Comparison of Transfer Learning Model Accuracy for Osteoporosis Classification on Knee Radiograph

Usman Bello Abubakar
Department of Computer Science,
Baze University
Abuja, Nigeria
usman.abu@bazeuniversity.edu.ng

Moussa Mahamat Boukar
Department of Computer Science,
Nile University of Nigeria
Abuja, Nigeria
musa.muhammed@nileuniversity.edu.ng

Steve Adeshina
Department of Computer Engineering,
Nile University of Nigeria
Abuja, Nigeria
steve.adeshina@ nileuniversity.edu.ng

2022 2nd International Conference on Computing and Machine Intelligence (ICMI) | 978-1-6654-7483-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICMI55296.2022.9873731

Abstract— **In terms of financial costs and human suffering, osteoporosis poses a serious public health burden. Reduced bone mass, degeneration of the microarchitecture of bone tissue, and an increased risk of fracture are its main skeletal symptoms. Osteoporosis is caused not just by low bone mineral density, but also by other factors such as age, weight, height, and lifestyle. Recent advancement in Artificial Intelligence (AI) has led to successful applications of expert systems that use Deep Learning techniques for osteoporosis diagnosis based on some modalities such as dental radiographs amongst others. This study uses a dataset of knee radiographs (i.e., knee-Xray images) to apply and compare three robust transfer learning model algorithms: GoogLeNet, VGG-16, and ResNet50 to classify osteoporosis. From the statistical analysis and scikit learn python analysis, the accuracy of the GoogLeNet model was 90%, the accuracy of the VGG-16 model was 87% and lastly, the accuracy of the ResNet-50 model was 83%.**

Keywords— Osteoporosis, Transfer Learning Models, Dual-Energy X-ray Absorptiometry, Bone Mineral Density

I. Introduction

Osteoporosis, which originates from Greek, is literally translated as porous bone. [1]. According to World Health Organization, low bone mass and microarchitectural degeneration of bone tissue are the characteristics of osteoporosis, a progressive systemic skeletal disease that increases bone fragility and fracture susceptibility. [1].

Osteoporosis is a metabolic bone condition in which osteoclastic bone resorption is not counteracted at the cellular level by osteoblastic bone synthesis. As a result, bones become brittle and weak, putting them at risk of fracture. Traditional osteoporosis pathophysiology centered on endocrine factors such as estrogen shortage or vitamin D deficiency, as well as secondary hyperparathyroidism. Although osteoporosis can affect persons of any age or gender, it is typically an age-related disease that affects women more than males [2].

Osteoporosis is diagnosed by dual-energy X-ray absorptiometry (DXA), which measures bone mineral density (BMD). Therefore, checking for osteoporosis can have a big impact on how patients turn out. However, because osteoporosis is hidden until severe fragility fractures, osteoporosis is mostly misdiagnosed, and DXA screening for osteoporosis has been underutilized [3]. Patients frequently underestimate the severity of the sickness and, as a result, refuse to volunteer for the screening program [4].

There is a growing consensus that other screening approaches are needed to overcome the shortcomings of DXA as an osteoporosis diagnosis method. Adults frequently undergo Abdominal-Pelvic Computed Tomography (APCT) to examine a variety of disorders during routine health check-ups or to follow up on previously identified conditions. Even if only a tiny percentage of these scans were for osteoporosis screening opportunistically, there would be a significant impact. APCT has shown promising results in opportunistic osteoporosis screening in several studies [5][9].

Artificial intelligence (AI) and Deep Learning (DL) is used for image interpretation for osteoporosis classification [7]. In a 2019 review paper, AI advancements have aided in the detection of osteoporosis [8]. The following methods were employed: dental radiographs [9], [13], spine radiographs[7], [14], hand and wrist radiographs [12], [13].

This study tends to apply and compare three transfer learning models to classify osteoporosis using a dataset of knee radiographs (i.e., knee X-rays). The transfer learning models applied were GoogLeNet, VGG-16, and ResNet-50. In addition, to compare the diagnostic performance of the three models, several state-of-the-art neural networks metric was used.

II. RELATED WORKS

In order to estimate the prevalence of osteoporosis in postmenopausal women, machine learning techniques were applied [14]. The researchers constructed a non-linear model using regression support vector machines (SVM) for a sample of 305 postmenopausal women to ascertain the association between BMD, diet, and lifestyle variables. A preliminary assessment of BMD in the study women was also used to decide whether densitometry testing was required (based on a questionnaire with questions largely regarding dietary habits). Regression trees were used to identify which elements were most crucial, and SVMs were used to build a mathematical model that reflected the relationship. The most important things for postmenopausal women to do to prevent bone density loss include consuming extra calcium, getting enough sun, managing their weight, exercising regularly, and eating enough calories [14].

The authors in [15], based on identified risk factors, established a modern, effective bone disease prediction model. Then, using Pre-training and fine-tuning, it was possible to identify the early risk factors for determining the start of bone problems. During the pre-training phase, the most important risk factors are combined with model parameters to calculate contrastive divergence, which reduces record size. The results of the previous phase were compared using the ground truth values "g1" and "g2," where g1 represented osteoporosis and g2 represented a rate of bone loss. Deep Belief Network (DBN) was used to generate the model, which was then compared to models created before and after critical feature identification. The study's findings suggested that including relevant variables could increase the prediction model's effectiveness. [15].

The authors in [16] created and assessed DL approaches for osteoporosis classification using Dental Panoramic Radiographs (DPR). In this work, various CNN models for osteoporosis discriminating accuracy were tested using panoramic radiograph pictures that had been categorized based on BMD value (T-score). The effects of transfer learning and fine-tuning a deep CNN model were also evaluated in terms of classification performance. Deep CNN's have been found to be useful for classifying images, but because they need a lot of training data, it is challenging to apply them to radiographic medical imaging data. Transfer learning is a popular strategy for training deep CNN's without "overfitting" when the target dataset is significantly smaller than the basis dataset [17].

III. MATERIALS AND METHODS

A. Data Acquisition

The dataset was gotten from a public dataset repository for machine learning called Kaggle. The name of the Kaggle repo is "Osteoporosis Knee X-ray Dataset", version 1, uploaded on the 16th of September, 2021 accessible via www.kaggle.com/stevepython/datasets. The number of images was increased using data augmentation. Fig 1 shows two images from the dataset indicating osteoporosis cases and normal cases.

Fig. 1. Osteoporosis case and normal case.

The dataset, after statical augmentation using python augmentation functions, comprises 323 normal knee radiograph images and 323 osteoporotic knee radiograph images of patients. Table I shows the splitting of image data into a train, test and validation data.

TABLE I. IMAGE DISTRIBUTION

Class	Total	Training	Validation	Testing
Normal(0)	323	207	52	65
Osteoporosis(1)	323	207	52	65

B. Image Scaling

In most image data, the pixel values are integers with values ranging from 0 to 255. Since neural networks only analyze inputs with modest weight values, inputs with large integer values can interfere with or slow down the learning process. Therefore, picture normalization is a recommended practice: pixel values range between 0 and 1. The images in the dataset were normalized(rescaled) using the python ImageDataGenerator method and passing rescale=1. /255 as its argument.

C. Data Augmentation

Overfitting can be reduced by using a technique called data augmentation. Overfitting occurs when a model learns a function with a relatively large variance to perfectly model the training data [21]. For this study, the Keras ImageDataGenerator python class was used to perform data augmentation using a variety of augmentation techniques as itemized below:

1. Standardization
2. Rotation
3. Shifts
4. Brightness changes, among others

The Keras ImageDataGenerator class is intended to give real-time data augmentation, which is said to be its key advantage. Every epoch, the model is given fresh versions of the images thanks to the ImageDataGenerator class.

D. Transfer Learning Model Architecture

Three transfer learning model architectures were applied: GoogLeNet, VGG-16, and ResNet-50. All the layers of the pre-trained model were made to be non-trainable. However, some of the layers could be re-trained to increase performance but at the cost of a higher chance of model overfitting. For this model, as the loss metric, binary_crossentropy was used as the dataset target has two classes (i.e., binary classification problem). RMSprop is the chosen optimizer, and its learning rate is 0.001. Each model underwent 30 epochs of training.

The GoogleNet is a 22-layer deep convolutional neural architecture that addressed computer vision issues such as object recognition and image classification in the ImageNet It has achieved 93.9% accuracy in the top 5 results [19]. Fig.2 shows the GoogLeNet model's architecture.

VGG16 is CNN architecture. Having 16 layers and is distinguished by its simplicity by having just a stack of 33 convolutional layers on top of each other, with the max-pooling layers handling the rising depth and volume size. A softmax layer comes after two fully linked layers with 4096 nodes each [21]. In ImageNet, the VGG16 model obtained top-5 test accuracy of 92.7%. Fig. 3. Shows the VGG-16 model architecture [21].

ResNet-50 is a 50-layer CNN. Five stages, each with a convolution and an identity block, make up the ResNet-50 model. As with the identity blocks, each convolution block has three convolution layers [22]. The identity connection between the layers is the single addition to the basic network that transforms it into a residual network. Fig 4 shows the architecture of the ResNet-50 model [22].

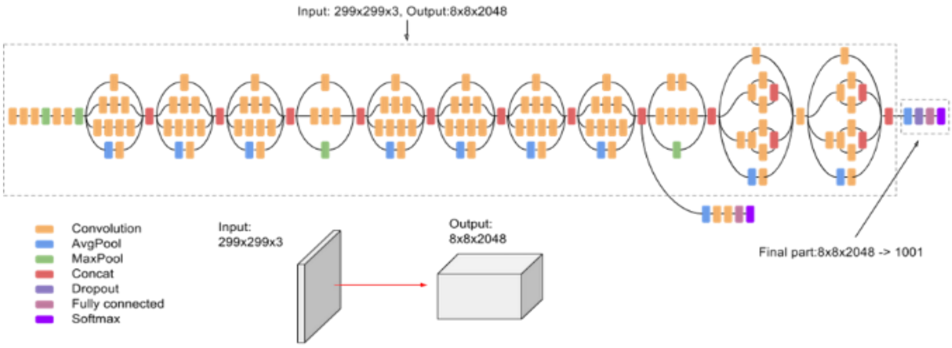

Fig. 2. GoogLeNet Model [20].

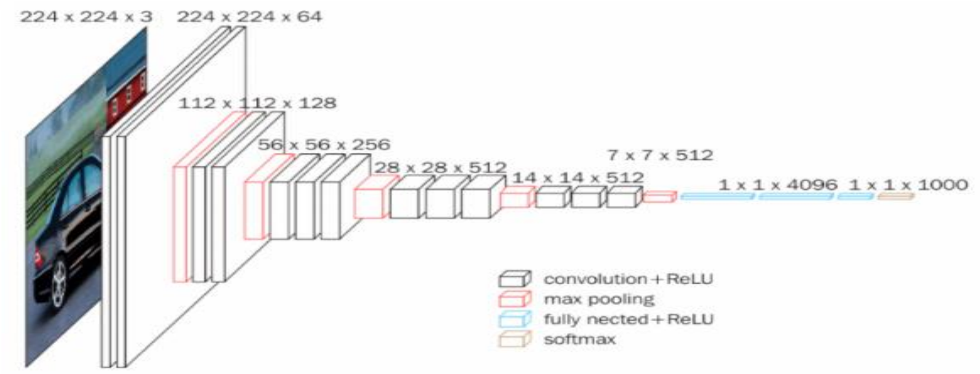

Fig. 3. VGG-16 Model [21].

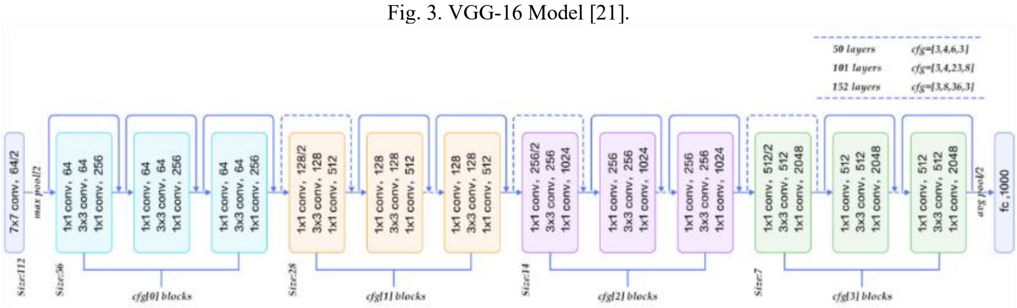

Fig. 4. ResNet-50 Model [22].

IV. RESULTS

We experimented with the osteoporosis patient knee x-rays dataset. In all transfer learning models, the dataset was split into 80:20 ratio for training and testing. The overall accuracy obtained for all the classifiers on the dataset is summarized in Table II. Each model underwent 30 epochs of training. For all models, as the loss metric, binary_crossentropy was used as the dataset target has two classes (i.e., binary classification problem). RMSprop is the chosen optimizer, and its learning rate is 0. 001. The Keras evaluate function was invoked on the compiled model with the test data as an argument to evaluate the accuracy of the models.

A. *Confusion Matrix*

The confusion matrix for the three transfer learning models is presented in Fig. 5 and 6.

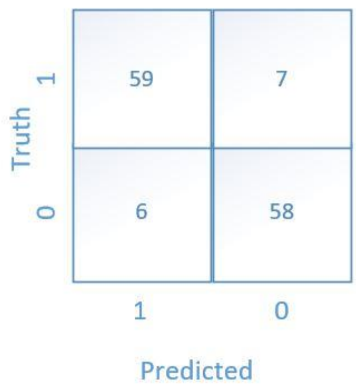

Fig. 5. Confusion Matrix for GoogLeNet Model.

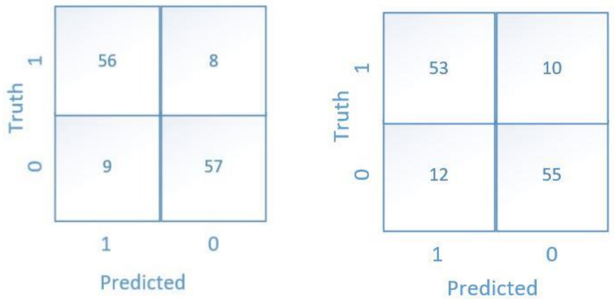

Fig. 6. Confusion Matrix for VGG-16 (left) and ResNet-50 (right) Models.

B. *Sensitivity, Specificity, and Accuracy*

Sensitivity is the ability to identify osteoporotic patients (The actual positives).

Specificity is the ability to identify non-osteoporotic patients (The actual negatives).

$$\text{Sensitivity} = \frac{true\ positive}{true\ positive + false\ negative} \quad (1)$$

$$\text{Specificity} = \frac{true\ negative}{true\ negative\ + false\ negative} \quad (2)$$

$$\text{Accuracy} \ = \frac{true\ negative + true\ positive}{all\ cases} \quad (3)$$

Table II shows the sensitivity, accuracy, and specificity of the models. Table III provides a comparison of our work with similar works

TABLE II. RESULTS OBTAINED FOR TRANSFER LEARNING

	Accuracy	**Sensitivity**	**Specificity**
GoogLeNet	0.90	0.91	0.91
VGG-16	0.87	0.86	0.86
ResNet-50	0.83	0.81	0.82

TABLE III. COMPARISON WITH OTHER SIMILAR WORK

Paper	**Dataset**	**Method**	**Ac**	**Se**	**Sp**
Our Paper	Knee	GoogLeNet	0.90	0.91	0.91
Our Paper	Knee	VGG-16	0.87	0.86	0.86
Our Paper	Knee	ResNet-50	0.83	0.81	0.82
[23]	Hip	ResNet-18	0.79	0.86	0.86
[23]	Hip	ResNet-34	0.84	0.88	0.86
[24]	Dental	VGG-16-Fine-Tuning	0.84	0.90	0.81
[24]	Dental	CNN with 3 layers	0.66	0.68	0.65
[25]	Panoramic radiograph	ResNet-50	0.83	0.75	0.90

*Ac: Accuracy, Se: Sensitivity, Sp: Specificity

C. *Accuracy Graph*

The behavior of the models in terms of accuracy over several iterations can be better visualized in the following figures. The test accuracy for GoogLeNet, VGG-16 and ResNer-50 are shown in Fig. 7, 8 and 9, respectively. We can see that overall the test accuracy for all models seems to behave similar on the given dataset.

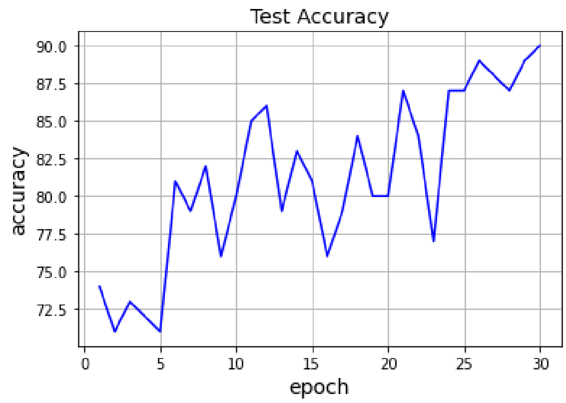

Fig. 7. Accuracy Graph for GoogLeNet Model.

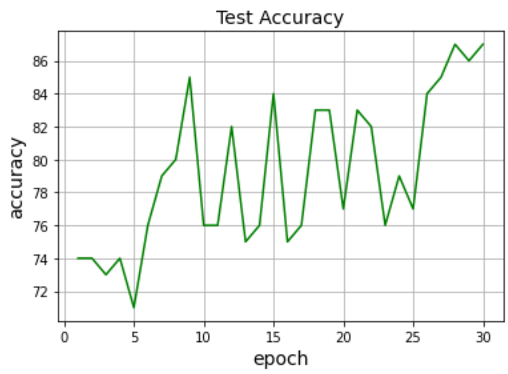

Fig. 8. Accuracy Graph for VGG-16 Model.

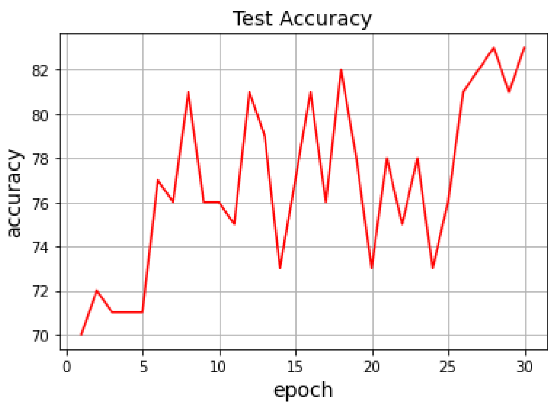

Fig. 9. Accuracy Graph for ResNer-50 Model.

V. CONCLUSION

This study demonstrated that Deep Learning Transfer Models can diagnose osteoporosis from knee radiographs. This paper uses a dataset of knee radiographs (i.e., knee-Xray images) to apply and compare three robust transfer learning model algorithms: GoogLeNet, VGG-16, and ResNet-50 to classify osteoporosis. ImageDataGenerator was used to augment the dataset and increase the number of training data to provide a variety of images for the models. The results gotten from the statistical analysis and scikit learn python analysis, show that the accuracy of the GoogLeNet model was 90%, the accuracy of the VGG-16 model was 87% and lastly, the accuracy of the ResNet-50 model was 83%. In addition to model accuracy, other performance metrics were used to evaluate the model such as confusion matrix, sensitivity, and specificity.

Osteoporosis is caused not just by low bone mineral density, but also by other factors such as age, gender, weight, height, and so on. These are clinically important risk factors for osteoporosis. For future work, we would like to extend our methods by adding patient variables such as age, and gender, amongst others, as clinical covariates to create an ensemble model with the transfer learning models.

REFERENCES

[1] S. Gschmeissner and S. Photo Library, "Diagnosis, assessment and management of osteoporosis," *Prescriber*, vol. 31, no. 1, pp. 14–19, Jan. 2020, doi: 10.1002/PSB.1815.

[2] U. Föger-Samwald, P. Dovjak, U. Azizi-Semrad, K. Kerschan-Schindl, and P. Pietschmann, "Osteoporosis: Pathophysiology and therapeutic options," *EXCLI Journal*, vol. 19, p. 1017, 2020, doi: 10.17179/EXCLI2020-2591.

[3] E. M. Curtis, R. J. Moon, N. C. Harvey, and C. Cooper, "The impact of fragility fracture and approaches to osteoporosis risk assessment worldwide," *Bone*, vol. 104, pp. 29–38, Nov. 2017, doi: 10.1016/j.bone.2017.01.024.

[4] J. R. Curtis *et al.*, "Longitudinal Trends in Use of Bone Mass Measurement Among Older Americans, 1999–2005," *Journal of Bone and Mineral Research*, vol. 23, no. 7, pp. 1061–1067, Jul. 2008, doi: 10.1359/JBMR.080232.

[5] H. K. Lim, H. il Ha, S. Y. Park, and K. Lee, "Comparison of the diagnostic performance of CT Hounsfield unit histogram analysis and dual-energy X-ray absorptiometry in predicting osteoporosis of the femur," *European Radiology 2018 29:4*, vol. 29, no. 4, pp. 1831–1840, Sep. 2018, doi: 10.1007/S00330-018-5728-0.

[6] S. Jang, P. M. Graffy, T. J. Ziemlewicz, S. J. Lee, R. M. Summers, and P. J. Pickhardt, "Opportunistic osteoporosis screening at routine abdominal and Thoracic CT: Normative L1 trabecular attenuation values in more than 20 000 adults," *Radiology*, vol. 291, no. 2, pp. 360–367, May 2019, doi: 10.1148/RADIOL.2019181648/ASSET/IMAGES/LARGE/RADIOL.2019181648.TBL1.JPEG.

[7] S. Lee, E. K. Choe, H. Y. Kang, J. W. Yoon, and H. S. Kim, "The exploration of feature extraction and machine learning for predicting bone density from simple spine X-ray images in a Korean population," *Skeletal Radiology*, vol. 49, no. 4, pp. 613–618, Apr. 2020, doi: 10.1007/S00256-019-03342-6.

[8] U. Ferizi, S. Honig, and G. Chang, "Artificial intelligence, osteoporosis and fragility fractures," *Curr Opin Rheumatol*, vol. 31, no. 4, pp. 368–375, Jul. 2019, doi: 10.1097/BOR.0000000000000607.

[9] J. J. Hwang *et al.*, "Strut analysis for osteoporosis detection model using dental panoramic radiography," *Dentomaxillofacial Radiology*, vol. 46, no. 7, 2017, doi: 10.1259/DMFR.20170006.

[10] K. S. Lee, S. K. Jung, J. J. Ryu, S. W. Shin, and J. Choi, "Evaluation of Transfer Learning with Deep Convolutional Neural Networks for Screening Osteoporosis in Dental Panoramic Radiographs," *undefined*, vol. 9, no. 2, Feb. 2020, doi: 10.3390/JCM9020392.

[11] H. P. Dimai *et al.*, "Assessing the effects of long-term osteoporosis treatment by using conventional spine radiographs: results from a pilot study in a sub-cohort of a large randomized controlled trial," *Skeletal Radiology*, vol. 48, no. 7, pp. 1023–1032, Jul. 2019, doi: 10.1007/S00256-018-3118-Y.

[12] A. S. Areeckal, N. Jayasheelan, J. Kamath, S. Zawadynski, M. Kocher, and S. David S, "Early diagnosis of osteoporosis using radiogrammetry and texture analysis from hand and wrist radiographs in Indian population," *Osteoporosis International*, vol. 29, no. 3, pp. 665–673, Mar. 2018, doi: 10.1007/S00198-017-4328-1.

[13] N. Tecle, J. Teitel, M. R. Morris, N. Sani, D. Mitten, and W. C. Hammert, "Convolutional Neural Network for Second Metacarpal Radiographic Osteoporosis Screening," *The Journal of Hand Surgery*, vol. 45, no. 3, pp. 175–181, Mar. 2020, doi: 10.1016/J.JHSA.2019.11.019.

[14] C. Ordóñez, J. M. Matías, J. F. de Cos Juez, and P. J. García, "Machine learning techniques applied to the determination of osteoporosis incidence in post-menopausal women," *Mathematical and Computer Modelling*, vol. 50, no. 5–6, pp. 673–679, Sep. 2009, doi: 10.1016/J.MCM.2008.12.024.

[15] M. Saranya, M. Sc, M. Phil, and K. Sarojini, "An Improved and Optimal Prediction of Bone Disease Based On Risk Factors." [Online]. Available: www.ijcsit.com

[16] K. S. Lee, S. K. Jung, J. J. Ryu, S. W. Shin, and J. Choi, "Evaluation of Transfer Learning with Deep Convolutional Neural Networks for Screening Osteoporosis in Dental Panoramic Radiographs," *undefined*, vol. 9, no. 2, Feb. 2020, doi: 10.3390/JCM9020392.

[17] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?".

[18] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, Dec. 2019, doi: 10.1186/S40537-019-0197-0.

[19] "Transfer Learning using Inception-v3 for Image Classification | by Tejan Irla | Analytics Vidhya | Medium." https://medium.com/analytics-vidhya/transfer-learning-using-inception-v3-for-image-classification-86700411251b (accessed Apr. 04, 2022).

[20] "Inception V3 Model Architecture." https://iq.opengenus.org/inception-v3-model-architecture/ (accessed Mar. 30, 2022).

[21] "What is VGG16? — Introduction to VGG16 | by Great Learning | Medium." https://medium.com/@mygreatlearning/what-is-vgg16-introduction-to-vgg16-f2d63849f615 (accessed Mar. 30, 2022).

[22] "Detailed Guide to Understand and Implement ResNets – CV-Tricks.com." https://cv-tricks.com/keras/understand-implement-resnets/ (accessed Jun. 21, 2022).

[23] N. Yamamoto *et al.*, "Deep learning for osteoporosis classification using hip radiographs and patient clinical covariates," *Biomolecules*, vol. 10, no. 11, pp. 1–13, Nov. 2020, doi: 10.3390/BIOM10111534.

[24] K. S. Lee, S. K. Jung, J. J. Ryu, S. W. Shin, and J. Choi, "Evaluation of Transfer Learning with Deep Convolutional Neural Networks for Screening Osteoporosis in Dental Panoramic Radiographs," *Journal of Clinical Medicine*, vol. 9, no. 2, Feb. 2020, doi: 10.3390/JCM9020392.

[25] S. Sukegawa *et al.*, "Identification of osteoporosis using ensemble deep learning model with panoramic radiographs and clinical covariates," *Scientific Reports 2022 12:1*, vol. 12, no. 1, pp. 1–10, Apr. 2022, doi: 10.1038/s41598-022-10150-x.