AMERICAN UNIVERSITY OF BEIRUT

UTILIZING TRANSFER LEARNING FOR VERTEBRAL FRACTURES DETECTION

by
RIM RADWAN CHERRI

A thesis
submitted in partial fulfillment of the requirements
for the degree of Master of Engineering
to the Department of Electrical and Computer Engineering
of the Maroun Semaan Faculty of Engineering and Architecture
at the American University of Beirut

Beirut, Lebanon
April 2023

AMERICAN UNIVERSITY OF BEIRUT

UTILIZING TRANSFER LEARNING
FOR VERTEBRAL FRACTURES DETECTION

by
RIM RADWAN CHERRI

Approved by:

Signature Ali Chehab

Dr. Ali Chehab, Professor — Advisor
Department of Electrical and Computer Engineering

Signature

Dr. Riad Chedid, Professor — Member of Committee
Department of Electrical and Computer Engineering

Signature

Dr. Fatima Abu Salem, Associate Professor — Member of Committee
Department of Computer Science

Date of thesis defense: April 28, 2023

AMERICAN UNIVERSITY OF BEIRUT

THESIS RELEASE FORM

Student Name:	Cherri	Rim	Radwan
	Last	First	Middle

I authorize the American University of Beirut, to: (a) reproduce hard or electronic copies of my thesis; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes:

☐ As of the date of submission

☒ One year from the date of submission of my thesis.

☐ Two years from the date of submission of my thesis.

☐ Three years from the date of submission of my thesis.

RimCh May 10, 2023

Signature Date

ACKNOWLEDGEMENTS

First, I would like to express my sincere gratitude to my supervisor Professor Ali Chehab, who provided me with invaluable guidance and support throughout my thesis project. His expertise, insights, and constructive feedback helped me to refine my research questions, develop my methodology, and analyze my findings. I am also grateful for his/her unwavering encouragement, patience, and flexibility, which allowed me to overcome the many challenges that I encountered along the way.

I would also like to thank Prof. Riad Chedid and Prof. Fatima Abu Salem for agreeing to serve on my committee. Furthermore, I would like to thank my family and friends for their unwavering support, love, and encouragement throughout my thesis journey.

ABSTRACT
OF THE THESIS OF

Rim Radwan Cherri for Master of Engineering
Major: Electrical and Computer Engineering

Title: Utilizing Transfer Learning for Vertebral Fractures Detection

Spinal fractures are a prevalent type of fracture that have led to serious health issues which include difficulty in movement and permanent pain. Vertebral fractures are frequently present in most CT scans taken for abdominal health issues, and so they have rarely been diagnosed. Moreover, manual detection in medical images is time-consuming and requires specialized training. Thus, aiming for early and automated vertebral fracture detection is crucial for effective and fast treatment. Machine learning automated techniques can be utilized for fracture detection while deep learning models have proven their power in diagnosing different types of diseases. Specifically, transfer learning models have proven their effectiveness in diseases' detection from limited medical data benchmarks. For this reason, the project suggests the use of transfer learning models to effectively diagnose vertebral fractures from a CT scan dataset at AUBMC.

Five different deep architectures models (ResNet26, ResNet-RS-50, Inception_ResNet_v2, Swin_S3_Tiny, and ConvNeXT_Tiny_in22k) have been selected for investigation to diagnose fracture presence from the AUBMC CT scan dataset after passing through a series of pre-processing. Segmenting the vertebra along with classical augmentation were two important pre-processing steps that have improved the binary classification performance in all models. The results have shown that ConvNeXT_Tiny_in22k model has attained the highest testing accuracy of 81% without segmentation and 84.8% with segmentation upon selecting the same number of CT scan images. The dataset was then extended and segmented where the ConvNeXT model outperformed with a testing accuracy of 96.4%.

TABLE OF CONTENTS

ILLUSTRATIONS

Figure

TABLES

Table

CHAPTER 1

INTRODUCTION

1.1. Motivation

Vertebral fractures are a common type of fracture which have serious consequences, including chronic pain, spinal deformity, and reduced mobility. Early detection and accurate diagnosis of vertebral fractures are crucial to offer the most critical treatment and recovery possible and help prevent further long-term complications. However, detecting vertebral fractures from CT scans can be challenging and time-consuming for doctors and radiologists, particularly in cases where the fracture is subtle or difficult to distinguish from other structures in the image.

Machine learning techniques offer a promising solution to this problem. By exploiting machine learning algorithms through training models on large dataset of annotated CT scans, it is possible to develop a reliable automated system for detecting vertebral fractures. Such a system can provide faster and more consistent results compared to human experts' performance, enabling radiologists and doctors to provide more timely and informed decisions about patient care.

Overall, machine learning-based vertebral fracture detection from CT scans has the potential to improve vertebral fracture diagnosis, and enhance the efficiency of healthcare delivery, making it a motivating area of research.

1.2. Contribution

The main objective of this thesis is to present an intelligent system that automates radiologists' work by developing a data-driven vertebral fracture detection model. The proposed model exploits transfer learning to train a scarce collection of anonymized CT scans of various patients collected from AUBMC to detect fractures with promising results. CT scans will pass through data pre-processing before being fed to the classification model. Classification results will be compared upon using different pre-trained models, as well as upon changing the percentages of training, validation and testing datasets.

1.3. Structure of the Thesis

In the rest of this report, Chapter 2 summarizes the existing related work done in the literature. Chapter 3 demonstrates the methods and steps taken from pre-processing the dataset to model creation and testing. Chapter 4 presents an overview of a possible alternative approach, and Chapter 5 shows the classification results through different evaluation metrics.

CHAPTER 2

LITERATURE REVIEW

2.1. Transfer Learning in Medical Imaging

One approach in deep learning for the problem of scarce data applications is transfer learning. In simple terms, transfer learning is reusing pretrained machine learning models on new task. It improves generalization by leveraging the knowledge learnt from a certain problem and transferring it to the other. This type of learning have shown to be an optimization for progress especially for computer vision applications [1]. This section presents the work done using transfer learning in the field of classification tasks for different types of medical images.

2.1.1. CNN-based Transfer Learning

Most of the pre-trained Convolutional neural network (CNN) models mentioned in the literature for image classification purposes and visual object recognition research, such as resnet50 and inception V3 models, are initially developed and trained on the generic ImageNet dataset, which contains more than 14 million non-medical images. Although trained on non-medical images, pre-trained CNN models have shown their effective use for fracture detection [2]. In fact, one of the approaches for bone fracture detection was utilizing transfer learning by retraining the first layer of the inception V3 model on a dataset of 1389 radiographs, and it achieved a testing AUC of 0.95 [2].

Moreover, Loey *et al.* proposed a detection method based on transfer learning for detecting COVID-19 from a limited benchmark dataset of 742 chest CT scan images. The proposed method tested different pretrained deep convolutional neural network models (AlexNet, VGGNet16, VGGNet19, GoogleNet, and ResNet50) where ResNet50 have outperformed the others with a testing accuracy of 82.91%, sensitivity 77.66% and specificity of 97.62% using classical augmentation techniques [3]. A similar work in [4] was done where the authors have employed transfer learning models VGG16 and Xception to detect Corona Virus from X-ray images achieving an accuracy of 97.34% with F-score of 98%. Another work utilizes transfer learning in a hierarchical classification technique and demonstrated it for detecting glaucoma disease from a small tomography dataset where the weighted accuracy for a randomized test dataset scored 83.9% [5]. Similarly, Maqsood *et al.* employed transfer learning, specifically by fine-tuning AlexNet model to detect Alzheimer disease from MRI images which an overall accuracy of 92.85% [6]. Arooj *et al.* proposed a customized CNN-AlexNet model empowered with transfer learning to detect breast cancer from ultrasound images and histopathology images, and the model have achieved a maximum accuracy of 99.4%, 96.70% and 99.10% for respectively three different datasets [7]. Rashid *et al.* proposed a deep transfer learning model using MobileNetV2 model for Melanoma (a skin cancer disease) detection and obtained AUC of almost 98.2% using the official dataset of the Melanoma Classification Challenge also known as ISIC-2020 dataset [8]. The work in [9] entails fine-tuning of the classification head of various pre-trained models where DenseNet201 have outperformed to detect pneumonitis (an acute infection of the lungs) from chest X-ray images by achieving an AUROC of 0.96.

2.1.2. ViT-based Transfer Learning

Recently, vision transformers have an effective performance in the computer vision field. Like pretrained CNN models, pre-trained ViT models are utilized for transfer learning. In [10], the authors demonstrated a ViT-based transfer learning model for diagnosing breast cancer from mammographs and have achieved a AUC of 1. Furthermore, Okolo *et al.* developed an enhanced ViT for chest X-ray image classification tasks and have achieved an F1-score between 96% and 100% on a series of dataset examined. Their work also used transfer learning as parts of the ViT model were being initialized with ImageNet pre-trained weights [11]. Likewise, Zhang *et al.* fine-tunned the weights of a pretrained Swin ViT model to diagnose COVID-19 from chest CT scan images to gain an F1 score of 0.93 and 0.84 on the validation and test datasets respectively [12]. Moreover, the work in [13] presents a ViT-CNN ensemble model trained by fine-tuning, a transfer learning approach, for Leukemia diagnosis achieving an accuracy of 99.03%.

CHAPTER 3

METHODOLOGY

3.1. Dataset

The dataset is a collection of anonymized CT scans of different patients collected from AUBMC.

3.1.1. CT Scan definition

A CT scan is an abbreviation for computed tomography scan which is a medical imaging technique that uses multiple X-ray images from different angles and a computer combines them to produce a detailed cross-sectional image of the body [14].

3.1.2. AUBMC Dataset

The AUBMC dataset consists of 1567 folders where each folder belongs to a different accession. An accession identified by the accession number belongs to a single CT scan report, thus an accession might refer to the same or different patient. For each accession, AUBMC have extracted 15 2D images of the sagittal view and embedded the label in the folder's name, where the 5 middle images, including the central image and the 4 images that surround it (central), are the clearest for radiologists to diagnose fracture from. Each accession is labeled by a radiologist from AUBMC as one of 4 classes, either absent (no fracture), or present (3 classes of fracture: mild, moderate, or severe).

3.2. Data Pre-processing

A series of pre-processing steps were taken before training the model. These steps include the selection of data from the original AUBMC data, conversion of the CT scan images from their original extension, splitting and resizing them for normalization. To increase the dataset, data augmentation was applied which allows the model to generalize even better.

3.2.1. Selection, Conversion, Splitting, Resizing

The first approach was to select the central CT scan image of every patient and converting all images from DICOM to PNG. Then, categorize these selected PNG images into normal and fractured folders, after which they are split to evaluate the performance of the models tested. The splitting process was done as follows:

- The dataset is randomly split into three sets training, validation, and testing. The training set contains 80% of the data. This set will be used to train the machine learning model. The validation set contains 10% of the data. This set will be used to evaluate the performance of the model during training and to tune the model's hyperparameters. The testing set will contain 10% of the data. This set will be used to evaluate the final performance of the trained model on unseen data.

It's worth noting its critical to choose a split that provides enough data for training and testing while also allowing for a meaningful evaluation of the model's performance.

The second approach was to extend the window by taking extra 4 images for each patient which surround the central image from both sides and converting them to

PNG to have a total of 7835 Ct scan images. In other words, the approach is to select 2 images that are exactly before the central CT image and 2 images exactly after the central CT image for each patient. Then, images are classified into fracture and normal folders, and then split in the same manner as in the first approach.

3.2.2. Segmentation

Segmentation was applied after converting all CT scan images to PNG by isolating the vertebral column alone of each image as illustrated in figures 1 and 2 . Specifically, U-net neural network was trained on the modified Verse2019 and Verse2020 datasets. Verse2019 and Verse2020 datasets were modified by taking the sagittal view of the images to fit the format and shape of AUBMC dataset. After that, prediction using the U-net on the AUBMC dataset was tested. The output generated by U-net are masks which are then overlayed with the original images to give overlayed CT-scan images showing only the vertebra.

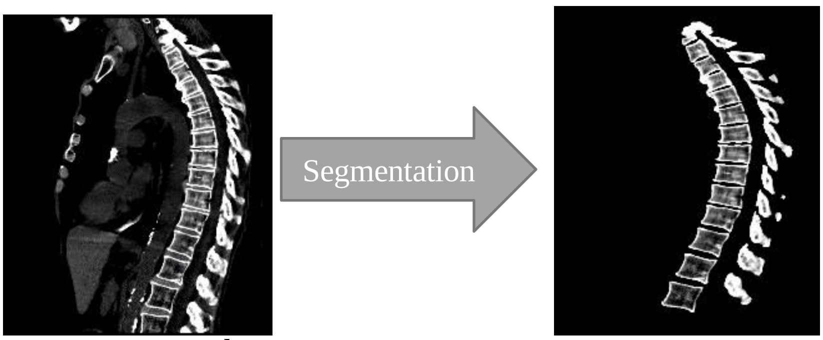

Figure 1: Original CT Scan

Figure 2: Segmented CT Scan

3.2.3. Data Transformation and Batch Augmentation in FastAi version

Data augmentation is perhaps the most crucial regularization technique for computer vision applications. Rather than repeatedly providing the model with identical images, data transformation is responsible to make minor random alterations that do not modify the image's content as perceived by the human eye but do alter its pixel values. By applying image augmentation, the model is more robust in classifying CT images [15]. Batch augmentation is applied on the AUBMC dataset (after being resized to 480 by 480 pixels) and applying a bunch of transformations to them like rotations, flipping, and so on. The FastAi built-in function responsible for augmentation automatically does random resized cropping after optionally providing it with 2 parameters: minimum scale of 0.75 and size of 224 [16]. What happens is the following:

- First, for each image in a batch, at least 75% of image pixels will be taken in the crop.
- Then, resizing each image to the standard image size chosen to be 224 by 224 pixels.

3.3. Transfer Learning

Due to the data limitation, transfer learning is the most appropriate type of deep learning that can effectively influence the accuracy of fracture detection. Transfer learning has the potential to improve the accuracy of the detection system by leveraging the pre-trained model's ability to extract relevant features from medical images [1]. Thus, for the sake of diagnosing fractures from the targeted AUBMC CT scan dataset available, transfer learning was applied on the generic ImageNet benchmark pre-trained models.

Fine-tuning is a transfer learning technique in which last layer(s) of pre-trained model are removed and weights of the network are tweaked by continuing the backpropagation [17]. Usually, only higher-level layers of the network are fine-tuned while keeping the earlier one. The particular reason for this is that features of earlier layers of pre-trained models learn common generic features such as edges, shapes, and textures, whereas the last layer(s) are more specialized toward the model's specific class.

Indeed, the FastAi approach (a python library) for fine-tuning, a transfer learning technique, was applied and works as follows. What was done first was to remove the last layer of the pre-trained model and adding a new layer with random weights for the fracture/normal classes. Then, freeze the pre-trained weights, and for one cycle train the last layer to allow the random weights to adjust to the CT scan dataset [18]. In that manner, the model can further learn to identify the specific features associated with vertebral fractures. After that, the entire network is unfrozen and is trained with a certain number of epochs. However, earlier layers don't require a lot of learning as they already have learnt the common attributes needed. Consequently, the learning rates used for early stages in the network are low but increments and become higher for subsequent layers. This method is called discriminative learning rate [19].

3.4. Deep Learning Architecture

Deep learning models that are frequently applied for medical computer vision applications and have proven to excel in their performance were selected with their revised updated versions [2][3][4][12].

3.4.1. ResNet-26, ResNet-RS-50

ResNet is short for Residual Network, a type of CNN that was presented in the paper "Deep Residual Learning for Image Recognition" by Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun in 2015 [20]. The success of this model is evident from the fact that its ensemble achieved the highest ranking in the ILSVRC 2015 classification competition, with a minimal error rate of 3.57%. Furthermore, in the ILSVRC competitions of 2015, it also achieved first place in ImageNet detection and ImageNet localization [20]. Residual blocks are utilized in deep residual networks to enhance the accuracy of the models. The primary advantage of this type of neural network lies in the concept of "skip connections," which is the foundation of the residual blocks [21]. ResNet has many variants that have different number of layers. ResNet26 and ResNet50 denote two variants that work with 26 and 50 neural network layers respectively ([number of layers-2] convolutional layers, one MaxPool layer, and one average pool layer). ResNet-26 attained a 75.29% top 1 accuracy for image classification on ImageNet with 16M parameters, while ResNet-50 achieved ImageNet top 1 accuracy of 79.04% with 26M parameters [22]. After ResNet-50 has been revised by applying improved training and scaling strategies, it was re-named ResNet-RS-50 and has achieved a top 1 accuracy of 78.8% with 36M parameters and 84.4% with 192M parameters [23][24].

3.4.2. InceptionResNetv2

The InceptionResNetv2 model, introduced by Szegedy et al. in 2016, is an extension of the Inception architecture. It replaces the filter concatenation stage of Inception with residual connections, which improve performance and simplify the

Inception blocks. With a depth of 572 and 55.9M parameters, the model attained a top accuracy of 80.3% and a top 5 accuracy of 95.4% [25][26].

3.4.3. Swin Transformer

The Swin Transformer is a variant of the Vision Transformer which creates hierarchical feature maps by combining image patches at deeper layers. It has a linear computational complexity relative to input image size because it only computes self-attention within each local window. This makes it a versatile backbone for image classification and dense recognition tasks. In contrast, previous vision Transformers produce feature maps of a single low resolution and have quadratic computational complexity relative to input image size because they compute self-attention globally [27].

The Swin Transformer is a model developed by the Microsoft research team which has a linear computational complexity with respect to input image [27]. Later, a proposed updated model named S3-tiny Transformer has attained 82.1% top 1 accuracy on ImageNet image classification with 28.1M parameters using 224 by 224 input images [28].

3.4.4. ConvNext

The authors Zhuang Liu *et al* introduced the ConvNeXT model in their paper "A ConvNet for the 2020s" [29]. ConvNeXT is a type of ConvNet that takes inspiration from the design of Vision Transformers but is composed entirely of convolutions. The authors assert that ConvNeXT performs better than Vision Transformers. ConvNeXTs exhibit comparable accuracy and scalability to Transformers, with an ImageNet top-1 accuracy of 87.8%. They also outperform Swin Transformers in COCO detection and ADE20K segmentation. Despite these achievements, ConvNeXTs retain the simplicity and efficiency of regular ConvNets [29]. ConvNeXT has a version called ConvNext-Tiny which have achieved a top-1 accuracy of 82.1% with 29M parameters [30].

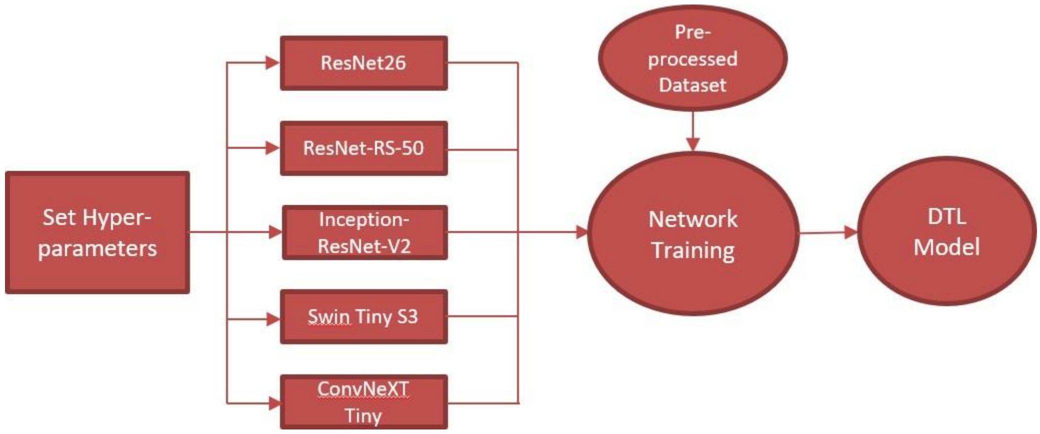

Figure 3: General Framework for Transfer Learning on Deep Learning Architectures

3.5. Evaluation Metrics

The evaluation of the suggested models was measured using the following two metrics:

- Accuracy: It measures the number of data instances that were correctly classified over the total number of data instances.
- F1-score: It is a harmonic mean of precision and recall, where precision is the ratio of correctly predicted positive observations to the total predicted positive observations and recall is the ratio of correctly predicted positive observations to all observations in the actual class.

Although accuracy is the most frequently used metric to assess the performance of classification models, F1-score is used alongside accuracy in cases where the dataset is imbalanced [31].

CHAPTER 4

POSSIBLE ALTERNATIVE APPROACH

Meta-learning is an approach to machine learning where a model learns how to learn from multiple tasks. Instead of learning a specific task, the model learns how to quickly adapt to new tasks based on its previous experience. This makes it a promising approach for addressing the problem of vertebral fractures detection.

In the context of vertebral fractures detection, meta-learning could be used to train a model to recognize patterns in CT scan images that indicate the presence of a fracture. The model would be trained on multiple datasets that contain images of different types of fractures, as well as images without fractures. Through this process, the model would learn to identify common features in CT scan images that indicate the presence of a fracture.

The advantage of using meta-learning for vertebral fractures detection is that it can enable the model to adapt to new datasets quickly, without requiring large amounts of labeled data. This is especially important in medical applications where labeled data is often scarce and expensive to obtain. By leveraging the knowledge learned from multiple datasets, a meta-learned model can achieve high performance on new datasets with minimal training.

One potential downside of using meta-learning is that it requires a large number of diverse datasets to train the model. This can be challenging in medical applications where there may be limited access to datasets due to privacy concerns and other regulatory issues. Additionally, meta-learning may be more computationally expensive than transfer learning, as it requires training multiple models on multiple datasets.

In summary, meta-learning is a promising approach for vertebral fractures detection that has the potential to overcome the limitations of traditional transfer learning methods. By learning how to quickly adapt to new datasets, a meta-learned model can achieve high performance with minimal training data. However, the approach also requires a large number of diverse datasets to train the model, which can be challenging in medical applications.

CHAPTER 5

EXPERIMENTAL RESULTS AND EVALUATION

A series of proposed models are trained on a high-end graphics processing unit (GPU). The GPU used is that of google colab and so it depends on their availability of GPUs. The deep learning package used is FastAi library using Python language. The deep learning models mentioned in chapter 4 were tested where the optimizer selected in all models was Adam optimizer. Moreover, all the models were trained on a batch size of 64. All experiments were done with data split percentage of 0.8, 0.1, 0.1 for training, validation, and testing respectively.

Different scenarios were tested depending on the data used:

- Dataset1: Data including only the 1567 central CT scan images
- Dataset2: Data including 1567 segmented central CT scan images
- Dataset3: Data extended and segmented to a total of 7835 CT scan images (central images with their corresponding surrounding 4 images).

First series of experiments were using Dataset1 and the results were reported in table 1.

	Validation Accuracy	Testing Accuracy	F1- Score
ResNet-26	67.3	75.3	75.3
ResNet-RS-50	66.7	72.2	72
Inception-ResNet-V2	61.5	74.1	74
Swin S3 Tiny	71.8	75.9	76
ConvNeXT Tiny in22k	69.9	**81**	81

Table 1: Summary of the 5 Pre-trained Models Perfomance on Dataset1

Second series of experiments were done with Dataset2. Results were presented in table 2.

	Validation Accuracy	Testing Accuracy	F1- Score
ResNet-26	68	77.2	77.2
ResNet-RS-50	70.5	74.1	74
Inception-ResNet-V2	69.2	81.7	81.6
Swin S3 Tiny	73.1	81	81
ConvNeXT Tiny in22k	65.4	**84.8**	84.8

Table 2: Summary of the 5 Pre-trained Models Performance on Dataset2

Third series of experiments were using Dataset3. The results were reported in table 3.

	Validation Accuracy	Testing Accuracy	F1- Score
ResNet-26	95.9	95.5	95.1
ResNet-RS-50	91.4	91.5	90.9
Inception-ResNet-V2	95.9	94.8	94.3
Swin S3 Tiny	95.5	94.5	94
ConvNeXT Tiny in22k	95.5	**96.4**	96.1

Table 3: Summary of the 5 Pre-trained Models Performance on Dataset3

ConvNeXT Tiny in22k model have surpassed all other models with testing accuracy of 81%, 84,8% and 96.4% when using Dataset1, Dataset2, and Dataset3. Inception-ResNet-V2 model comes in the second place for both Dataset2 and Dataset3 followed by Swin S3 Tiny model while they switch their ranking for Dataset1. Finally, ResNet-26 ranked in the 4th place before ResNet-RS-50 for all datasets. One can deduce that segmentation does improve the performance of all the models when we compare the accuracy of Dataset1 (before segmentation) and Dataset2 (after segmentation). The highest performance metrics of all models as for accuracy and F1 are when the data was extended further to be a total of 7835 segmented CT scan images.

To gain a complete understanding of the confusion matrix for this binary class classification problem, it's important to become familiar with certain terms.

- True Positive (TP) is used to describe a sample belonging to the positive class that is correctly classified.
- True Negative (TN) refers to a sample belonging to the negative class that is correctly classified.
- False Positive (FP) occurs when a sample belonging to the negative class is incorrectly classified as belonging to the positive class.
- False Negative (FN) occurs when a sample belonging to the positive class is incorrectly classified as belonging to the negative class. The confusion matrix for a binary class dataset summarizes these classification results.

Confusion matrix of the highest accuracy of all datasets (Dataset1, Dataset2, Dataset3) were presented.

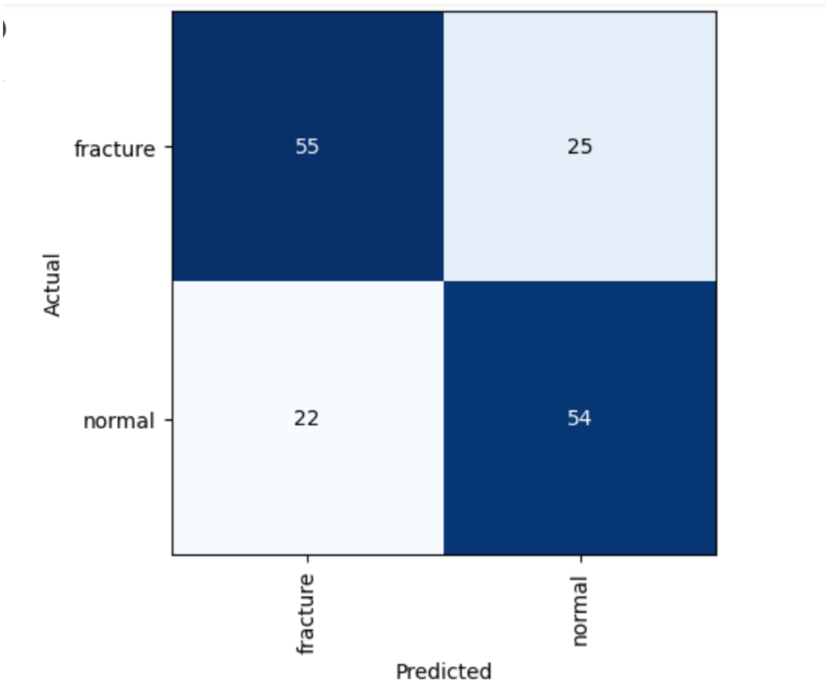

Figure 4: Confusion Matrix for ConvNext on Dataset1

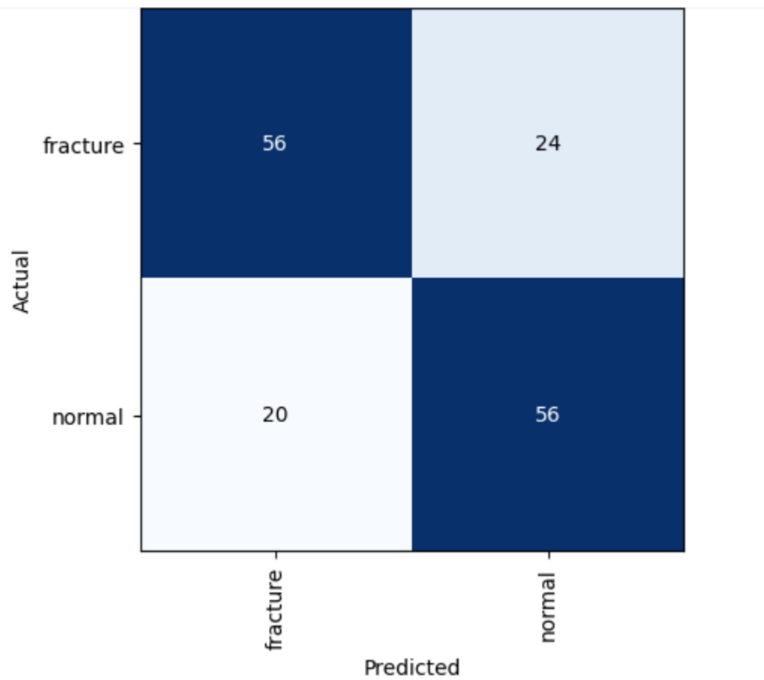

Figure 5: Confusion Matrix of ConvNext on Dataset2

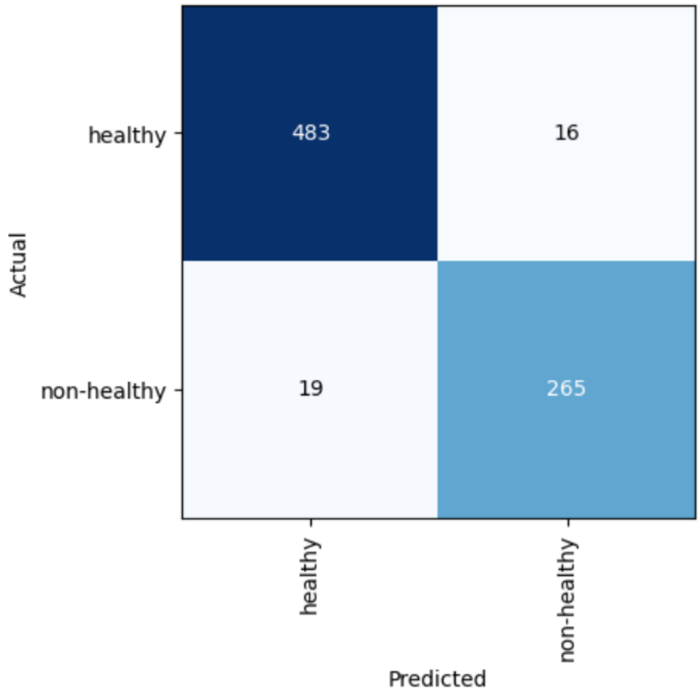

Figure 6: Confusion Matrix of ConvNext on Dataset3

From the confusion matrix, we can evaluate the precision. Precision is calculated for the positive class (in our case: Presence of fracture/ non-healthy) and it is the number of samples actually belonging to the positive class over the whole samples that were predicted to be positive by the model.

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

- Precision for Dataset1: 0.714
- Precision for Dataset2: 0.737
- Precision for Dataset3: 0.943

Moreover, specificity can be evaluated. It is the number of correctly predicted samples to be negative (in our case: the absence of fracture) over all the samples that actually belong to the negative class.

Specificity = $\frac{TN}{TN + FP}$

- Specificity for Dataset1: 0.711
- Specificity for Dataset2: 0.737
- Specificity for Dataset3: 0.968

Both sensitivity and specificity are important to be calculated especially that we are dealing with medical sensitive data, and so radiologists, doctors and even engineers would be concerned about the values of these two metrics, as they give a critical evaluation of our models. It is clearly shown that segmentation has increased the value of both sensitivity and specificity as we compare the values for Dataset1 (before segmentation) and Dataset2 (after segmentation). The best values were attained when extending the dataset and segmenting it where indeed they attained their highest values for Dataset3.

CHAPTER 6

CONCLUSION

Transfer learning models have been particularly effective in detecting diseases from limited medical data benchmarks. Therefore, this project proposed the use of transfer learning models to diagnose vertebral fractures from a CT scan dataset at AUBMC.

Classical data augmentation along with vertebra segmentation from images have proven improved performance using five different deep architecture models. The ConvNeXT_Tiny_in22k model has achieved the highest testing accuracy in all scenarios: with and without segmentation, and even after extending the data to achieve 96.4%.

As for future work, we plan to detect the fracture per class (depending on the level of severity of the fracture), and thus the task will become a multi-class detection, and this is done by getting more data from AUBMC to train models more robustly.

REFERENCES

[1] Donges, Niklas. "What Is Transfer Learning? Exploring the Popular Deep Learning Approach." *Built In*, https://builtin.com/data-science/transfer-learning.

[2] Meena, Tanushree, and Sudipta Roy. "Bone Fracture Detection Using Deep Supervised Learning from Radiological Images: A Paradigm Shift." *Diagnostics*, vol. 12, no. 10, 2022, p. 2420., https://doi.org/10.3390/diagnostics12102420.

[3] Loey, Mohamed et al. "A deep transfer learning model with classical data augmentation and CGAN to detect COVID-19 from chest CT radiography digital images." *Neural computing & applications*, 1-13. 26 Oct. 2020, doi:10.1007/s00521-020-05437-x

[4] Mahanty, Chandrakanta, et al. "Covid-19 Detection with X-Ray Images by Using Transfer Learning." *Journal of Intelligent & Fuzzy Systems*, vol. 43, no. 2, 2022, pp. 1717–1726., https://doi.org/10.3233/jifs-219273.

[5] An, Guangzhou, et al. "Hierarchical Deep Learning Models Using Transfer Learning for Disease Detection and Classification Based on Small Number of Medical Images." *Scientific Reports*, vol. 11, no. 1, 2021, https://doi.org/10.1038/s41598-021-83503-7.

[6] Maqsood, Muazzam, et al. "Transfer Learning Assisted Classification and Detection of Alzheimer's Disease Stages Using 3D MRI Scans." *Sensors*, vol. 19, no. 11, June 2019, p. 2645. Crossref, https://doi.org/10.3390/s19112645.

[7] Arooj, Sahar, et al. "Breast Cancer Detection and Classification Empowered with Transfer Learning." *Frontiers in Public Health*, vol. 10, 2022, https://doi.org/10.3389/fpubh.2022.924432.

[8] Rashid, Javed, et al. "Skin Cancer Disease Detection Using Transfer Learning Technique." *Applied Sciences*, vol. 12, no. 11, June 2022, p. 5714. Crossref, https://doi.org/10.3390/app12115714.

[9] Krishnamurthy, Surya, et al. "Evaluating Deep Neural Network Architectures with Transfer Learning for Pneumonitis Diagnosis." *Computational and Mathematical Methods in Medicine*, vol. 2021, 2021, pp. 1–12., https://doi.org/10.1155/2021/8036304.

[10] Ayana, Gelan, et al. "Vision-Transformer-Based Transfer Learning for Mammogram Classification." *Diagnostics*, vol. 13, no. 2, Jan. 2023, p. 178. *Crossref*, https://doi.org/10.3390/diagnostics13020178.

[11] Okolo, Gabriel Iluebe, et al. "IEVIT: An Enhanced Vision Transformer Architecture for Chest X-Ray Image Classification." *Computer Methods and Programs in Biomedicine*, vol. 226, 2022, p. 107141., https://doi.org/10.1016/j.cmpb.2022.107141.

[12] Zhang, Lei, and Yan Wen. “A Transformer-Based Framework for Automatic COVID19 Diagnosis in Chest CTS.” *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2021, https://doi.org/10.1109/iccvw54120.2021.00063.

[13] Jiang, Zhencun, et al. “Method for Diagnosis of Acute Lymphoblastic Leukemia Based on VIT-CNN Ensemble Model.” *Computational Intelligence and Neuroscience*, vol. 2021, 2021, pp. 1–12., https://doi.org/10.1155/2021/7529893.

[14] Mayo Foundation for Medical Education and Research. (2022, January 6). *CT Scan*. Mayo Clinic. Retrieved April 1, 2023, from https://www.mayoclinic.org/tests-procedures/ct-scan/about/pac-20393675

[15] Vasani, Dipam. “Data Augmentations in Fastai.” *Medium*, Towards Data Science, 28 Nov. 2019, https://towardsdatascience.com/data-augmentations-in-fastai-84979bbcefaa.

[16] Bello, Olabisi. *Image Classification Using FASTAI — Tutorial Pt.1*, 30 Jan. 2022, https://medium.com/@FastAIButMakeItSlow/image-classification-using-fastai-tutorial-pt-1-acea74cd7faa.

[17] Singh, Kamaldeep. “How to Improve Class Imbalance Using Class Weights in Machine Learning.” *Analytics Vidhya*, 2 Dec. 2022, https://www.analyticsvidhya.com/blog/2020/10/improve-class-imbalance-class-weights/.

[18] “Fastai fit_one_cycle & fine_tune and Super-Convergence (Leslie Smith).” *Mldurga*, 14 Oct. 2021, https://mldurga.github.io/easydl/paper_reading/2021/10/14/super_convergence.html.

[19] Shastri, Akash. “Why I Use FASTAI and You Should Too.” *Medium*, Towards Data Science, 10 Nov. 2020, https://towardsdatascience.com/why-i-use-fastai-and-you-should-too-a421f6c99508.

[20] He, Kaiming, et al. “Deep Residual Learning for Image Recognition.” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, https://doi.org/10.1109/cvpr.2016.90.

[21] Boesch, Gaudenz. “Deep Residual Networks (ResNet, RESNET50) - 2023 Guide.” *Viso.ai*, 2 Jan. 2023, https://viso.ai/deep-learning/resnet-residual-neural-network/.

[22] “ResNet.” *Papers With Code*, https://paperswithcode.com/model/resnet?variant=resnet26.

[23] “Papers with Code - Imagenet Benchmark (Image Classification).” *The Latest in Machine Learning*, https://paperswithcode.com/sota/image-classification-on-imagenet.

[24] Bello, Irwan, et al. "Revisiting ResNets: Improved Training and Scaling Strategies." *NeurIPS*, 2021, https://doi.org/arXiv:2103.07579v1.

[25] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A.A. (2017). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. AAAI

[26] “Improving inception and image classification in TensorFlow,” Google AI Blog, 31-Aug-2016. [Online]. Available: https://ai.googleblog.com/2016/08/improving-inception-and-image.html. [Accessed: 1-Apr-2023].

[27] Liu, Ze, et al. “Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows.” *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, https://doi.org/10.1109/iccv48922.2021.00986)

[28] Chen, Minghao, et al. *Searching the Search Space of Vision Transformer*, 29 Nov. 2021, https://doi.org/ https://doi.org/10.48550/arXiv.2111.14725

[29] Liu, Zhuang, et al. “A Convnet for the 2020s.” *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, https://doi.org/10.1109/cvpr52688.2022.01167

[30] Hou, Qibin et al. “*Conv2Former: A Simple Transformer-Style ConvNet for Visual Recognition”*, 22 Nov. 2022, https://doi.org/arXiv:2211.11943v1.

[31] Al Sahili, Zahraa, and Mariette Awad. “The Power of Transfer Learning in Agricultural Applications: AgriNet.” *Frontiers in Plant Science*, vol. 13, 2022, https://doi.org/10.3389/fpls.2022.992700.