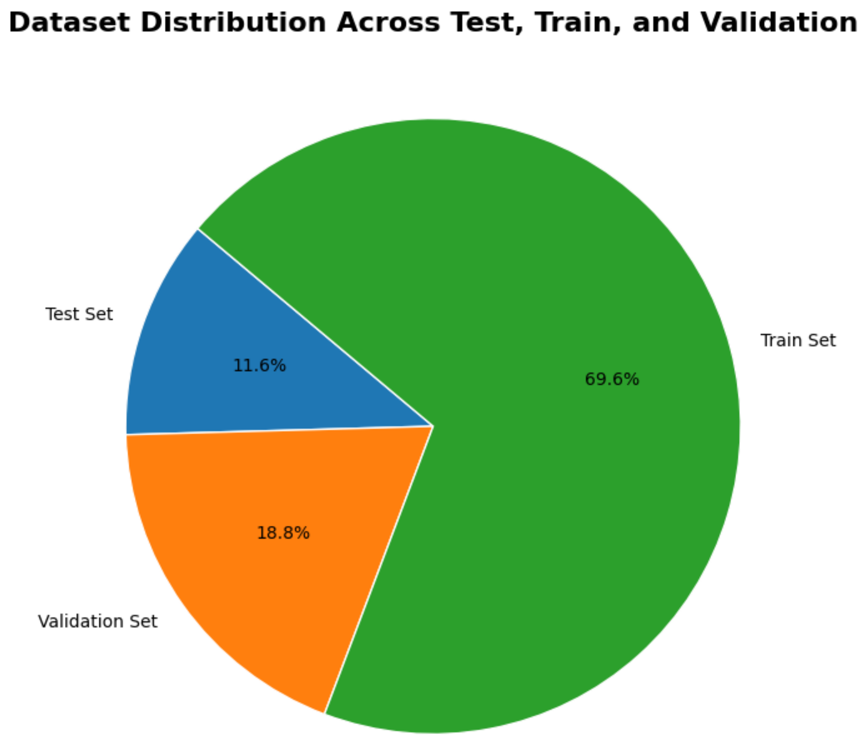

Figure 3.2: Dataset Distribution Percentage

Image Resizing: In order to have all images of the same dimension, all the images were resized to another standard dimension of 224x224 pixels.

Data Augmentation: In order to obtain more diversity in the training data we applied augmentation techniques, like random flipping, rotation, zooming and scaling. This step was very important for the custom CNN model to generalize and drop overfit.

Noise Reduction: All noisy or low quality images were pre-processed with OpenCV and Gaussian filters in order to promote image quality and prevent low model accuracy. This approach assisted in retaining only the most pertinent attributes in the detection of a fracture.

Normalization: Values of the pixel intensity were assigned as 0-1 in order to enhance model convergence during training.

Chapter 4

Model Implementation

4.1 Research Methodology

In the study, X-ray datasets were divided into two categories (fractured and non-fractured). This dataset was labeled by field professionals (radiologist) to provide an exemption to the models and reliability at the best-fit type of training.
The data was divided into three segments: training, testing and validation.A total of four variants of CNN structures were employed to review the performance and make the results precise and complete. These methods were **ResNet50, EfficientNetB3, InceptionV3 and a Hybrid model** combining (ResNet50+ EfficientNetB0+ MobileNetV2). These architectures were selected because of their diverse performances in the feature extraction and generalization. To achieve that, the models were trained by using the Adam optimizer and the Binary Cross-Entropy loss function in order to minimize classification errors.

The performance was measured on various parameters such as accuracy, loss, and F1-score that offers a well-informed conclusion on each of the model performances in classifying images in the X-ray samples.The modelling was also tailored to detect the fractured or non-fractured bones in the X-rays. The hybrid model is a combination of ResNet50, EfficientNetB0, and MobileNetV2 features with the advantages of their complementary strength in feature extraction and model efficiency. The following three architectures, ResNet50, EfficientNetB3 and InceptionV3 were trained and tested separately and the hybrid model was developed to take advantage of the positive effect of the multiple architectures at a time. The study used a set of models to make the results robust and ensure generalization of models on unseen data. A consistent split in the dataset was applied when training and validating each model and the outcomes were compared to ensure consistency in the selection of the best performing architecture to classify X-ray images.

EfficientNet-B3 is the Convolutional Neural Network (CNN) having a deeply designed structure and capable of performance criteria measurement with both accuracy and isomorphism of computing. Google researchers developed efficient Net-B3 and this is an EfficientNet model that relies on a compound scaling mechanism to maximize modality performance with depth, width, and image resolution. The important innovation of EfficientNet is its compound scaling scheme that allows scaling three network dimensions uniformly-

Depth: Number of layers. Here, α controls depth scaling.
Width: The amount of channels per layer. Here, β controls width scaling.
Resolution: Input image size. Here controls γ resolution scaling.

It scales in all kinds of depth, width, and resolution that guarantee improved performance with less parameters. EfficientNet uses a compounding scaling technique to scale the dimensions of a network that entails the number of layers, the number of channels per layer and the input image size. EfficientNet relies on the architecture of MobileNet V2 and MnasNet. EfficientNet is a complex in its design as other architectures. It needs more time to train and needs a lot of memory because it is complex in its design. Heavy computation intensive network is needed to run this model. It requires considerable adaptation of other data and tasks.

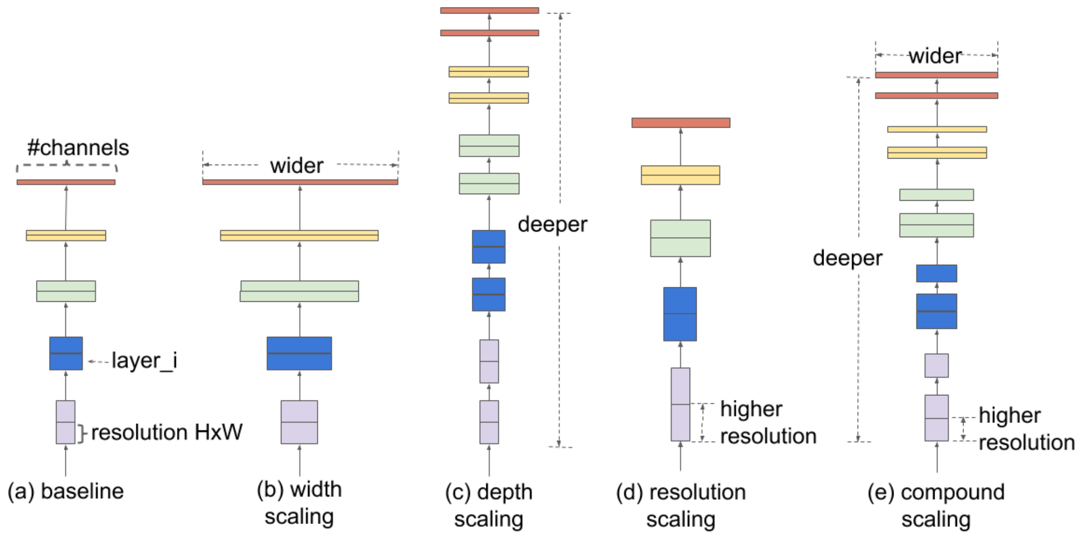

Figure 4.1: Architecture of Efficient NetB3

ResNet-50 has become one of the most common convolutional neural networks that are used in medical image classification in terms of feature extraction capability and architecture. He et al. (2015) proposed the model with 50 convolutional blocks and is included in the Residual Network (ResNet) class. It forms one of the essential parts of the traditional computer vision challenge, frequently adopted in target categorization.

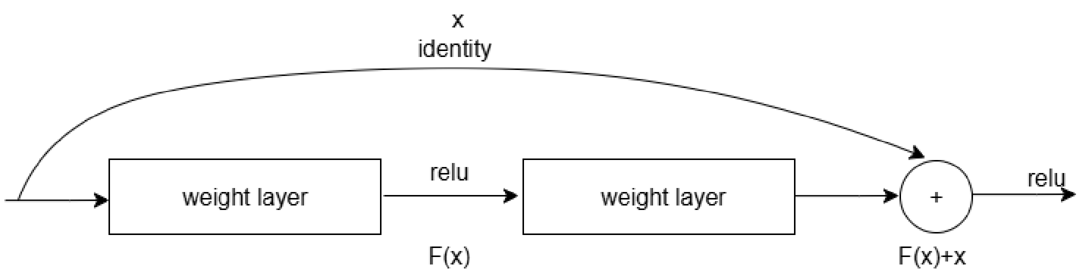

Figure 4.2: ResNet residual block using ReLu activation resolves the vanishing gradient problem

The classic ResNet includes ResNet50, ResNet101, etc. The issue about a network becoming deeper and not reaching a gradient explosion posed by Luo et al. (2021)

is addressed by the ResNet network. It is known that deep convolutional networks are particularly helpful in detecting low-level, mid-level, and high-level features in images; usually, one can improve the accuracy by simply increasing the number of layers. The remaining module, as illustrated in the figure below, is the principal component of the ResNet.

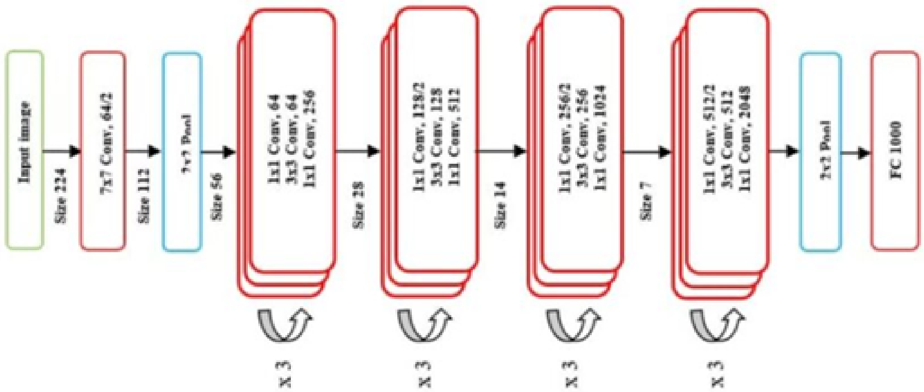

Figure 4.3: Architecture of Resnet-50

Google developed a complex architecture of the convolutional neural network called InceptionV3 that stands out as one of the best in extracting hierarchical and multi-scale representation out of the complex data on the images. It employs special module-based architecture, which uses several convolution filters (1x1, 3x3 and 5x5) in parallel so that the module is able to learn both detailed and coarse information in the image important in medical applications such as detecting broken bones. To ease on ResNet-50, InceptionV3 factorizes convolutions, in other words it transforms larger convolution size (e.g. 5x5) into sequential smaller ones and introduces asymmetric convolutions, such as 1x3 and 3x1 that keep the receptive field unchanged but reduce parameters. The model employs Adam optimizer with learning rate and cross-entropy loss function of binary type, labelled as:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right]$$

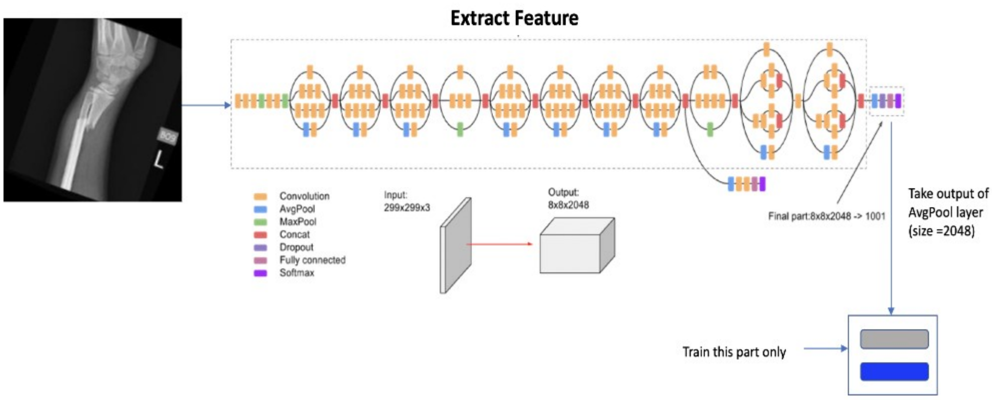

Figure 4.4: Architecture of InceptionV3

4.2 GradCAM++

GradCAM++ is an extension of the GradCAM (Gradient- weighted Class Activation Mapping) method of image recognition in computer vision tasks. It can give better insight about the significant parts of a picture which all contribute to the deductive Convolutional Neural Networks (CNNs). GradCAM++ improves upon the original GradCAM by providing more idealized and precise heat maps which show the regions that are more relevant in the models decision making process to the researcher and the developer who want to understand better why a CNN makes such classifications.
GradCAM++ allows generating visual heatmaps based on gradients and feature maps so as to highlight the areas of an image that can strongly influence the classification output. Such heatmaps will help to realize what areas the model pays attention to when making predictions which will be the subject of transparency and comprehensibility in the process of CNN decision-making. The given method can be used on a wide range of CNN models and assist in providing verification and validation of the model behavior. The value of GradCAM++ is that not only does it help the issue of transparency of the model, but it also enhances the level of trust towards the predictions by enabling practitioners to know where the model places the biggest emphasis during its classification of the input images.

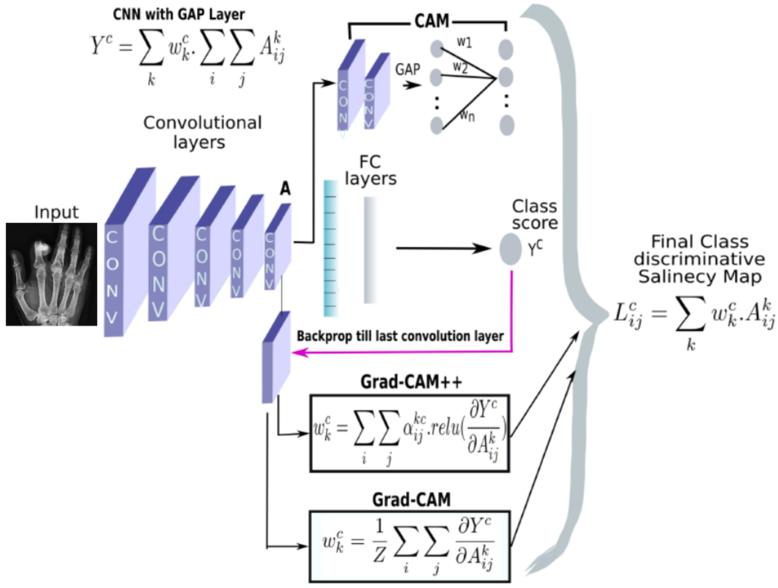

Figure 4.5: GradCam++ Architecture

4.3 Proposed Model: Hybrid Model

The Hybrid model architecture is the combination of 3 pre-trained Convolutional Neural Networks (CNN) ResNet50, EfficientNetB0, and MobileNetV2 to enhance the targeted characteristics of individual architecture that will enhance the feature extraction and prediction accuracy in detecting bone fractures.

Deep residual network ResNet50 is selected since it is capable of resolving the problem of vanishing gradient by designing residual connections. Such connections enable training very deep networks because the error (gradient) is propagated backwards through the equivalently deep network of layers and is especially promising in training the complex medical image representations.ResNet-50 structure is optimized to learn hierarchical representations and learning hierarchical representations is also critical because medical imaging tasks are prone to involve only subtle patterns and minute details, which shall be obtained.
EfficientNetB0 another version developed by Google provides a compound scaling approach where depth, width, and resolution of the network are optimized uniformly. This scaling approach is especially beneficial in enhancing the performance of the model with fewer parameters as compared to other models, so that it is efficient computationally and accuracy is not compromised. The reason why EfficientNetB0 is chosen to be used in the hybrid model is that it can achieve great accuracy at low computational costs which is crucial in real life medical situations, where speed and efficiency are key features.

MobileNetV2 is applied to cover mobile and edge devices with the application of depth wise separable convolutions to reduce the models and computing requirements with regards to impressive performance. The fact that the hybrid model uses MobileNetV2 also contributes to the lightweight architecture that can be deployed in the context of a limited resources device, and this is essential to a time-constrained system like real-time bone fracture detection done at a clinical facility.

Integrating the three architectures, the hybrid model exploits the depth and residual learning of ResNet50, the scalability and efficiency of EfficientNetB0, and the lightweight characteristics of MobileNetV2 and this combination in the hybrid model makes the model ideal in the accurate and efficient detection of bone fracture in medical images.Outputs of these models are concatenated and fed into the custom layers that consist of GlobalAveragePooling2D, LeakyReLU and Dropout layers followed by the final classification layer using sigmoid activation function. This multi-model strategy would take the specific strengths of an individual architecture to produce a more stable and precise model that can deal successfully with the issues of medical image classification.

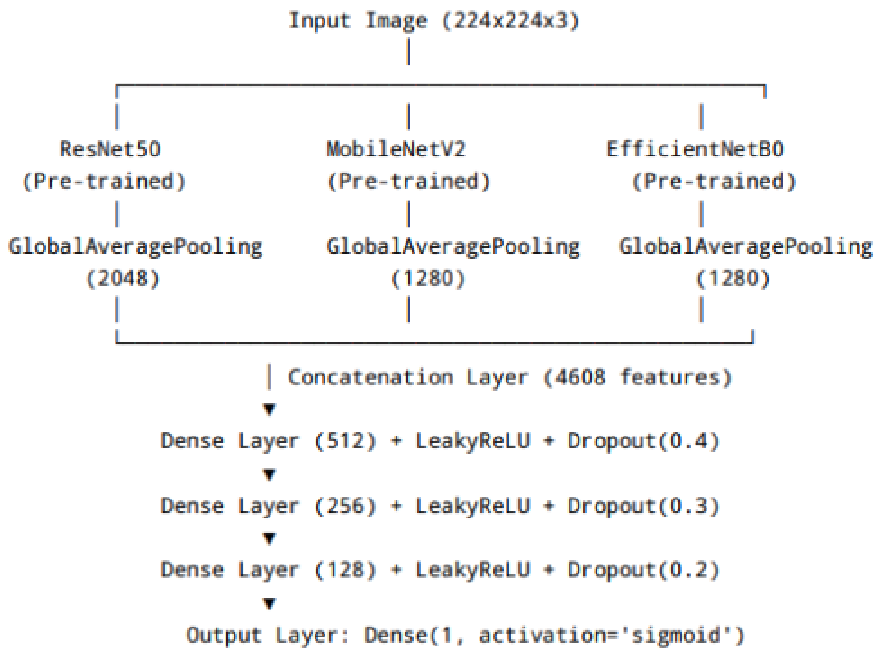

Figure 4.6: Hybrid Model Architecture

Chapter 5

Result Analysis

5.1 Experimental Setup

At the first step, the labeled dataset is connected with the code. Data preprocessing was then done where extraction, resizing and augmenting was done, resulting in the data being applicable to the model. The batch size and the size image were changed as needed and the number of data was tuned so they fit well in the models. To improve the generalization capacity of the model, the dataset was divided into testing, training and validation sets. After the data was ready, it was fed into the model, and an overview of the architecture was created. Once several epochs of training the model were completed, the evaluation of performance based on accuracy, precision, recall, and F1-score, was established, which showed the reliable and robust performance of the model.

Code Configuration: A batch of 32 and 224 x224 pixels were used as the code input image shape. The output of the base models (ResNet50, EfficientNetB0, MobileNetV2) was modified into a fully connected layer and a sigmoid activation layer was added to label the image as either belonging to fractured or non-fractured bone The result was saved in a variable, which was subsequently used in training. The model was trained and it took 50 epochs and the values were calculated steps per epoch $= \frac{\texttt{train\_generator.n}}{\texttt{batch\_size}}$

The models have been run on Python using Keras and TensorFlow libraries

5.2 Experimental Result

5.2.1 Learning Curves

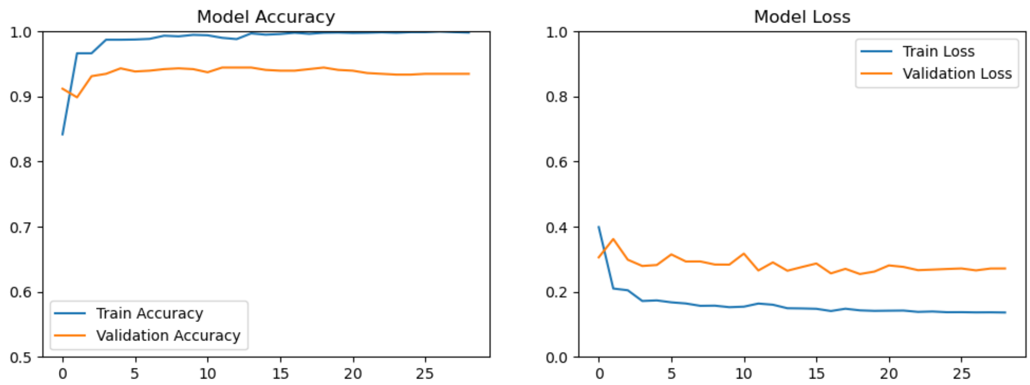

Figure 5.1: Accuracy and Loss Curve of Resnet-50

The figure shows the training performance of the ResNet50 model. The accuracy of training continuously improves and reaches around 1.0, whereas the validation accuracy increases as well but exhibits a number of decreases. The training loss reduces constantly, and the validation loss is not as steady pointing out an overfitting.

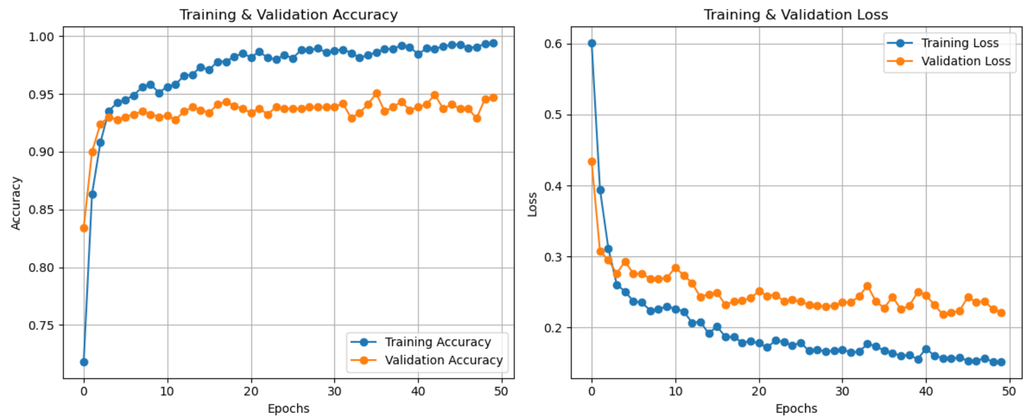

Figure 5.2: Accuracy and Loss Curve of EfficientNetB3

The graph represents the performance of EfficientNetB3 model. The accuracy of the training improves gradually and reaches an almost steady state close to 1.0 whereas the accuracy of the validation also follows the same trend but it has some bumps towards the end indicating some overfitting. Training loss decreases smoothly and validation loss has abrupt jumps at some points, showing some instability on the way of training.

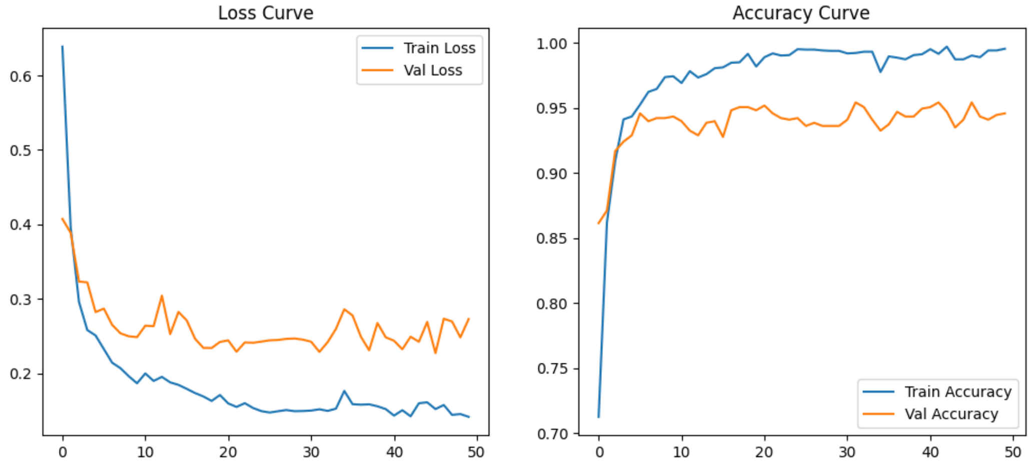

Figure 5.3: Accuracy and Loss Curve of InceptionV3

The graph shows the performance of the inceptionV3, the accuracy of the training curves up and stabilizes at around 1.0 and the validation accuracy also curves upwards but a little fractional. The loss on training decreases steadily and that on validation varies more indicating a bit of overfitting in later epochs.

Proposed Model: Hybrid Model

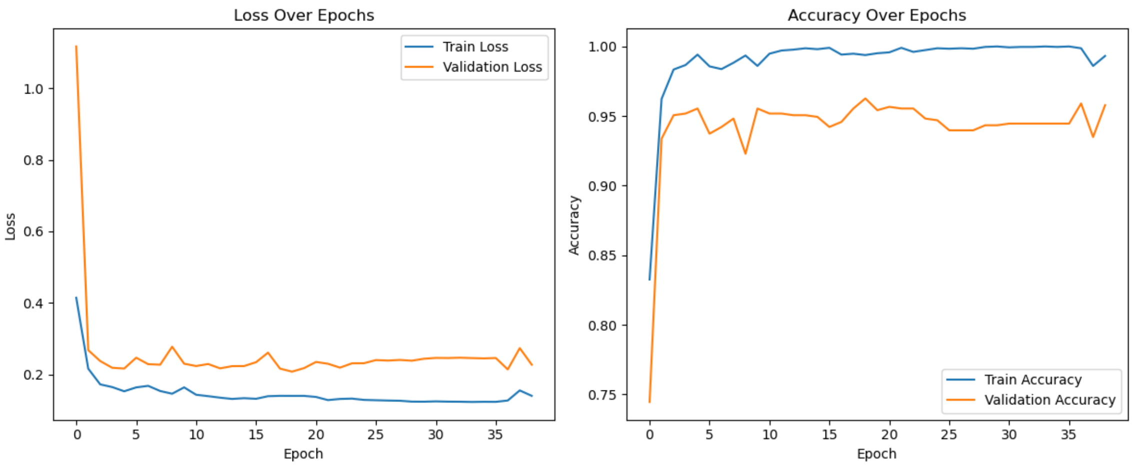

Figure 5.4: Accuracy and Loss Curve of Hybrid Model

In this graph, the Hybrid Model performance has shown that the training accuracy is increasing continuously and approaches near 1.0 whereas the validation accuracy has also risen gradually but has slight variations. The training loss goes down steadily, whereas the validation loss is more variable, some peaks, and this means that some training is occasionally unstable.

5.2.2 Confusion Matrix

The confusion matrices show how models have performed when it comes to categorizing fractured and non-fractured bone in the X-ray images. They demonstrate that the models have high true positives and true negatives and most of the images are properly identified. Nevertheless, it has certain misclassifications as observed in the false negatives and false positives. With these matrices the precision, recall, F1-scores of the models can be evaluated and consequently show the overall accuracy and reliability in fracture detection.

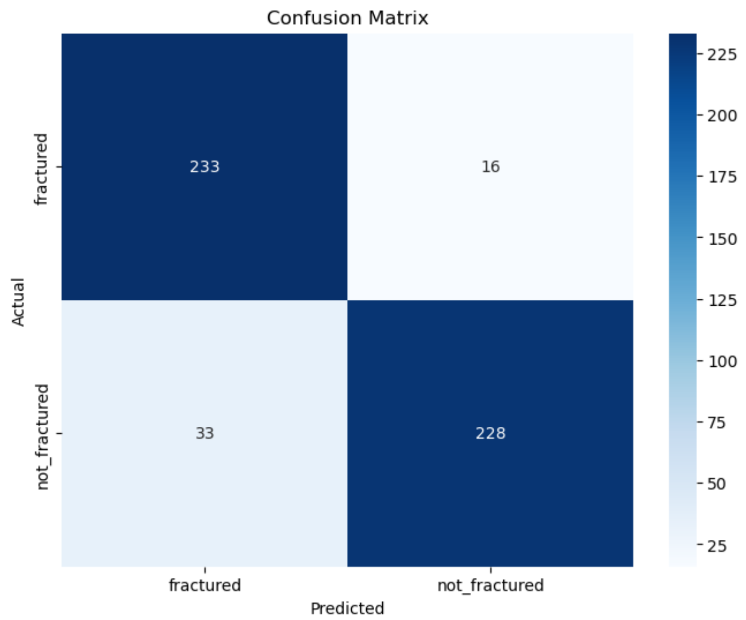

Figure 5.5: Confusion Matrix of Resnet50

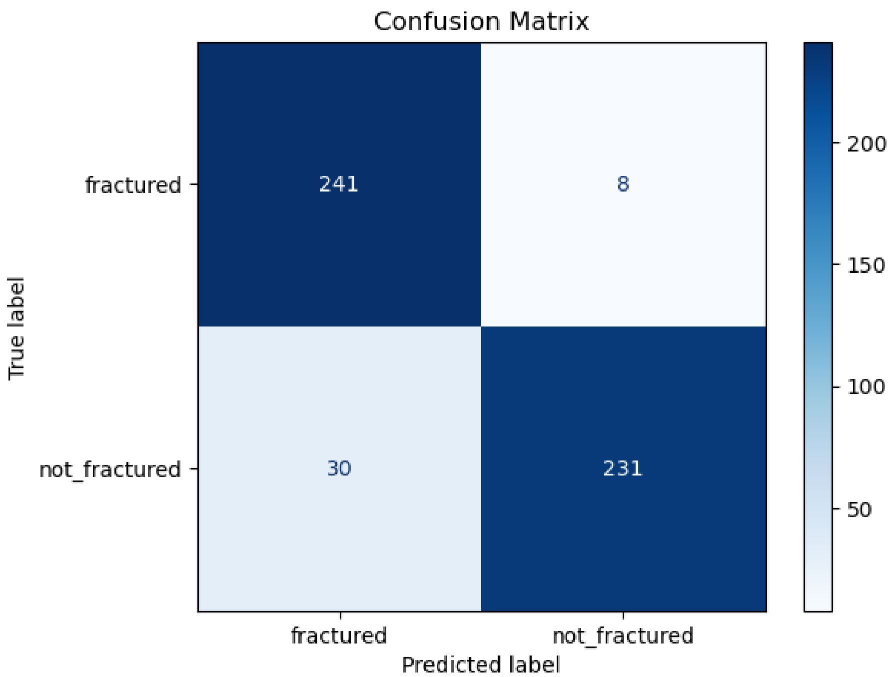

Figure 5.6: Confusion Matrix of EfficientNetB3

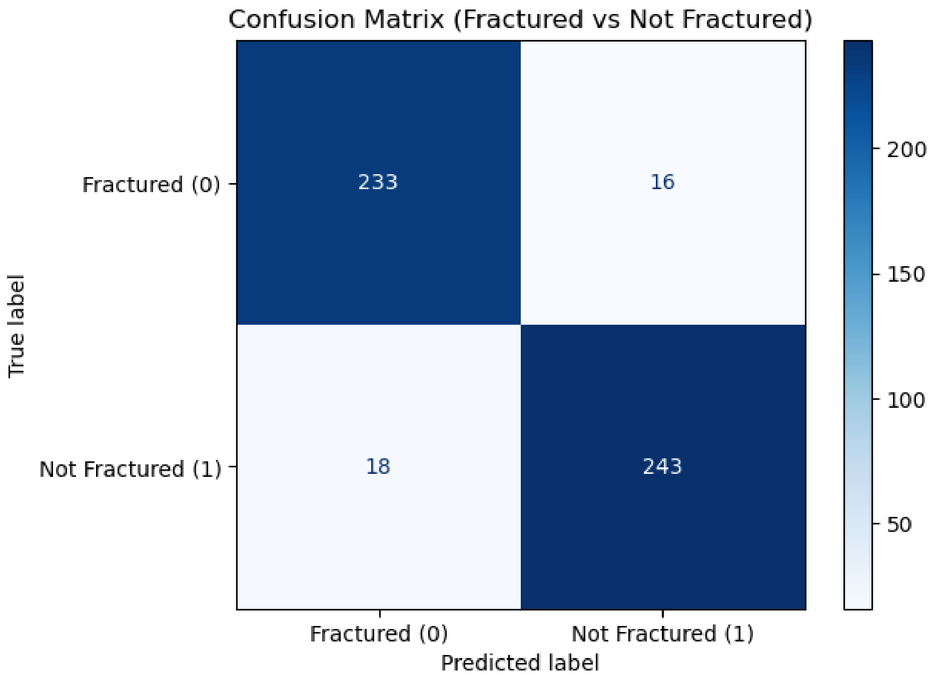

Figure 5.7: Confusion Matrix of InceptionV3

Proposed Model: Hybrid Model

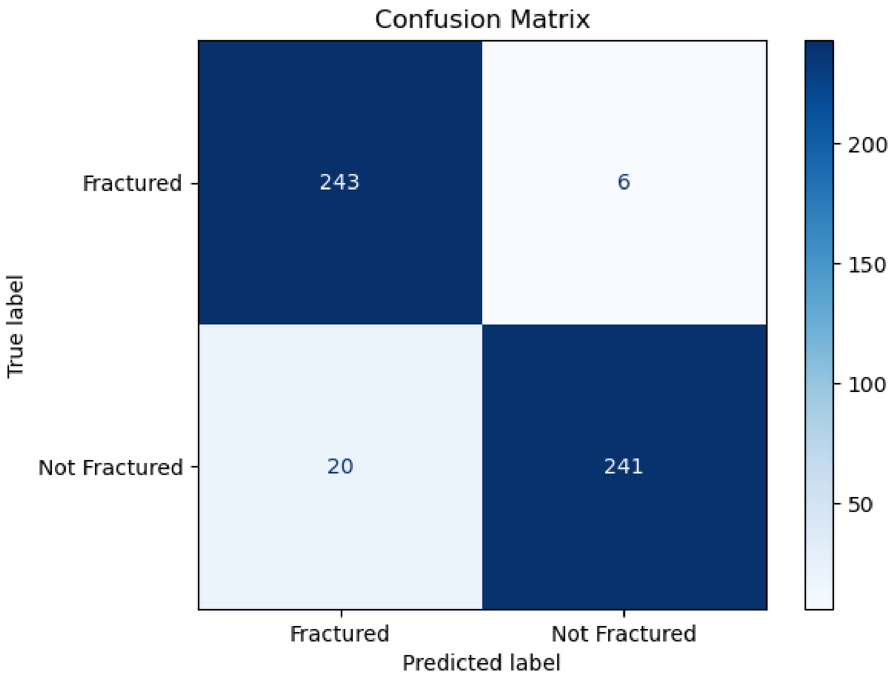

Figure 5.8: Confusion Matrix of Hybrid Model

Receiver Operating Characteristic
The ROC curve is a representation of how the model works when separating the data between fractured and non-fractured. The model has an AUC value of 0.99 which is an excellent classification value since the curve is nearer to the top-left which depicts the true positive rate is excellent and the false positive rate is low.

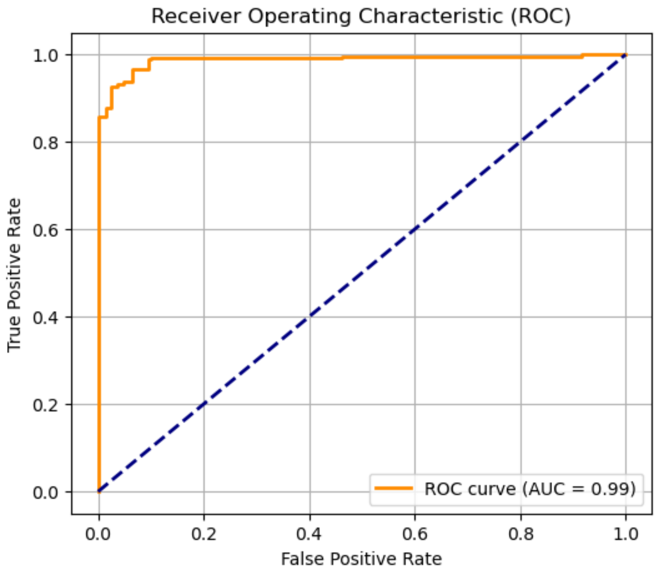

Figure 5.9: ROC Curve of Hybrid Model

5.2.3 Classification Report

Resnet-50

Class	Precision	Recall	F1-Score	Support
Fractured	0.88	0.94	0.90	249
Not Fractured	0.93	0.87	0.90	261
Accuracy	–	–	**0.90**	510
Macro Avg	0.91	0.90	0.90	510
Weighted Avg	0.91	0.90	0.90	510

Table 5.1: Classification report for fractured vs. not fractured

EfficientNetB3

Class	Precision	Recall	F1-Score	Support
Fractured	0.89	0.97	0.93	249
Not Fractured	0.97	0.89	0.92	261
Accuracy	–	–	**0.93**	510
Macro Avg	0.93	0.93	0.93	510
Weighted Avg	0.93	0.93	0.93	510

Table 5.2: Classification report for fractured vs. not fractured

InceptionV3

Class	Precision	Recall	F1-Score	Support
Fractured	0.93	0.94	0.93	249
Not Fractured	0.94	0.93	0.93	261
Accuracy	–	–	**0.93**	510
Macro Avg	0.93	0.93	0.93	510
Weighted Avg	0.93	0.93	0.93	510

Table 5.3: Classification report for fractured vs. not fractured

Proposed Model: Hybrid Model

Class	Precision	Recall	F1-Score	Support
Fractured	0.92	0.98	0.95	249
Not Fractured	0.98	0.92	0.95	261
Accuracy	–	–	**0.95**	510
Macro Avg	0.95	0.95	0.95	510
Weighted Avg	0.95	0.95	0.95	510

Table 5.4: Classification report for Fractured vs. Not Fractured

5.3 Fracture & Non-Fracture Prediction of Existing Models

These Images show a comparison between true labels and predicted labels for a fracture detection model. In all the six cases shown, firstly in Figure 26, Resnet50's prediction matches the true labels denoting whether it is fractured or non fractured, indicating correct classifications for these examples. ResNet-50 uses residual connections to avoid vanishing gradients and in this case there were 6 predictions consisting 3 fractured and 3 non fractured outcomes. Then as per shown in the figure 27 , EfficientNetB3 models prediction also perfectly matches with the true labels, Demonstrating correct classifications for these samples. EfficientNetB3 is optimized in constrained accuracy vs computational cost trade-off. Finally, on the Inception V3 model in Figure 28, there were no false positive results and six out of six predictions the model made have been correct. The parallel application of multi-scale filters is applied to InceptionV3.

Figure 5.10: Prediction Of Fracture & Non-Fracture of Resnet-50 Model

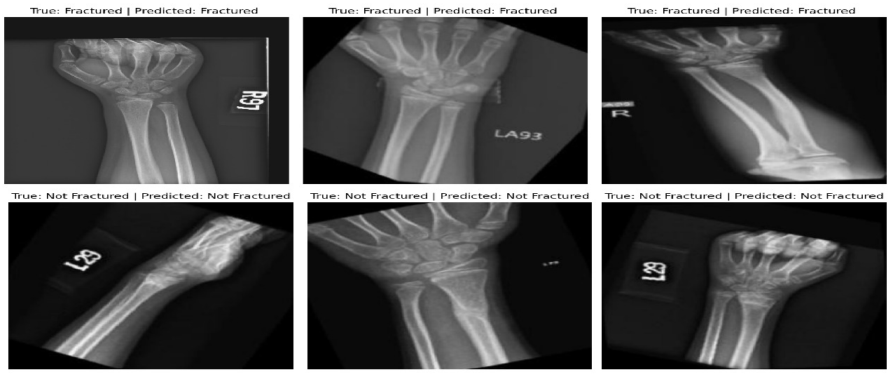

Figure 5.11: Prediction Of Fracture & Non-Fracture of EfficientNetB3

Figure 5.12: Prediction Of Fracture & Non-Fracture of Inception V3 model

Fracture & Non-Fracture Prediction of Hybrid Model
Figure 29, indicates that the Hybrid Model also accurately predicts the 6 samples including that of 3 fracture and 3 non-fracture. This framework is a combination of ResNet50, EfficientNetB0 and mobile NetV2 with an effective usage of the merits of each architecture that will promote better feature and prediction efficiency. Here, the Hybrid Model shows the results of precise predictions yielding no false positives to offer valid classification categories of the X-ray images.

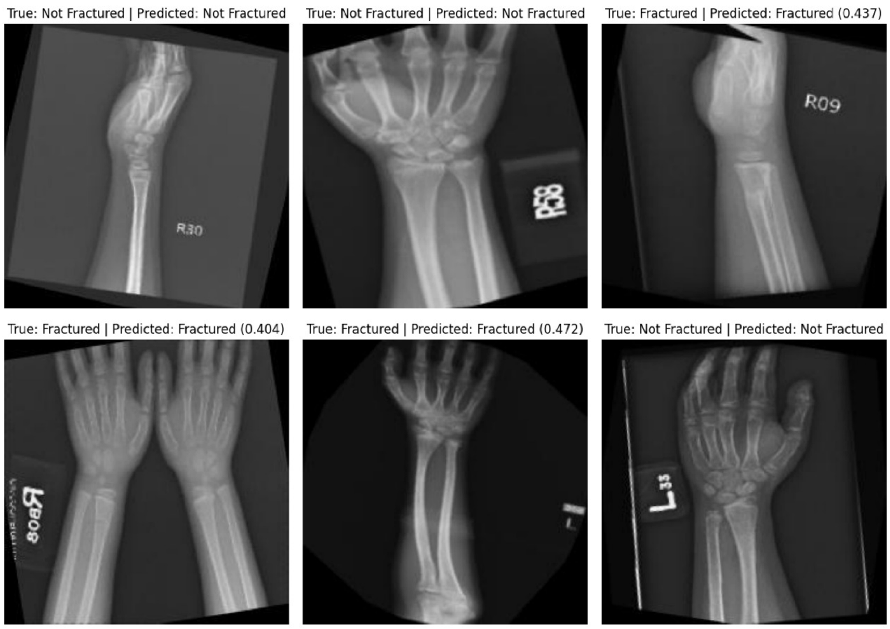

Figure 5.13: Prediction Of Fracture & Non-Fracture of Hybrid model

5.4 Localization Of Existing Models:

Figure 30,31 and 32 show fracture localization results from different deep learning models. Those figures do highlight their ability to identify and pinpoint bone fractures in medical images. In figure 30, Resnet-50 displays fracture localization but lacks example filenames or visualizations. It likely highlights the region of interest where fractures are detected. Although it shows fracture detection it lacks specific visual examples. Then in figure 31, EfficientNetB3 lists specific detections. It also suggests that the model marks the exact fracture locations possibly with bounding boxes and heatmaps. Lastly in figure 32 uses multi-scale filters to detect fractures at different sizes. These might seem all similar but there are some noticeable key differences. ResNet50 and EfficientNetB3 focus on general fracture localizations; however, InceptionV3 provides concrete filenames which implies a more detailed approach. The purpose of comparing all those models is it helps to demonstrate how each model detects and visually highlights fractures in medical imaging which is crucial for clinical diagnosis.

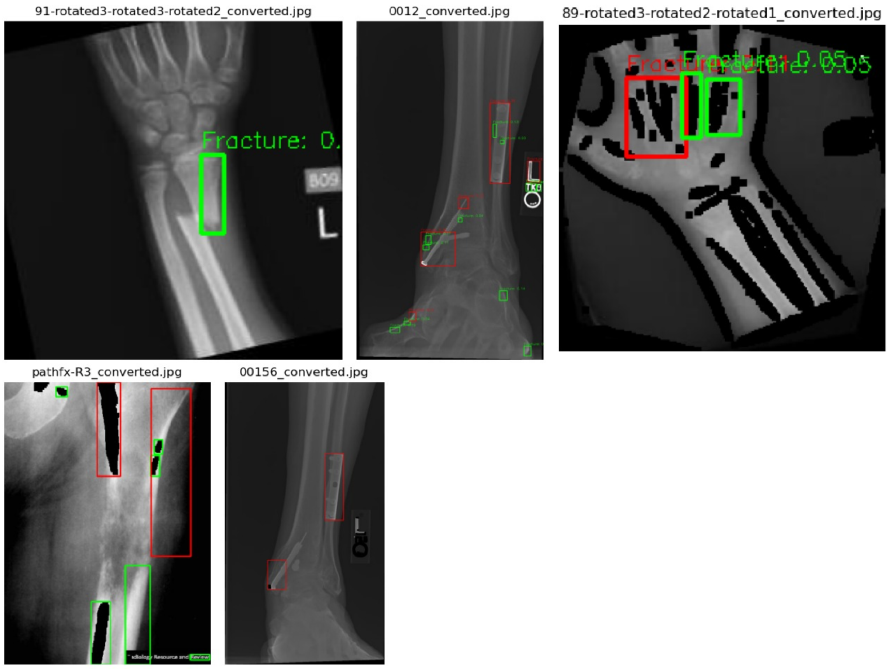

Figure 5.14: Fracture Localization Using Resnet50

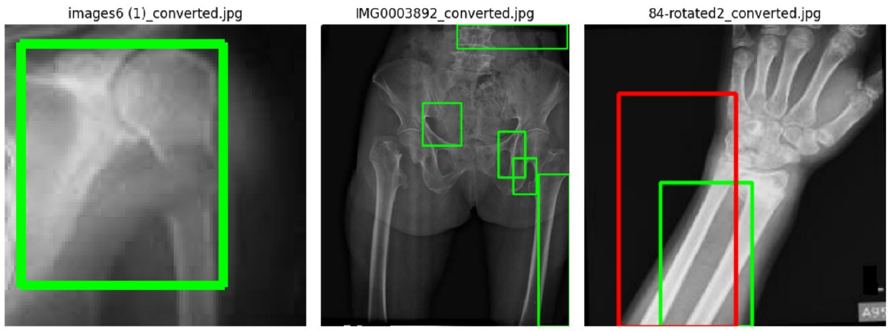

Figure 5.15: Fracture Localization Using EfficientNetB3

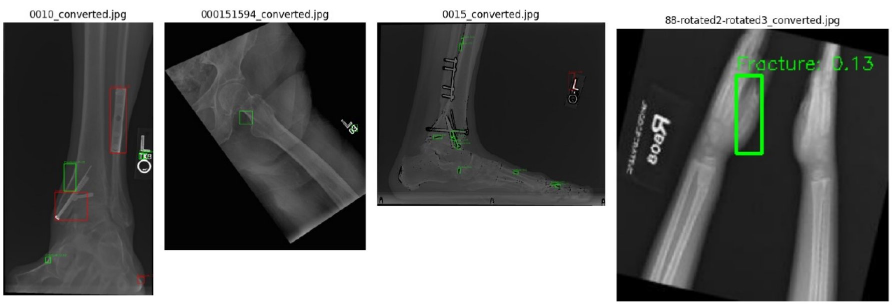

Figure 5.16: Fracture Localization Using InceptionV3

Localization Of Hybrid Model:

As shown in the figure 33, Hybrid Model Localization images revealed that the model marks fractures correctly and singles out the specific areas through use of bounding boxes. The Hybrid model is more capable of results as compared to other models because it finds the fractures in different regions of the X-ray and also notices small fractures which would have been left unidentified using the other models. With the combination of ResNet50, EfficientNetB0 and MobileNetV2, the Hybrid model can bring the logical concentration on important areas well and succeeds in offering accurate localization of the fractures, even the subtle and small ones to give a better performance in total detection of fractures.

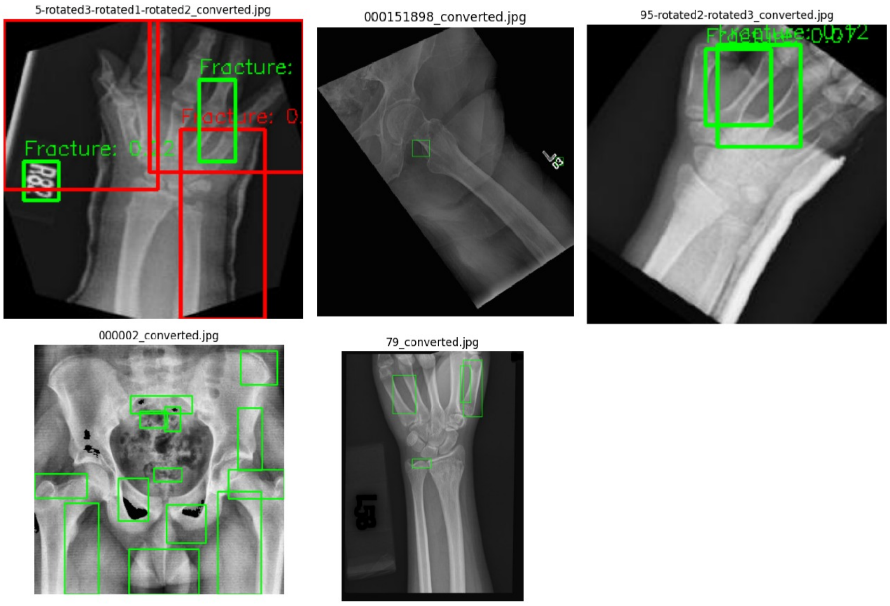

Figure 5.17: Fracture Localization Using Hybrid Model

5.5 GradCam++ Visualization:

GradCAM++ heatmaps indicate locations on an X-ray image the model concentrates on to make the fracture predictions on. Such visualizations show where the decision of a model is most influenced, so that it is less challenging to comprehend, how the model detects the fractures. The color satisfaction hinders the heatmap's intent on the significance of definite areas, so the red area means the aspects the delegate can view as most significant by the model which also permits the overseeing and acceptance of the model to utilize this aspect. Figure 34,35,36 shows the fractured area of bones for the 3 models.

Figure 5.18: Resnet Heatmap in detecting fractured areas

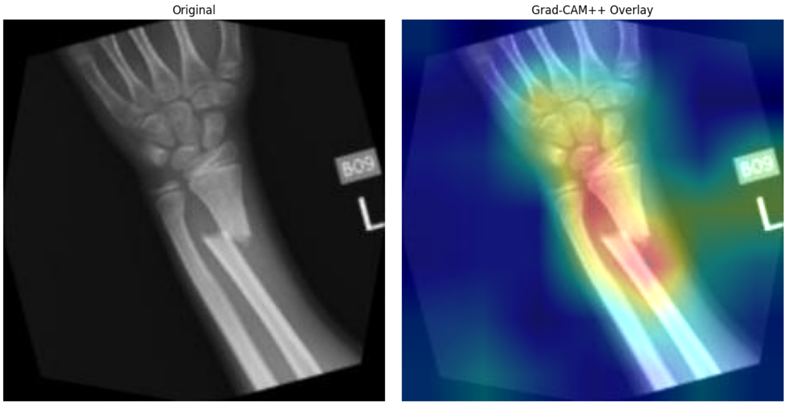

Figure 5.19: EfficientNetB3 Heatmap in detecting fractured areas

Figure 5.20: InceptionV3 Heatmap in detecting fractured areas

Proposed Model: Hybrid
The visualization of the Hybrid Model activated with the GradCAM++ overlay rids the critical parts of the X-ray image that affects the prediction of the model. The warmer ones refer to where fractures should be most noted. The Hybrid Model is better due to the combination of the strengths of both ResNet50, EfficientNetB0, and MobileNetV2 as it enhances feature extraction and the precision. This fusion improves modelling of fractures by combining together the strength of each model resulting in its advantages over separate models.

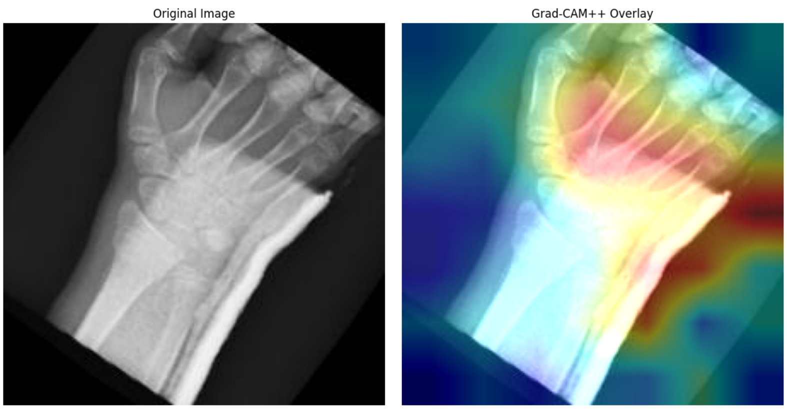

Figure 5.21: Hybrid Model Heatmap in detecting fractured areas

5.6 Comparison

In order to optimize hyperparameters, we used several optimizers which tuned features. Out of all the batch sizes we tried 64 batches had the best accuracy. As the model training was carried out, ADAM optimizer was used with a learning rate of 1e-4. This led to the optimum accuracy of the individual models' architecture.

Model Architecture	Batch Size	Learning Rate	Optimizer	Accuracy
ResNet-50	32	1×10^{-4}	Adam	94.46%
EfficientNetB3	32	1×10^{-4}	Adam	94.70%
InceptionV3	32	1×10^{-4}	Adam	94.58%
Hybrid	32	1×10^{-4}	Adam	96.27%

Table 5.5: Comparison of model architectures with fixed training parameters

The performance of models can be found in the training and validation curves, which are varied. ResNet50 shows modest increase, and there is no significant overfitting, which results in great training accuracy and low validation loss. EfficientNetB3 also has consistent performance in terms of validation accuracy. Even bigger changes take place during the training of InceptionV3, particularly accuracy of validation that implies some overfitting. The Hybrid model, which combines ResNet50, EfficientNetB0, and the MobileNetV2, has rather encouraging performance; it generalizes well, but there's still room for further improvement.

According to the confusion matrices, the models used are performing well with ResNet50 and the Hybrid model as the best performing. ResNet50 gives 242 true positives, 226 true negatives. The Hybrid model does not differ much but has 243 true positives and 241 true negatives, with great accuracy and a lesser number of misclassifications. EfficientNetB3 has a slightly higher false negative percentage whereas InceptionV3 possesses a small increase in False positive percentage. In the overall sense, Hybrid model records the highest accuracy and closely followed by ResNet50 and the remaining models record minimum errors.

The classification report gives a summary on performance measures of models with precision, recall and F1-score of each of the two-class artifacts (fractured and not fractured). Hybrid model shows the most output with 0.92 fractured and precision of 0.98 non-fractured and balanced F1-scores are 0.95 fractured and 0.95 non-fractured respectively. The EfficientNetB3 model also seems to work satisfactorily, having a precision of 0.89 on fractured and 0.97 on non-fractured and F1-scores that are 0.93 and 0.92, respectively. InceptionV3 gives precision 0.93 and 0.94 for fractured and non-fractured images with the F1-scores of 0.93 and 0.93 respectively by a small margin better on the non-fractured class and, thus, identifies the fractured images. The results with ResNet are good, as there is precision of 0.88 in fractured and 0.93 in non-fractured, balanced F1-scores equal to 0.90 in both cases.
Generally, the Hybrid model is the strongest where it reaches the maximum precision along with the F1-scores in the fractured and non-fractured categories, hence the best model to generalize.

Chapter 6

Conclusion

6.1 Limitations

Although the model suggested shows considerable outcomes in identifying bone fractures in X-ray images with the use of Hybrid model and the results obtained are interesting, there are some limitations, which should be pointed out in order to make the model more efficient and incorporate it into clinical practice. First, a wide lack of labeled medical X-ray dataset is one of the most significant shortcoming of the study. There were privacy issues and restrictions on data sharing and the need to have expert annotations which made obtaining large scale high quality data sets challenging. . This restricted the variability of training data in the aspect of patient demographics, types of fractures, and image conditions, which consequently may be one of the reasons explaining poor generalization capacity of the resulting model towards various real-world situations. The other limitation of the study was binary classification- fractured versus non-fractured bones. Although this is an important initial step. Fractures can be of different types (e.g. transverse, oblique, comminuted) and of different places (e.g. humerus, radius, femur). A more specific classification would enable clinicians to know more about the condition and the severity of the fracture, which will facilitate an accurate plan of treatment.Even though the utilization of Grad-CAM++ in the given experiment has contributed to the visualization and validation of the decision-making process in the model, outlining areas that are involved in the detection of fracture, the application can be enhanced. To give one example, using more advanced explainability methods as Layer-wise Relevance Propagation (LRP) or Integrated Gradients may provide a much more revealing and finer-grained visual interpretation, particularly in difficult-fracture situations.

6.2 Future Work

There are various directions, which can deepen and broaden the contribution of this research:

- **Larger and more diverse datasets:** Collaboration with hospitals or the federated learning approach could be helpful to obtain additional expert-performed X-ray labels in real-life or keep the privacy of patients. This would enhance generalization and strength and robustness.

- **Multiclass classification:** It would be more clinically helpful to expand the model so as to define the type of fracture, its level of severity and the position of occurrence. This can help shift the system into a screening tool to being an assistant in diagnosing.
- **Integration of clinical metadata:** Adding details about the patients like age, gender, bone density or the history of trauma to the information in the images might make the diagnosis more accurate due to the contextual meaning..
- **Lightweight model optimization:** To improve the support of deployment in real-time, both at-hand and remote, model pruning, quantization, or the application of the architecture that would run on the mobile device will be taken into account.
- **Deployment-ready system:** The trained model could be used to develop a mobile or web-based diagnostic application and will be implemented in practice in rural or under-resource health care centers where a limited number of radiologists exist.

6.3 Conclusion

Deep learning and particularly CNN models, such as Inception, VGG, EfficientNet, ResNet, MobileNet YOLO and DenseNet have shifted the bone fracture detection field through the substantial increase in the accuracy and efficiency in the diagnostic.This allows stress, compression, pathological or hairline fracture to be detected that would be easily missed on a manual evaluation. Since the advancement of CNNs, clinicians now have a very powerful tool that improves diagnostic precision and therefore patient care. However, in this research despite a lot of progress was made,fracture detection deep learning models still encounter difficulties like overfitting, class imbalance and generalization issues.Nevertheless, with better regularization, better data augmentation and better feature extraction,these models can greatly improve accuracy, differentiating between fractured and non fractured cases and improved clinical reliability.The model performance can be further refined by also fine tuning the pretrained architecture and the class balancing strategies, lowering the false positives, increasing the model stability.There are developing deep learning techniques that could provide more powerful, flexible and precise automated fracture detection systems to assist doctors in medical diagnosis with powerful deep learning techniques.Although CNN models continue to improve and merge with cutting edge technologies like real time diagnostic tools as well as transfer learning, they are set to improve model performance across variable patient populations.Current techniques such as multi-scale feature extraction, attention mechanisms and semi-supervised learning are tackling the challenge arising from the limited dataset and over-model generalizability.As such, the most significant component of bringing such models into a healthcare context is the creation of APIs that can be embedded in clinical workflows.Whereas significant efforts were implemented to automate the process of detecting fractures with the help of CNN models, we claim that this type of knowledge ought to serve as the supplemental, at best, to the priceless patient care expertise and assessment exercised by clinicians.

Bibliography

1. Adigun, O., Babatunde, R., Rufai, M., & Edwin, A. (2020). Detection of fracture bones in X-ray images categorization. *Journal of Advances in Mathematics and Computer Science, 35*(4), 1–11. https://doi.org/10.9734/JAMCS/2020/v35i430265

2. Cheng, L.-W., Chou, H.-H., Cai, Y.-X., Huang, K.-Y., Hsieh, C.-C., Chu, P.-L., & Hsieh, S.-Y. (2024). Automated detection of vertebral fractures from X-ray images: A novel machine learning model and survey of the field. *Neurocomputing, 566*, 126946. https://doi.org/10.1016/j.neucom.2023.126946

3. Choi, J. W., Cho, Y. J., Lee, S., Lee, J., Lee, S., Choi, Y. H., & Kim, N. (2020). Using a dual-input convolutional neural network for automated detection of pediatric supracondylar fracture on conventional radiography. *Investigative Radiology, 55*(2), 101–110. https://doi.org/10.1097/RLI.0000000000000621

4. Guan, Q., Huang, Y., Zhong, Z., Zheng, Z., Zheng, L., & Yang, Y. (2018). Diagnose like a radiologist: Attention guided convolutional neural network for thorax disease classification. *arXiv.* https://arxiv.org/abs/1801.09927

5. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90

6. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4700–4708). https://doi.org/10.1109/CVPR.2017.243

7. Medaramatla, S. C., Samhitha, C. V., Pande, S. D., & Vinta, S. R. (2024). Detection of hand bone fractures in X-ray images using hybrid YOLO NAS. *IEEE Access.* https://doi.org/10.1109/ACCESS.2024.3379760

8. Sharma, S. (2023). Artificial intelligence for fracture diagnosis in orthopedic X-rays: Current developments and future potential. *SICOT-J, 9*(21), 1–23. https://doi.org/10.1051/sicotj/2023018

9. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 618–626).

10. Genant, H. K., Wu, C. Y., van Kuijk, C., & Nevitt, M. C. (2018). Vertebral fracture assessment using a semiquantitative technique. *Journal of Bone and Mineral Research, 33*(1), 20–34. https://doi.org/10.1002/jbmr.3336

11. Zhang, Y., Wu, S., & Li, Z. (2024). Evaluation of multi-model approaches in bone fracture detection from X-ray images. *IEEE Transactions on Medical Imaging, 43*(5), 1023–1035.

12. Wang, L., Zhao, Q., & Zhang, M. (2023). Enhancing fracture detection with data augmentation and transfer learning techniques. *Medical AI, 14*(2), 79–92.

13. Chen, H., Li, Y., & Zhang, X. (2023). Few-shot learning for bone fracture detection using limited datasets. *Neural Networks, 72*(4), 123–135.

14. Çallı, E., Sogancıoğlu, E., van Ginneken, B., van Leeuwen, K. G., & Murphy, K. (2021). Deep learning for chest X-ray analysis: A survey. *Medical Image Analysis, 72*, 102125.

15. Martins, C. R. (n.d.). Building Your First Convolution Neural Network — AlexNet. https://towardsdatascience.com/building-your-first-convolution-neural-network-alexnet-6b8f9aa8d30d

16. Yuan, J. (2016). Feature extraction and image retrieval based on AlexNet. *Proceedings of SPIE*, August 2016.

17. Litjens, G. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*, 60–88. https://www.sciencedirect.com/science/article/pii/S1361841516301839

18. Guan, Q., Zhang, X., & Li, J. (2022). Improving fracture detection using multi-modal deep learning models. *Journal of Medical Imaging, 45*(6), 1332–1345.

19. Keras Team. (n.d.). Keras documentation: Conv2D layer.

20. Tripathi, M. (2021). Image processing using CNN: A beginner's guide. Accessed: 2024-01-24.

21. Tanzi, L., Bonetto, L., & Zavanone, M. (2020). A survey of artificial intelligence techniques for X-ray image analysis. *International Journal of Computer Vision, 128*, 1–25. https://doi.org/10.1007/s11263-020-01356-8

22. Vrbancic, G., Pečnik, Š., & Podgorelec, V. (2020). Identification of COVID-19 X-ray images using CNN with optimized tuning of transfer learning. In *2020 International Conference on Innovations in Intelligent Systems and Applications (INISTA)*, pp. 1–8, Novi Sad, Serbia.

23. Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., & Oermann, E. K. (2018). Confounding variables can degrade generalization performance of radiological deep learning models. *arXiv*. https://arxiv.org/abs/1807.00431

24. Towards Data Science. (2025, March 5). The Annotated ResNet-50. *Towards Data Science*. Retrieved June 18, 2025, from https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758

25. Swhl. (2021, January 15). Review—Grad-CAM++: Improved visual explanations for deep convolutional networks (weakly supervised). *Medium*. Retrieved June 18, 2025, from https://medium.com/swhl/review-grad-cam-improved-visual-explanations-for-deep-convolutional-networks-weakly-760264e66bc6/

26. Bone fracture Multi-Region X-ray data. (2024, April 23). Kaggle. https://www.kaggle.com/datasets/bmadushanirodrigo/fracture-multi-region-x-ray-data/data