Preprints.org

Article

Enhancing Finger Fracture Diagnosis: A Deep Learning Approach Using ResNet and VGG

Shinawar Naeem , Asma Naseer , Shafiq ur Rehman * , Volker Gruhn , Sheeraz Akram

Article

Enhancing Finger Fracture Diagnosis: A Deep Learning Approach Using ResNet and VGG

Shinawar Naeem [1], Asma Naseer [1], Shafiq ur Rehman [2,*], Volker Gruhn [3] and Sheraz Akram [2]

[1] Department of Computer Science, FAST School of Computing, National University of Computer and Emerging Sciences, Lahore, Pakistan; l228183@lhr.nu.edu.pk (S.N.); asma.naseer@nu.edu.pk (A.N.)

[2] College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Kingdom of Saudi Arabia; sakram@imamu.edu.sa

[3] Department of Software Engineering, Universität Duisburg-Essen, Essen Germany; volker.gruhn@uni-due.de

* Correspondence: srehman@imamu.edu.sa

Abstract: Our daily activities hinge on the flexibility of our fingers, and a fractured finger can significantly disrupt these routines. The finger bones enable us to bend and fold the fingers and thumb to pick up or grasp objects and do all of our daily activities. A broken finger can cause adverse effects on our daily life activities. It is important to treat broken finger as soon as possible. Swift and precise treatment begins with capturing multiple X-rays, followed by the critical step of fracture detection in these images. Relying on the naked eye for this task carries the risk of overlooking small fractures. To address this issue an automated diagnoses of fractured fingers from images is required for which the current research employs advanced deep learning models—ResNet34, ResNet50, ResNet101, ResNet152, VGG-16, and VGG-19—to classify finger images from the Musculoskeletal Radiographs (MURA) dataset as either fractured or non-fractured. The results emphasize the consistently strong performance of ResNet models, attaining an impressive accuracy of 81.9%. This surpasses VGG models by 3.4% and establishes ResNet as the new benchmark for state-of-the-art accuracy.

Keywords: classification; transfer learning; deep learning; finger fracture; X-Rays

---

1. Introduction

Fingers are vital for doing daily chores. One of the most often used and most sensitive appendages, fingers are vulnerable to injury. Anyone could become slower if they suffered from finger injury. Phalanges are the name for the bones in your fingers. All fingers have three phalanges and The thumb consists of two phalanges, and the bones in the fingers are interconnected and held together by ligaments. Tendons link bones and muscles.The muscles in the forearm that exert force on the tendons of the fingers play a role in controlling the movement of the fingers. The distance between the finger bones and the rest of the body is what distinguishes them. The proximal phalanx is the bone that lies nearest to the palm. The middle phalanx is the next bone after the proximal phalange. The distal phalanx is the last bone, the smallest and the furthest from the hand. There is no middle phalanx on the thumb. When one or more of these bone breaks, it referred as broken or fractured finger. Your knuckles, which are the joints where your finger bones converge, are another place where fractures can happen. You're more likely to shatter a finger if you have diseases like osteoporosis and malnutrition. A fracture might occasionally happen because of atypical bone in the finger. A condition that weakens the bone and makes it more susceptible to fracture results in this form of fracture, known as a pathologic fracture. It happens frequently for someone to believe they have sprained their finger when in fact they have a fracture that needs to be treated surgically. A delayed diagnosis can result in a worse long-term outcome if an injury is treated as a sprain when it is actually a fracture that needs specialist care. This frequently refers to restricted mobility or a chronically deformed finger. Because of this, any suspicious finger injury should be examined by a doctor to see if an X-ray is necessary. Your doctor will likely order an X-ray if you exhibit signs of a fractured finger to see if there is a fracture.

If the diagnosis is uncertain, it might be essential to take additional X-rays in various orientations because not all fractures will show up clearly on a single image. To determine the best course of therapy, the wounded finger should be examined. If the digit has been rotated or shortened as a result of the injury, this might be discovered by examination. You'll discover that these are significant guiding variables for treatment. To be sure there is no tendon damage or other injury that could change the recommended course of treatment, examination can be helpful.When X-ray is done, finding fractures with the naked eye is a laborious technique that carries a larger chance of missing small fractures. Therefore, the goal is to design a finger fracture classification system that can be used regularly in any busy orthopaedic clinic to detect finger fractures and to make decision taking easy on a course of treatment, without formal criticism. In this project we were aimed to classify and detect finger fractures using deep learning architecture. We use the dataset called MURA, and images were preprocessed using resizing and data augmentation to address data imbalance issues. Multiple deep learning models such as Resnet34, ResNet50, ResNet101, ResNet152, VGG-16 and VGG-19 are trained and tested for classification and detection of finger fractures. Standard machine learning techniques are used to optimize the model's performance. The models are assessed using metrics including F1 score, recall, accuracy, and precision. Lastly, a test dataset is used to evaluate the generalization capacity of the top-performing model. The outcomes are examined in order to make judgments regarding the viability of the suggested strategy for classification and detection of finger fractures.

Major Research Contributions:

The major research contributions of this study are given as below:

- **Automated Fracture Diagnosis:** The research introduces an automated approach to diagnose fractured fingers from images, mitigating risks associated with manual examination.
- **Advanced Deep Learning Models:** The study employs cutting-edge models, including ResNet34, ResNet50, ResNet101, ResNet152, VGG-16, and VGG-19, showcasing a commitment to leveraging state-of-the-art technologies for medical image analysis.
- **Comprehensive Evaluation Metrics:** The research pioneers the use of accuracy, precision, recall, and F1-score as thorough evaluation metrics, contributing to the methodology of assessing deep learning models in the medical domain.
- **Performance Comparison:** The findings reveal the consistent and robust performance of ResNet models, with a commendable test accuracy of up to 81.9%, surpassing VGG models (78.5%) in the context of finger fracture classification.
- **Insights for Future Research:** The study provides valuable insights for future research in the field of medical image analysis, particularly in optimizing deep learning models for enhanced accuracy and efficiency in diagnosing finger fractures.

The rest of the article is organised in such a way that Section 2 provides an over view of the existing research in the field of finger fracture classification. Methodology of the current study is elaborated in Sections 3 while Section 4 provides statistics regarding the results of multiple experiments. Conclusion and Future Work are provided in Section 5.

2. Literature Review

In this study, the researcher employed a multi-network model and utilized the "MURA" dataset, achieving an accuracy of 90.77%. The model comprised three sub-networks designed to identify distinct categories of abnormalities in radiographs. These sub-networks were dedicated to detecting abnormalities related to bones, joints, and implants, respectively.[1]. In this study, By using the BoneNet model and a dataset of bone X-rays, the author was able to achieve an astounding accuracy of 98.6%. The model proved adept at identifying subtle distinctions among closely related bone categories, surpassing the capabilities of human observation. Notably, the model showcased resilience in the face of changes in bone positioning, angles, and image quality, highlighting its consistent accuracy across diverse conditions. [2]. In this research, the researcher utilized a publicly accessible dataset comprising 3000 femur X-ray images, achieving an accuracy of 93.8% through the implementation of

the Vision Transformation (ViT) model. Scientists utilizing the ImageNet dataset's pre-trained weights. The training process involved a dataset consisting of approximately 3,000 femoral X-ray images. The model strategically prioritizes regions around the fracture site and areas with distinctive bone texture, as per the authors' analysis of the relevance of different image regions in the classification of femur fractures. [3]. The authors of this article personally collected the X-ray images. from 1280 patients in an Indian hospital, utilizing a CNN to achieve an impressive accuracy of 98.31%. Through ablation research, the authors discerned the crucial role of convolutional layers in learning valuable features for fracture detection. This experimentation involved analyzing the contributions of different components within the model. [4]. In this study, the author employed CNN-based models, denoted as E1 and E2, incorporating transfer learning on the "MURA" dataset, attaining, respectively, test accuracies of 0.8455 and 0.8472. The model's accuracy and robustness were exemplified by its proficient classification of various shoulder X-ray abnormalities, including acromioclavicular joint separation, rotator cuff tears, proximal humerus fractures, and glenohumeral joint osteoarthritis. The research underscores that the model excels when transfer learning is applied, using the ImageNet dataset's pre-trained weights. This method allows the model to understand important features essential for the categorization of shoulder X-ray images. [5].

In this study, the author delves into various strategies for bone fracture classification, with a focus on X-ray images as the primary modality. Emphasizing the significance of data preprocessing and augmentation techniques, the article highlights their role in enhancing the performance of classification models.Deep learning techniques, especially Convolutional Neural Networks (CNNs), have become the standard method in recent years. The author underscores the necessity for standardized evaluation metrics and benchmark datasets to facilitate a fair comparison of diverse methods. While acknowledging the potential applications of bone fracture classification in clinical settings, the article acknowledges existing challenges and limitations. These include the demand for substantial amounts of annotated data and the complexity of detecting subtle or intricate fractures. To address these issues, the paper proposes exploring the utilization of multimodal imaging data and incorporating explainable AI techniques, aiming to enhance both the accuracy and interpretability of classification models.[6]. This paper highlights how deep learning can improve tasks in musculoskeletal radiology, including bone fracture detection, disease diagnosis, and joint segmentation. Data preprocessing and augmentation are crucial for model performance. Deep learning enhances diagnostic accuracy, reduces variability, and automates patient screening. Lack of imaging protocol standardization is a challenge. Future research could explore multiple imaging modalities, integrating clinical and imaging data, and personalized medicine in musculoskeletal radiology using deep learning. [7]. In order to investigate the application of deep learning models for the detection and classification of fractures using diverse imaging modalities, this paper performs a comprehensive review and meta-analysis of 14 studies. The combined sensitivity and specificity of these models are reported as 0.91/0.95 for fracture detection and 0.92/0.94 for fracture classification, respectively. The accuracy of these deep learning models varies depending on the type of fracture and the imaging technique employed. The potential clinical applications include enhancing diagnostic precision and minimizing variations between different observers. However, to confirm that deep learning models for orthopedic fractures accurately diagnose patients, more standardized evaluation metrics and larger studies are essential. [8]. In this article, the author proposes the utilization of a deep learning model built upon a CNN architecture. The model is specifically trained for the classification of hip fractures into stable or unstable categories and the prediction of functional outcomes based on preoperative radiographs. The dataset employed in the study comprises 374 hip fracture cases from a single institution. The model demonstrates an overall accuracy of 87.2% for fracture identification and 71.4% for functional classification.[9]. The author of this work presents an attention-based cascade R-CNN model intended for X-ray image detection of sternum fractures. The model operates in two stages, focusing on Region of Interest (ROI) detection and subsequent classification. Using a dataset comprising 380 X-ray images, the authors attained noteworthy results, including an F1-score of 0.905 and a precision of 0.947 on the test set.

These results outperform those of other cutting-edge models. The findings of this research suggest promising potential for enhanced accuracy and efficiency in sternum fracture detection within clinical practice.[10].

This study recommends using a deep learning model for the identification and categorization of bone abnormalities that is based on the VGG-16 architecture in radiographic images, employing the MURA dataset. High accuracy rates for binary and multi-class classification tasks are reported by the study, distinguishing between normal and abnormal images. These findings underscore the model's potential utility in aiding the detection and diagnosis of bone abnormalities. However, the paper acknowledges certain limitations, such as the necessity for larger and more diverse datasets, and the exploration of alternative deep learning techniques.[11]. In this study, With a mean average precision of 85.6%, the author argues in favor of using the YOLOv5 model to automatically identify bone fractures in X-ray images. This proposed approach holds promise for enhancing clinical outcomes through the accurate and efficient identification of bone fractures in X-ray data. The authors underscore the significance of automated fracture detection systems, particularly in regions with high incidence rates of bone fractures, where manual interpretation of X-ray images may be prone to errors. [12]. In this article, In order to identify arm fractures in X-rays, the author proposes an improved deep CNN that is derived from the R-CNN model. The proposed method surpasses other state-of-the-art deep learning approaches, attaining a notable average precision of 62.04%. The authors attribute this achievement to various enhancements, including the implementation of a new backbone network, image preprocessing techniques for contrast enhancement, and adjustments to the receptive field. The research showcases the potential practical applicability of the proposed method in clinical settings, aiming to enhance the efficiency and accuracy of detection of arm fracture in X-rays.[13]. In this research, the author proposes the exploration of deep learning models for various musculoskeletal disease-related tasks, including lesion detection, progression prediction, and bone age assessment. The results indicate the potential for these models to achieve high accuracy, offering possibilities for improved clinical outcomes through early detection and diagnosis, prediction of disease progression, and more precise assessments of bone age. However, the author emphasizes the necessity for additional validation and testing to assess the models' applicability and reliability in real-world clinical scenarios. [14]. In this study, the The author suggests and illustrates the use of CNN-based models to identify cartilage lesions in knee MR images. The model underwent training and evaluation using a dataset from the Osteoarthritis Initiative (OAI) database, surpassing the diagnostic performance of radiologists with an AUC-ROC of 0.92. This suggests that deep learning methods can serve as valuable tools to assist clinicians in accurately detecting musculoskeletal abnormalities, potentially enhancing patient outcomes. But it's important to remember that validation of the model's performance is necessary. on larger and more diverse datasets before considering its application in clinical practice. [15]. Deep learning (DL) has emerged as a game-changing technology for medical applications in recent years, bringing about significant changes in diagnostic and treatment approaches. With its advanced neural network architectures, deep learning has particularly excelled in tasks like medical imaging for improved image segmentation and disease classification. This study specifically focuses on the Alzheimer's disease segmentation and classification through Magnetic Resonance Imaging (MRI). The innovative approach integrates both transfer learning and customization of Convolutional Neural Networks (CNNs). The methodology involves the utilization of pre-segmented brain images, with a specific focus on the Gray Matter. The researchers use a pre-trained DL model as the initial framework and then apply transfer learning, rather than creating the model from scratch. Through assessing the model's accuracy at various epochs (10, 25, and 50), the study achieves an outstanding overall accuracy of 97.84%. This demonstrates how their method of applying state-of-the-art deep learning techniques to advance the analysis of Alzheimer's disease is effective [16]. In this study, a computer-aided diagnosis (CAD) system is developed using images from chest X-rays (CXRs) tailored for a specific disease. The method entails training two deep learning networks: the Long Short-Term Memory Network (LSTM) and the Convolutional Neural Network (CNN), in three phases. Initially, the CNN undergoes training

with raw CXR images, subsequently pre-processed images, and lastly, using augmentation techniques, improved CXR images. The final CNN-LSTM model achieves an impressive accuracy of 99.02%, surpassing benchmark models. Notably, this approach enhances true positive rates, addressing the issue of false negatives encountered when using raw CXR images [17]. Predicting road traffic is vital for intelligent transportation systems, yet it presents challenges given diverse roads, speed fluctuations, and interdependencies among segments. In order to handle dynamic spatial dependencies, this work incorporates attention mechanisms into the Graph WaveNet model. When evaluated against the Graph WaveNet for a 60-minute prediction, On the PEMS-BAY and METR-LA datasets, the improved model showed a reduction in root-mean-square error of 3.4% and 4.76%, respectively.It is noteworthy, though, that the additional computation of attention scores resulted in an increase in the model's training time[18]. Another study underscores the critical importance of early detection in Acute Pancreatitis (AP). This research leverages advanced machine learning techniques to replace traditional scoring methods, overcoming challenges such as small datasets and class imbalance. The incorporation of augmented datasets from sources like MIMIC-III, MIMIC-IV, and eICU enhances the training of the classifier. Effective handling of missing values is achieved through iterative imputation. The study compares the performance of downsampling on small test sets, cautioning about its potential for being misleading on larger sets. Upsampling techniques, including CTGAN, TGAN, CopulaGAN, CTAB, TVAE, and SMOTE, were explored. Among these, Random Forest demonstrated exceptional performance with an F-Beta of 0.702 and a recall of 0.833 in a 50-50 class split by CTGAN. Random Forest also exhibited strong performance on the TVAE dataset with an F-Beta of 0.698. In the case of SMOTE-based upsampling, the Deep Neural Network emerged as the top performer, achieving the best result with an F-Beta of 0.671. [19].

3. Methodology

The steps of the methodology we applied in this study are as shown in Figure 1 take X-ray images of finger from MURA dataset, data preprocessing and data augmentation for data imbalance of classes, apply transfer learning based deep learning models, show the results in class 0 and class 1, where class 0 is negative and class 1 is positive.
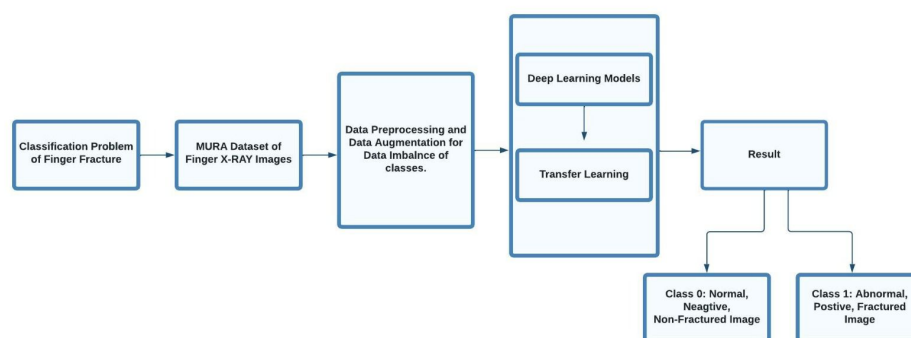

Figure 1. Architecture Diagram of Classification and Detection of Finger Fracture by Using Deep Learning.

3.1. Date Set

The MURA dataset, among the most extensive openly accessible radiography datasets, includes X-rays showing the finger, elbow, wrist, hand, forearm, humerus, and shoulder bones, among other skeletal regions. For this study, the primary focus is on utilizing finger bone X-ray images, specifically those presented in Table 1. The MURA dataset includes a total of 2142 finger images, with 1389 identified as normal and 753 as abnormal. Radiologists themselves have labeled these images based on their assessment of normalcy or abnormality.

Table 1. Total Images in Finger Class.

Total Images	Normal	Abnormal
2142	1389	753

3.2. Data Augmentation

In this study, the MURA dataset for finger bone X-ray images was augmented to amplify the dataset's size. This practice is crucial in deep learning, especially for classification studies, as the variety and volume of data the model is trained on have a significant impact on its performance. The augmentation technique involved creating additional images by rotating the original X-ray photographs, with rotations ranging up to a maximum of 10 degrees to both the right and left. The objective of this procedure is to enhance the network's training phase by incorporating a larger collection of images, positively impacting the model's ability to learn and generalize effectively.

3.3. Data Preprocessing

The study examined X-ray images that exhibit noise and a dark background, which could potentially hinder the precision of the fracture detection and classification procedures. To address these issues and improve the precision of classification results, Several pre-processing procedures were used on the dataset. Following are the preprocessing steps:

1. Detection of the Corresponding Area: A significant portion of the X-ray images in the dataset lacked sufficient size-related semantic information. To address this issue, a series of preprocessing steps were implemented. Initially, the images were converted to grayscale and then double thresholded. Using Otsu's thresholding value approach, an adaptive threshold value was found, aiming to rectify this inadequacy. This involved assessing within-class variance for the foreground and background color classes in the grayscale pictures. The threshold value chosen was the one that minimized this variance, utilizing Otsu's thresholding value method. Following this, the thresholded image underwent edge detection techniques to identify and highlight its edges. After this, the original images underwent a cropping process to fine-tune and boost their overall quality.
2. CLAHE Transformation: In this step, the transformation was achieved by employing contrast-limited adaptive histogram equalization (CLAHE) with the OpenCV library. The input image was divided into segments, each having its own histogram. Users defined limits for histogram cropping, and based on these specifications, adjustments were applied to the histogram of each segment, resulting in the creation of CLAHE-enhanced versions of the images.
3. Normalization and Standardization: In this final phase, ImageNet values were employed for the normalization and standardization of the images.

3.4. Models

3.4.1. ResNet-34

It starts by loading a pre-trained ResNet-34 model, which has undergone training on an extensive dataset.. The code then modifies the fully connected (fc) layer of the model. The original fc layer is replaced with a new one that consists of a linear layer and a sigmoid activation function. This modification enables the model to undergo fine-tuning specifically for a binary classification task, where it predicts one of two classes. By replacing the fc layer, the network can be customized to suit the specific classification problem at hand. The utilization of the sigmoid activation function guarantees that the model produces output probabilities within the range of 0 to 1, which is commonly used in binary classification tasks. The equations for original [1] and combined (original+modified) [2] are given below: Where,

- W' is the weight matrix for the new linear layer.
- b' is the bias vector for the new linear layer.
- x is the input.
- σ is the sigmoid activation function.
- $f_{\text{ResNet-34}}(x)$ represents the feature extraction process of the ResNet-34 model.

$$y_{\text{original, ResNet-34}}(x) = f_{\text{ResNet-34}}(x) \tag{1}$$

$$y_{\text{combined, ResNet-34}}(x) = \sigma(W' f_{\text{ResNet-34}}(x) + b') \tag{2}$$

In the following diagrams, the Figure 2 meticulously captures the architecture of the original model, depicting its intricate layers and connections. The Figure 3 serves as a visual representation of the modified model, highlighting specific changes made to the fully connected (FC) layer. These alterations are instrumental in adapting the model for binary classification tasks, offering a clear comparison between the original and tailored architectures.
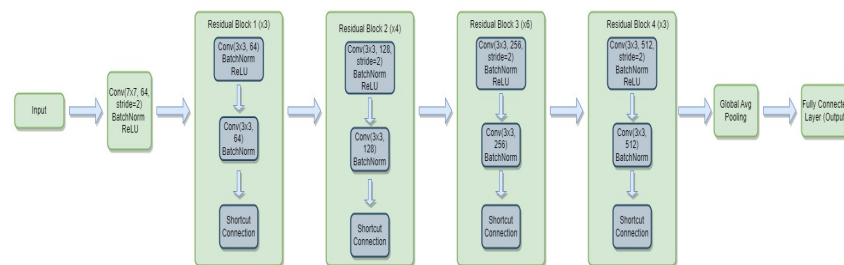

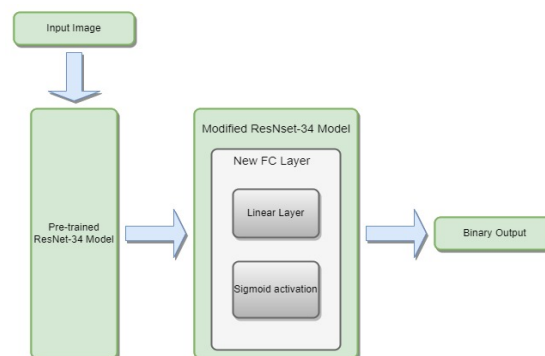**Figure 2.** Block Diagram of ResNet-34.

Figure 3. Block Diagram of Modifications in ResNet-34.

3.4.2. ResNet-50

The model is initialized with weights learned from an extensive dataset, and the subsequent step involves reconfiguring the model's fully connected (fc) layer. In this adjustment, the original fc layer is substituted with a new structure comprising a linear layer and a sigmoid activation function. The linear layer is tailored to accept a defined number of input features and generate 2 output features, representing the two classes in a binary classification task. Employing the sigmoid activation function ensures that the model's output is a probability falling within the standard 0 to 1 range, a widely adopted approach in binary classification scenarios. This code streamlines the fine-tuning of the ResNet-50 model for a specific binary classification task by tailoring the fc layer to suit the task's specific requirements. The equations for original [3] and combined (original+modified) [4] are given below: Where'

- W' is the weight matrix for the new linear layer.
- b' is the bias vector for the new linear layer.
- x is the input.
- σ is the sigmoid activation function.
- $f_{\text{ResNet-50}}(x)$ represents the feature extraction process of the ResNet-50 model.

$$y_{\text{original, ResNet-50}}(x) = f_{\text{ResNet-50}}(x) \tag{3}$$

$$y_{\text{combined, ResNet-50}}(x) = \sigma(W' f_{\text{ResNet-50}}(x) + b') \tag{4}$$

In Figure 4, the intricate architecture of the original model unfolds, revealing the layers and interconnected components. The Figure 5, in contrast, visually encapsulates the evolution of the model, highlighting targeted modifications made to the fully connected (FC) layer. These purposeful changes are instrumental in adapting the architecture for binary classification, providing a visual juxtaposition of the initial and customized designs.
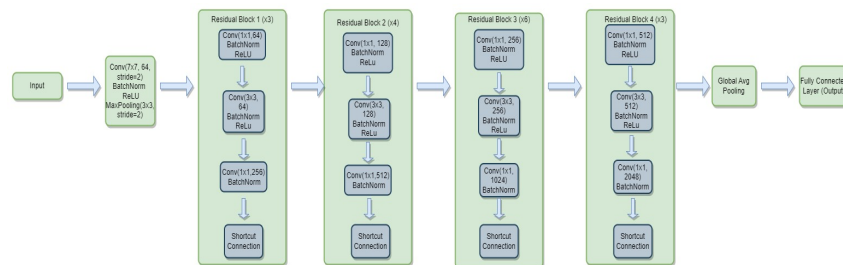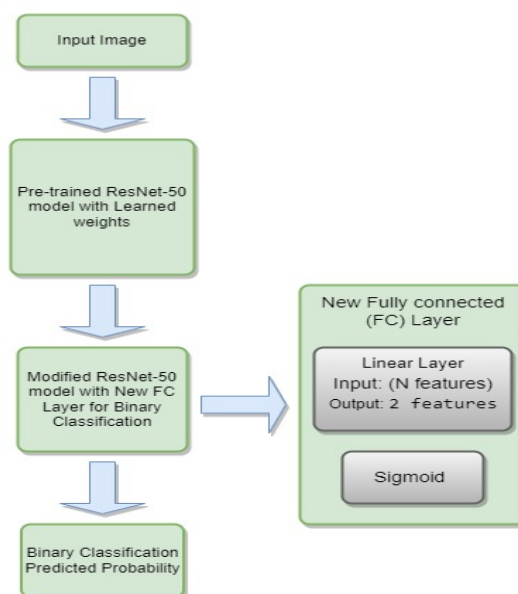

Figure 4. Block Diagram of ResNet-50.

Figure 5. Block Diagram of Modifications in ResNet-50.

3.4.3. ResNet-101

For image recognition, a pre-trained ResNet-101 model is utilized, benefiting from weights learned on a comprehensive dataset to adeptly recognize diverse objects and patterns in images. The subsequent focus in the code is on adapting the fully connected (fc) layer of the model, facilitating fine-tuning for a binary classification task. This entails replacing the original fc layer with a novel

configuration comprising a linear layer and a sigmoid activation function. The linear layer, tailored to a specific number of input features, yields 2 output features representing the two classes in question. The inclusion of the sigmoid activation function ensures that the model's output aligns within the probabilistic range of 0 to 1, rendering it well-suited for binary classification tasks. The equations for original [5] and combined (original+modified) [6] are given below: Where'

- W' is the weight matrix for the new linear layer.
- b' is the bias vector for the new linear layer.
- x is the input.
- σ is the sigmoid activation function.
- $f_{\text{ResNet-101}}(x)$ represents the feature extraction process of the ResNet-101 model.

$$y_{\text{original, ResNet-101}}(x) = f_{\text{ResNet-101}}(x) \tag{5}$$

$$y_{\text{combined, ResNet-101}}(x) = \sigma(W' f_{\text{ResNet-101}}(x) + b') \tag{6}$$

Figure 6 thoughtfully represent nuanced details of the original model's architecture, revealing the intricate interplay of its various layers. In contrast, the Figure 7 visually communicates the transformative journey of the model, emphasizing modifications to the fully connected (FC) layer. These purposeful adjustments are essential for customizing the architecture to meet the requirements of binary classification, providing a clear contrast between the original and adapted designs.
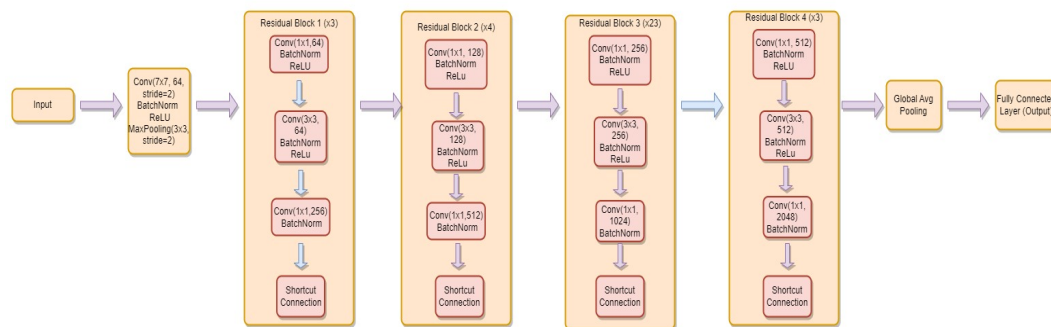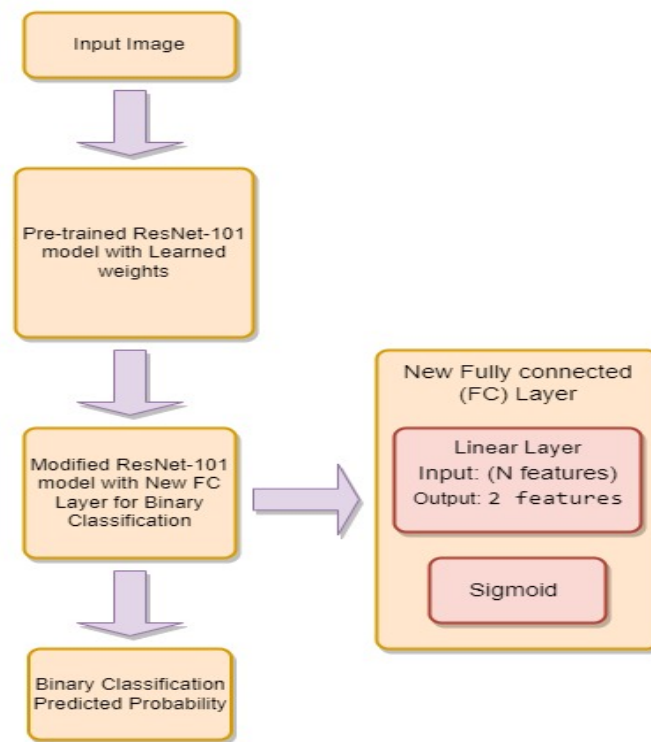

Figure 6. Block Diagram of ResNet-101

Figure 7. Block Diagram of Modifications in ResNet-101

3.4.4. ResNet-152

In image recognition, we leverage a pre-trained ResNet-152 model, renowned for its exceptional performance in visual tasks due to its deep convolutional neural network architecture. This model is initialized with pre-trained weights, indicating prior learning from an extensive dataset. Our study concentrates on customizing the fully connected (fc) layer of the ResNet-152 model, responsible for the final classification of input images. By replacing the original fc layer, the model can be tailored for a specific binary classification task. In this context, the new fc layer comprises a linear layer succeeded by a sigmoid activation function. The linear layer accommodates a specific number of input features, as determined by the original fc layer, and generates two output features representing the binary classification classes. The sigmoid activation function ensures the model's output is a probability value ranging between 0 and 1, aligning it with the requirements of binary classification. The equations for original [7] and combined (original+modified) [8] are given below: Where'

- W' is the weight matrix for the new linear layer.
- b' is the bias vector for the new linear layer.
- x is the input.
- σ is the sigmoid activation function.
- $f_{\text{ResNet-152}}(x)$ represents the feature extraction process of the ResNet-152 model.

$$y_{\text{original, ResNet-152}}(x) = f_{\text{ResNet-152}}(x) \tag{7}$$

$$y_{\text{combined, ResNet-101}}(x) = \sigma(W' f_{\text{ResNet-101}}(x) + b') \tag{8}$$

The figure [8] meticulously delineates the architectural intricacies of the original model, capturing the layered structure and intricate connections. Transitioning to the figure [9] , it visually encapsulates the model's transformation, emphasizing specific adjustments to the fully connected (FC) layer. These intentional modifications are pivotal for tailoring the model to excel in binary classification scenarios, offering a compelling visual comparison between the original and adapted architectures.
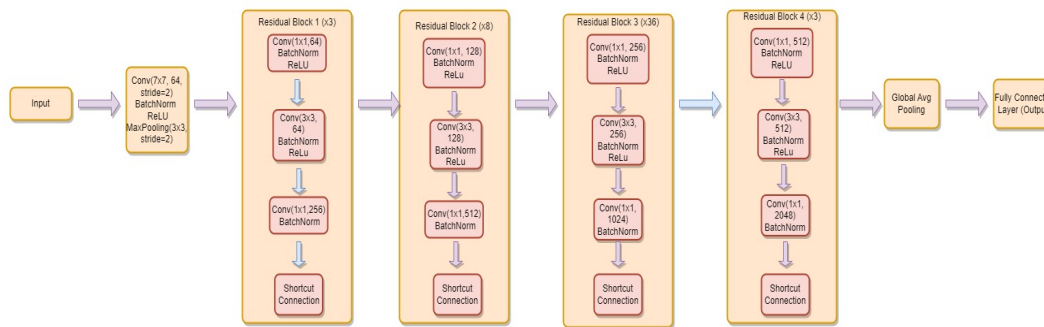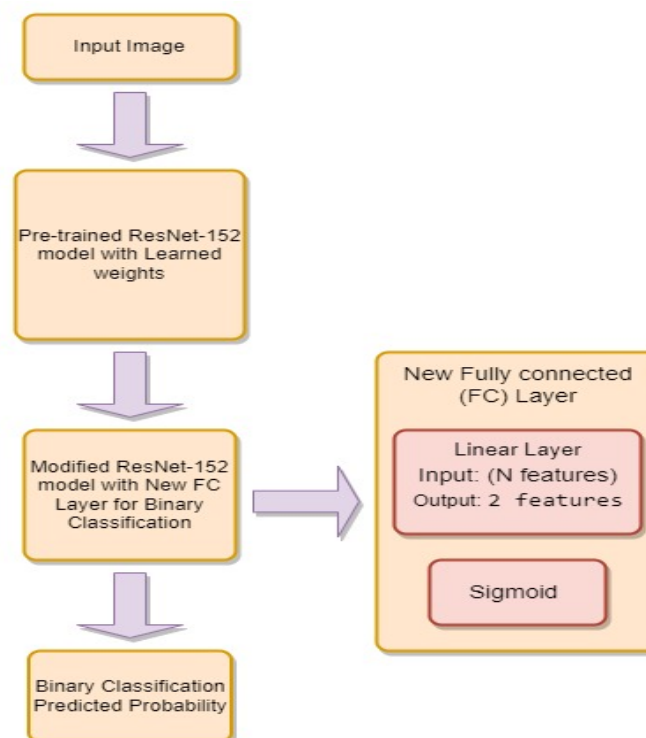
Figure 8. Block Diagram of ResNet-152.

Figure 9. Block Diagram of Modifications in ResNet-152.

3.4.5. VGG-16

Initialized with weights from a comprehensive dataset, the primary focus is on adapting the classifier layer of the VGG-16 model. This layer, pivotal for making final predictions based on extracted features, undergoes customization by replacing the last fully connected layer with a new linear layer. To determine the requisite number of input features for the new linear layer, the in-features attribute of the original fully connected layer is accessed. This attribute value is then employed to construct the new linear layer, specifically tailored for a binary classification task. This redesigned layer involves a linear transformation that takes the extracted features as input and produces two output features, representing the classes to be predicted. Through this classifier layer modification, the VGG-16 model is finely tuned for precise performance in a binary classification context, allowing it to learn and predict with tailored accuracy for the specific classes of interest. The equations for original [9] and combined (original+modified) [10] are given below: Where'

- W' is the weight matrix for the new linear layer.
- b' is the bias vector for the new linear layer.
- x is the input.

- σ is the sigmoid activation function.
- $f_{\text{VGG-16}}(x)$ represents the feature extraction process of the VGG-16 model.

$$y_{\text{original, VGG-16}}(x) = f_{\text{VGG-16}}(x) \tag{9}$$

$$y_{\text{combined, VGG-16}}(x) = \sigma(W' f_{\text{VGG-16}}(x) + b') \tag{10}$$

In the visual depictions given below, the Figure 10 intricately delineates the structure of the original model, portraying its complex layers and interconnections in detail. The Figure 11 functions as a graphical portrayal of the adapted model, emphasizing particular modifications introduced to the fully connected (FC) layer. These adjustments play a crucial role in customizing the model for binary classification assignments, providing a distinct contrast between the original and customized architectural configurations.
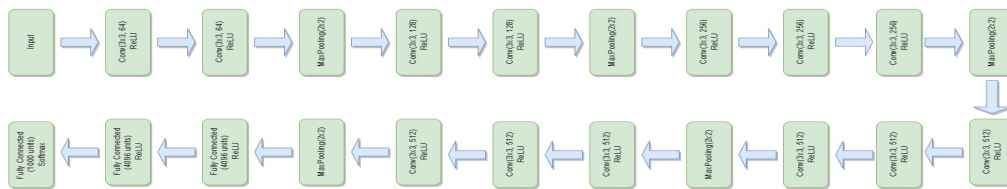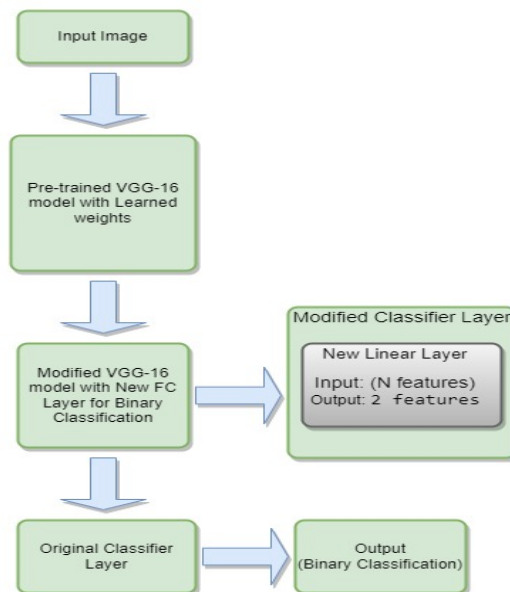
Figure 10. Block Diagram of VGG-16.

Figure 11. Block Diagram of Modifications in VGG-16.

3.4.6. VGG-19

Utilizing the VGG-19 model renowned for its prowess in visual tasks, this image classification task benefits from pre-trained weights on a comprehensive dataset. The core focus of the adaptation lies in the classifier layer, responsible for the conclusive classification step based on extracted features. Specifically, the final fully connected layer in the classifier undergoes transformation, being replaced with a novel linear layer. The determination of input features for this new linear layer is guided by accessing the in-features attribute of the original fully connected layer. Tailored for a binary classification task with two classes to predict, this new linear layer acts as the conclusive layer, mapping extracted features to the specified classes. Through this meticulous adjustment of the

classifier layer, the VGG-19 model is finely tuned for the nuances of a specific binary classification problem. This refinement allows the model to adeptly learn and tailor predictions to the classes of interest, elevating its performance in the designated task. The equations for original [11] and combined (original+modified) [12] are given below: Where'

- W' is the weight matrix for the new linear layer.
- b' is the bias vector for the new linear layer.
- x is the input.
- σ is the sigmoid activation function.
- $f_{\text{VGG-19}}(x)$ represents the feature extraction process of the VGG-19 model.

$$y_{\text{original, VGG-19}}(x) = f_{\text{VGG-19}}(x) \tag{11}$$

$$y_{\text{combined, VGG-19}}(x) = \sigma(W' f_{\text{VGG-19}}(x) + b') \tag{12}$$

The Figure 12 intricately delineates the nuanced architecture of the original model, elucidating its intricate layers and connections. The Figure 13 functions as a visual portrayal of the adjusted model, accentuating specific modifications made to the fully connected (FC) layer. These changes are pivotal in tailoring the model for binary classification tasks, presenting a distinct comparison between the initial and customized architectural configurations.
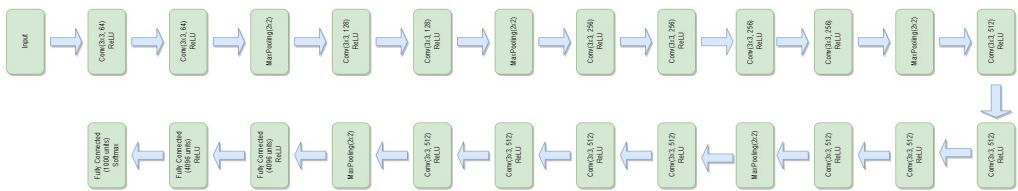

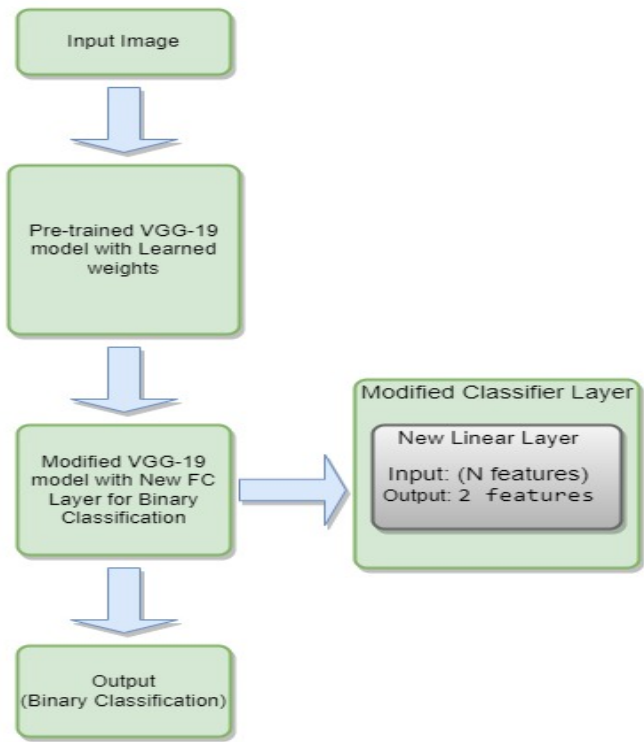**Figure 12.** Block Diagram of VGG-19.

Figure 13. Block Diagram of Modifications in VGG-19.

4. Results

The categorization of X-rays of the finger bones in this study was made easier by online servers serving as the hardware infrastructure. The ResNet and VGG deep learning models, alongside packages such as torchvision, seaborn, scikit-learn, matplotlib, and torch, were instrumental in executing the classification procedures. The cross-entropy loss function, the Adam optimizer, and a fixed learning rate of 0.0001 were employed in the procedure, spanning across 40 epochs. Notably, the initial learning rate was dynamic, undergoing a reduction by a factor of 10 every 10 epochs to enhance the learning success of the network.

4.1. Evaluation Metrics

In the domain of machine learning, the efficacy of classification models is commonly appraised using a suite of metrics derived from a confusion matrix. This matrix encompasses pivotal elements like True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). The training/testing accuracy acts as a thorough indicator of a model's accuracy, representing the proportion of correctly predicted cases to the total dataset. By highlighting the proportion of true positives to all positive predictions, precision provides insight into the accuracy of positive predictions. Recall, which is a measure of the model's ability to capture all real positive instances, is also referred to as sensitivity or the true positive rate. The F1-score provides a balanced metric that takes into account both false positives and false negatives. It is a harmonic mean of precision and recall. Collectively, these metrics provide varied perspectives on a classification model's performance, enabling practitioners to comprehensively evaluate its capabilities from diverse angles. By inputting specific values into these formulas derived from a confusion matrix, one can derive precise assessments of accuracy, precision, recall, and F1-score, customized to the distinct qualities of the model's predictions. Formulae of these evaluation metrics are the following:

1. **Training/Testing Accuracy:**

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

2. **Precision:**

$$\text{Precision} = \frac{TP}{TP + FP} \tag{14}$$

3. **Recall (Sensitivity or True Positive Rate):**

$$\text{Recall} = \frac{TP}{TP + FN} \tag{15}$$

4. **F1-score:**

$$\text{F1-score} = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{16}$$

4.2. Classification Results of Our Models

The outcomes of the classification, including training accuracy, test accuracy, precision, recall, F1-score, and Cohen's kappa scores for each model, are illustrated in Figure 14 and detailed in Table 2.

Table 2. Classification Results.

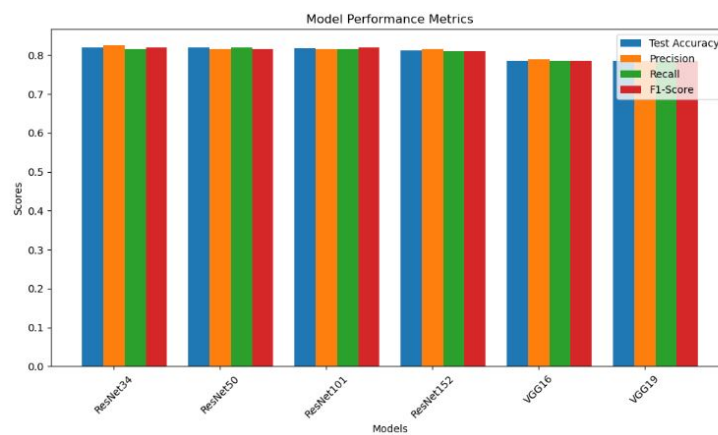| Models	Accuracy	Precision	Recall	F1-Score
ResNet34	0.8190	0.8246	0.8146	0.82
ResNet50	0.8188	0.8146	0.820	0.8149
ResNet101	0.8171	0.8152	0.8152	0.82
ResNet152	0.8117	0.815	0.81	0.81
VGG16	0.7851	0.79	0.7849	0.7849
VGG19	0.7851	0.7849	0.7849	0.7849

Figure 14. Results Visualization after testing.

The evaluation results indicate that the ResNet models, including ResNet34, ResNet50, ResNet101, and ResNet152, exhibit similar and strong performance across Test Accuracy, Precision, Recall, and F1-Score metrics. With accuracy ranging from approximately 81.1% to 81.9%, these models consistently achieve high precision (81.5% to 82.5%), recall (81.0% to 82.0%), and F1-Score (81.0% to 82.0%). In comparison, the VGG models (VGG16 and VGG19) demonstrate slightly lower but consistent performance, with accuracy, precision, recall, and F1-Score values around 78.5%. Overall, the ResNet models provide reliable and accurate results, making them strong choices for the given task, while the VGG models offer a computationally efficient alternative with slightly lower accuracy.

4.3. Discussion

The proposed study classifies finger bone X-ray images using deep learning models ResNet, VGG-16, and VGG-19. To thoroughly assess the models' performance, key evaluation metrics such as accuracy percentage, confusion matrix, precision, recall, and F1-score are employed. Notably, the ResNet models consistently demonstrate robust results, achieving accuracy levels ranging from 81.1% to 81.9%, accompanied by high precision, recall, and F1-Score values. In comparison, the VGG models exhibit slightly lower but consistently strong performance, with accuracy, precision, recall, and F1-Score hovering around 78.5%. The ResNet models are affirmed as reliable and accurate, while the VGG models present a computationally efficient alternative. Overall, the study highlights the commendable performance of ResNet models, encompassing ResNet34, ResNet50, ResNet101, and ResNet152, in the classification of X-ray images of finger bones. These models consistently deliver reliable and accurate results. On the other hand, VGG models, although exhibiting slightly lower accuracy, present a computationally efficient alternative for the same task. The decision between these Models should follow certain guidelines while taking trade-offs into account, between the pursuit of high accuracy and the consideration of computational efficiency.

In comparing the results of the present study with prior investigations on the MURA dataset, notable distinctions emerge, particularly concerning the scope of image classification. A study, which conducted binary classification on the entire MURA dataset utilizing VGG16 and RESNET50 architectures, yielded a commendable 78% accuracy [20] Subsequently, a study focused exclusively on finger classification achieved a higher accuracy of 81%, underscoring the significance of specialization within the dataset [21]. Furthermore, a study narrowed its scope to finger fracture classification, reporting accuracy of 71.4% for VGG19 and 70.1% for RESNET. Notably, this study highlighted the challenges associated with the nuanced task of distinguishing specific pathologies within the MURA dataset [22]. In contrast, a study demonstrated an 81% accuracy for the complete MURA dataset, emphasizing the robustness of models when confronted with a diverse range of musculoskeletal conditions [23]. The current investigation delved into finger fracture classification, yielding an accuracy of 78.5% for VGG models and 81.9% for RESNET models. The outcomes underscore the nuanced

nature of classifying finger fractures within the broader MURA dataset, demonstrating competitive accuracy rates in comparison to studies with a more extensive focus. The following Table 3 highlights each study's major conclusions:

Table 3. Comparative Analysis of Musculoskeletal Image Classification Studies on MURA Dataset.

Architecture	Dataset Scope	Accuracy
VGG-16	Whole MURA	78
ResNet50	Whole MURA	78
MSFMR50	Finger Fracture	81
VGG-19	Finger Fracture	71.4
ResNet	Finger Fracture	70.1
Naive DL network	Whole MURA	81
VGG(ours)	Finger Fracture	78.5
ResNet(ours)	Finger Fracture	81.9

5. Conclusion and Future Work

The ResNet architectures, in particular ResNet34, ResNet50, ResNet101, and ResNet152, outperform VGG models in terms of test accuracy, precision, recall, and F1-Score when evaluating deep learning models for finger fracture classification. ResNet models' reliable and powerful performance indicates their applicability for using X-ray images to automatically diagnose finger fractures. These findings can direct the choice of optimal models for practical implementation in computer aided diagnosis, highlighting the importance of model selection that prioritize both accuracy and reliability in healthcare applications. The research concludes the following findings:

- **ResNet Model Consistency:** ResNet models (ResNet34, ResNet50, ResNet101, ResNet152) exhibit consistent and robust performance across all evaluation metrics. The minimal variation in results indicates the reliability of ResNet architectures for the specific task of finger fracture classification.
- **High Accuracy:** All ResNet models achieve test accuracies above 81%, showcasing their effectiveness in distinguishing between fractured and non-fractured finger images. This high level of accuracy is crucial for reliable medical diagnoses.
- **Balanced Precision and Recall:** The ResNet models maintain a balance between precision and recall, with values ranging from 81.0% to 82.5%. This equilibrium shows how well the models can distinguish between positive (fractured) and negative (non-fractured) cases.
- **Comparable VGG Model Performance:** VGG16 and VGG19 models demonstrate comparable but slightly lower performance compared to ResNet models. The test accuracy of approximately 78.5% suggests a slightly reduced ability to correctly classify fractured and non-fractured finger images.

In future, the focus can be on improving the accuracy of finger fracture classification by fine-tuning ResNet architectures and employing auto-encoders, ensemble learning approaches, data augmentation, transfer learning with pre-trained Models and Continuous Model Monitoring and Updation.

References

1. Liang, S.; Gu, Y. Towards robust and accurate detection of abnormalities in musculoskeletal radiographs with a multi-network model. *Sensors* **2020**, *20*, 3153.
2. Pradhan, N.; Dhaka, V.S.; Chaudhary, H. Classification of human bones using deep convolutional neural network. In Proceedings of the IOP conference series: materials science and engineering. IOP Publishing, 2019, Vol. 594, p. 012024.

3. Tanzi, L.; Audisio, A.; Cirrincione, G.; Aprato, A.; Vezzetti, E. Vision transformer for femur fracture classification. *Injury* **2022**, *53*, 2625–2634.

4. Meena, T.; Roy, S. Bone fracture detection using deep supervised learning from radiological images: A paradigm shift. *Diagnostics* **2022**, *12*, 2420.

5. Uysal, F.; Hardalaç, F.; Peker, O.; Tolunay, T.; Tokgöz, N. Classification of shoulder x-ray images with deep learning ensemble models. *Applied Sciences* **2021**, *11*, 2723.

6. Karanam, S.R.; Srinivas, Y.; Chakravarty, S. A systematic review on approach and analysis of bone fracture classification. *Materials Today: Proceedings* **2021**.

7. Chea, P.; Mandell, J.C. Current applications and future directions of deep learning in musculoskeletal radiology. *Skeletal radiology* **2020**, *49*, 183–197.

8. Yang, S.; Yin, B.; Cao, W.; Feng, C.; Fan, G.; He, S. Diagnostic accuracy of deep learning in orthopaedic fractures: a systematic review and meta-analysis. *Clinical Radiology* **2020**, *75*, 713–e17.

9. Barhoom, A.; Al-Hiealy, M.R.J.; Abu-Naser, S.S. Bone Abnormalities Detection and Classification Using Deep Learning-Vgg16 Algorithm. *Journal of Theoretical and Applied Information Technology* **2022**, *100*, 6173–6184.

10. Krogue, J.D.; Cheng, K.V.; Hwang, K.M.; Toogood, P.; Meinberg, E.G.; Geiger, E.J.; Zaid, M.; McGill, K.C.; Patel, R.; Sohn, J.H.; et al. Automatic hip fracture identification and functional subclassification with deep learning. *Radiology: Artificial Intelligence* **2020**, *2*, e190023.

11. Jia, Y.; Wang, H.; Chen, W.; Wang, Y.; Yang, B. An attention-based cascade R-CNN model for sternum fracture detection in X-ray images. *CAAI Transactions on Intelligence Technology* **2022**, *7*, 658–670.

12. Sserubombwe, R. Automatic bone fracture detection in x-ray images using deep learning. PhD thesis, Makerere University, 2022.

13. Guan, B.; Zhang, G.; Yao, J.; Wang, X.; Wang, M. Arm fracture detection in X-rays based on improved deep convolutional neural network. *Computers & Electrical Engineering* **2020**, *81*, 106530.

14. Kijowski, R.; Liu, F.; Caliva, F.; Pedoia, V. Deep learning for lesion detection, progression, and prediction of musculoskeletal disease. *Journal of magnetic resonance imaging* **2020**, *52*, 1607–1619.

15. Liu, F.; Zhou, Z.; Samsonov, A.; Blankenbaker, D.; Larison, W.; Kanarek, A.; Lian, K.; Kambhampati, S.; Kijowski, R. Deep learning approach for evaluating knee MR images: achieving high diagnostic performance for cartilage lesion detection. *Radiology* **2018**, *289*, 160–169.

16. Raza, N.; Naseer, A.; Tamoor, M.; Zafar, K. Alzheimer Disease Classification through Transfer Learning Approach. *Diagnostics* **2023**, *13*, 801.

17. Naseer, A.; Tamoor, M.; Azhar, A. Computer-aided COVID-19 diagnosis and a comparison of deep learners using augmented CXRs. *Journal of X-ray Science and Technology* **2022**, *30*, 89–109.

18. Karim, S.; Mehmud, M.; Alamgir, Z.; Shahid, S. Dynamic Spatial Correlation in Graph WaveNet for Road Traffic Prediction. *Transportation Research Record* **2023**, p. 03611981221151024.

19. Hameed, M.A.B.; Alamgir, Z. Improving mortality prediction in Acute Pancreatitis by machine learning and data augmentation. *Computers in Biology and Medicine* **2022**, *150*, 106077.

20. Reddy, K.N.K.; Cutsuridis, V. Deep Convolutional Neural Networks with Transfer Learning for Bone Fracture Recognition using Small Exemplar Image Datasets. In Proceedings of the 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW). IEEE, 2023, pp. 1–5.

21. Kumar, K.; Pailla, B.; Tadepalli, K.; Roy, S. Robust MSFM Learning Network for Classification and Weakly Supervised Localization. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 2442–2451.

22. Kandel, I.; Castelli, M.; Popovič, A. Musculoskeletal images classification for detection of fractures using transfer learning. *Journal of imaging* **2020**, *6*, 127.

23. Singh, R.B.; Kumar, G.; Sultania, G.; Agashe, S.S.; Sinha, P.R.; Kang, C. Deep learning based MURA defect detection. *EAI Endorsed Transactions on Cloud Systems* **2019**, *5*, e6–e6.