Utilizing Deep Learning for Osteoporosis Diagnosis through Knee X-Ray Analysis

Mengyuan Shen[1]

[1] The department of computer arts, School of visual arts, New York, NY 10010, USA
shenmgy@usf.edu

Abstract. A common progressive disease called osteoporosis is defined by a steady decline of bone density that weakens bones and raises the risk of fractures. This condition significantly impacts the quality of life of affected individuals, particularly among the elderly. The goal of this study is to identify osteoporosis by analyzing knee x-rays with powerful deep learning models. By leveraging artificial intelligence technology, this approach aims to enhance diagnostic accuracy and efficiency, providing a more convenient and non-invasive method for early detection and treatment of osteoporosis. Ultimately, this can help lower the risk of fractures and enhance the overall health outcomes for patients. Specifically, this paper employed the Visual Geometry Group (VGG) 19 model, known for its ability to extract detailed features from 2D images. Using datasets from Kaggle and Mendeley, the model achieved an accuracy of 89% after 17 epochs of training, demonstrating its effectiveness in identifying osteoporosis traits in knee x-rays. This approach provides an alternative to the traditional hip x-ray diagnosis, potentially easing the diagnostic process for patients. Furthermore, this method could help in the early detection and intervention of osteoporosis, thereby reducing fracture risks. The outcomes of the study highlight the possibilities of deep learning models in improving diagnostic accuracy and patient outcomes in osteoporosis management.

Keywords: Osteoporosis, Deep Learning, Knee X-Rays, VGG.

1 Introduction

The illness osteoporosis develops over time. It is marked by a reduction in bone density. It is usually detected after a fracture. Early detection and diagnosis of osteoporosis is therefore very important. Osteoporosis can lead to fractures, restricted movement, and reduced quality of life [1]. Screening for osteoporosis in postmenopausal women is something that can greatly help prevent fractures. In clinical practice, identifying osteoporosis, the Dual-energy X-ray absorptiometry (DXA) technique is frequently utilized [1]. The body is exposed to x-rays with two distinct wavelengths. bone mineral density (BMD) is calculated by comparing the difference between soft tissue and bone passing through the energy of the rays. The

Y. Wang (ed.), *Proceedings of the 2024 International Conference on Artificial Intelligence and Communication (ICAIC 2024)*, Advances in Intelligent Systems Research 185,
https://doi.org/10.2991/978-94-6463-512-6_58

BMD can indicate the progress of the disease and determine the final treatment. The most prevalent type of arthritis is knee osteoarthritis (OA) [2]. It is a common condition in the older age group. x-rays are used to diagnose OA. therefore, x-rays of the knee are very common.

Since the 1970s, radiographs of the hip have been used to perform osteoporosis and fracture risk assessments. Individual bone risk assessment is done using several algorithms, such as the Q-fracture algorithm and the Fracture Risk Assessment Tool (FRAX). FRAX was calculated based on age weight and dichotomous risk factors [3]. The probability of fracture varies according to region. FRAX is calibrated with six million calculations per year in different regions. It was launched in April 2008. Q-fracture is a risk prediction algorithm. It predicts the likelihood of osteoporotic fracture in people with primary care [4]. These algorithms utilize patient physiologic data, past medical history, lifestyle habits, and family history to predict future fracture probability. Artificial intelligence has been used in medical scenarios to aid in judgment. For osteoporosis, hip, and dental x-rays have been used as analysis. Knee osteoarthritis can likewise be diagnosed using x-rays [5]. Convolutional neural network (CNN) models were used to learn dental panoramic radiographs and BMD measures. According to the findings, osteoporosis and other medical problems may be properly diagnosed using deep learning and CNN. In a research, research project, hip radiographs were used to classify and predict osteoporosis using deep neural networks.

The aim of this research is to use CNN to predict osteoporosis from knee x-ray images. Specifically, this research aims to build and evaluate a CNN-based model for osteoporosis prediction and categorization. The experimental results demonstrate the effectiveness of CNN modeling in accurately identifying osteoporosis, showcasing its potential as a reliable diagnostic tool. This work allows for early detection and timely therapeutic interventions for osteoporosis, which offers major contributions to the fields of medical imaging and bone health. By employing advanced CNN techniques, the research not only enhances the precision of osteoporosis detection but also helps in reducing the risk of fractures. Consequently, this approach can lead to substantial improvements in the quality of life for patients by enabling proactive management of the condition. The findings emphasize the crucial role of machine learning in developing medical diagnostics and the practical uses of artificial intelligence (AI) in improving patient outcomes.

2 Methodology

2.1 Dataset Description and Preprocessing

The dataset used in this study mainly are x-ray images from Kaggle [6]. The data in this study comes from different datasets. All the datasets are summarized and categorized under two different labels. This dataset was categorized into normal and osteoporosis labels before it was fed into the model. Each image was re-seized into 284*284 to ensure it was the same size. The pre-processed images are saved locally for model learning. 780 images are under normal label. 793 images are under osteoporosis label. Fig. 1 presents the images of normal label and osteoporosis label.

Fig. 1. Images from the Kaggle [6].

2.2 Proposed Approach

The goal of this study was to perform image classification for osteoporosis of the knee. A CNN model was used for classification. After introducing the dataset, the model use Data Frame as its data generator. The CNN model in this study is Visual Geometry Group 19 (VGG19) for classification by feature extraction of planar images. This technique employs the use of small components in order to facilitate the extraction of the image's features [7]. The Fig. 2 illustrate the pipeline of the model.

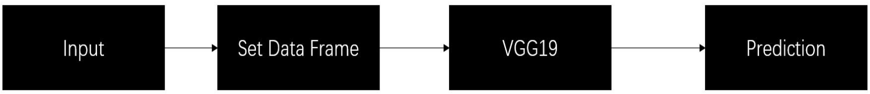

Fig. 2. The pipeline of the model (Photo/Picture credit: Original).

VGG19

There are 19 layers in VGG19, comprising 3 fully connected layers and 16 convolutional layers. It mainly uses conv2d as its convolutional layer which is used to extract local features such as texture of a 2D image. The initial convolutional layer contains 128 filters with a kernel size of 8x8 and strides of 3x3.This can quickly reduce the dimensionality of the input. In the later stage, the model shifts to using smaller kernel sizes such as 5 by 5, and smaller strides such as 3 by 3. This allows for smaller features to be extracted. These convolutional layers are activated using the Rectified Linear Unit (Relu) function. After each convolutional layer, it added a Batch Normalization layer to make a more stable distribution of data in each layer. It standardized size of every output from convolutional layers and smoothen the gradient of the data. This can decrease the internal covariate shift and epoch in a deep learning model. The time of calculating are shortened. After the fourth, twelfth, sixteenth and nineteenth layers. max pooling layers are added. Max pooling layer can help decrease the dimensionality of the output data meanwhile keep the essential data [8]. It only reserves the max value of the data. The first max pooling size is 3 plus 3 to reduce the size of feature map quickly. Size of the rest max pooling layer is 2 plus 2. Those finer details and features will be kept. Those layers would be used to reduce the amount of data in the model, reduce the possibility of overfitting and enhance the stability of the model. Flatten layer used to adjust the form of output from the late max pooling layer to connect to the dense layers. The dense layers both use Relu function to activate. The benefit of Relu is it can learn and express non-liner relationship. Dropout layer is a kind of regularization [9]. It can control the overfitting problem in the CNN model and enhanced generalization capabilities. The first two dropout layers are activated using the Relu function. The last dropout layer is activated using the SoftMax function. The SoftMax function transforms the input into a probability distribution. This model uses this function for classification of images. The construction of VGG19 is displayed in Fig. 3.

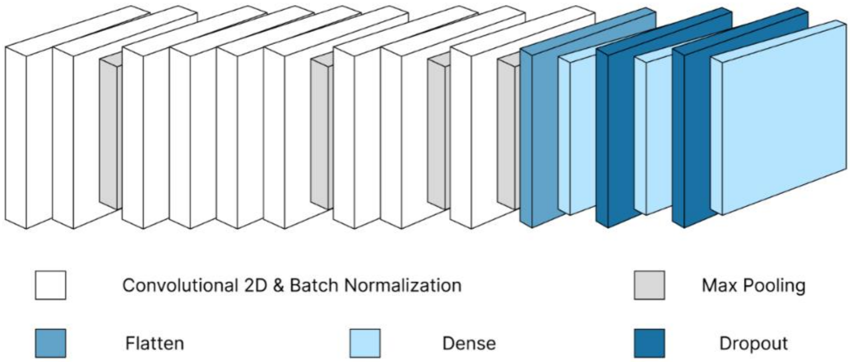

Fig. 3. The construction of VGG19 (Photo/Picture credit: Original).

Loss Function

Sparse Categorical Cross entropy is a kind of loss function which is suited for multiple classification questions [10]. It can be used to determine the gap between the results from model and the real label. It is usually paired with the soft max activation function and applies to the case of integer category labeling, as:

$$Loss = -\,log\,(p[y]) \tag{1}$$

p is the model's anticipated probability distribution. y is integer label for actual items. During training, the loss function measures the discrepancy separating the predicted label from the actual label. After gradient descent, this loss is passed to the model.

2.3 Implementation Details

This model deals mainly with images in mono color. These images have a very high level of detail in gray space. The knee is usually located right in the middle of the picture. In this study, the learning rate is set at 0.001. This model has 17 epochs. With a set learning rate, the accuracy eventually achieves 0.93 from 0.67. The loss decreases from 1.36 to 0.16.

3 Results and Discussion

In this study, the VGG19 model was used for classification and prediction of knee radiographs regarding osteoporosis. More than one thousand four hundred images were used in this study. Each image has been labeled.

3.1 Model Accuracy Analysis

The accuracy of the model was shown in Fig. 4. The accuracy of this model on the training set is depicted by the blue curve, while the accuracy on the validation set is represented by the orange curve. The model in this study underwent 17 epochs of training. As the amount of training epochs rises and the model's precision on the training set gets better, the model undergoes optimization. Accuracy on the validation set serves as a measure of the generalizability of the model. This data is also improving gradually. However, it fluctuates on some epochs. In this study, the accuracy on both the training and validation sets is gradually increasing. This proves that the model is not overfitting for the time being.

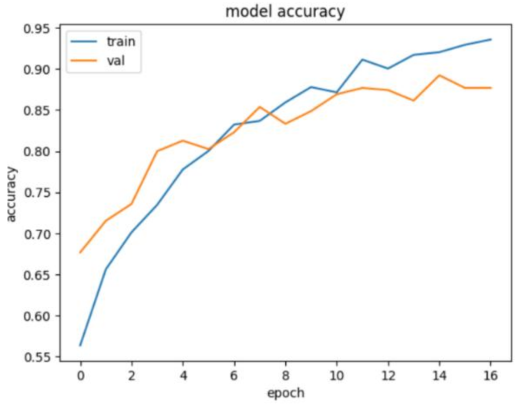

Fig. 4. The curve of model accuracy (Photo/Picture credit: Original).

3.2 Model Loss

Fig. 5 illustrates the loss incurred by the model. The blue curve indicates this model's loss on the training set, while the orange curve indicates the loss on the validation set. The model was trained for 17 epochs. In the first few epochs, both the training loss and the validation loss are higher. Both the training loss and the validation loss diminish throughout the training process. This also indicates that the model is not overfitting for the moment.

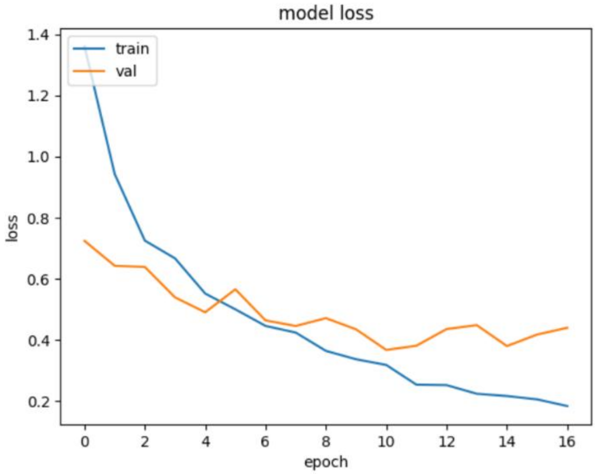

Fig. 5. Model loss (Photo/Picture credit: Original).

3.3 Confusion Matrix Analysis

Fig. 6 shows the performance of this model through the confusion matrix. This figure demonstrates that the current model performs effectively in the prediction and classification of osteoporosis. However, there is still some misclassification for osteopenia.

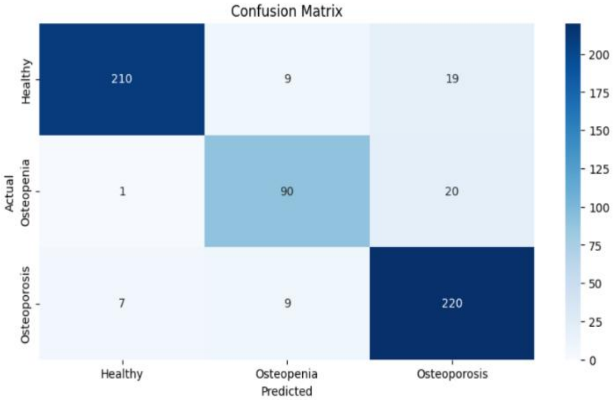

Fig. 6. Confusion matrix (Photo/Picture credit: Original).

As illustrated in Table 1, the false positive rate for the healthy label was less accurate than that for the Osteoporosis label. In contrast, healthy label has lower rate of recall than osteoporosis label. For the F1-score, the healthy label outperformed the osteoporosis label as a whole.

Table 1. The quantitative results of the model on four indicators.

	Precision	Recall	F1-score	Support
Healthy	0.96	0.88	0.92	238
Osteopenia	0.83	0.81	0.82	111
Osteoporosis	0.85	0.93	0.89	236

4 Conclusion

A study shows how to diagnose osteoporosis by using deep learning techniques to identify and categorize knee radiographs. This paper employed the VGG19 model, which features 19 convolutional layers, to analyze the degree of osteoporosis in these images. This model is adept at capturing the intricate features of 2D images and outputs results as probabilities. Notably, the VGG19 model boasts a simple architecture with a minimal number of convolutional kernels and small pooling layers, making it efficient for this task. In order to assess the effectiveness of the proposed method, a set of experiments was performed. The outcomes demonstrated that the model effectively learned the characteristics of osteoporosis in X-rays, achieving an accuracy rate of 89%. Importantly, even after 17 epochs pf training, the model displayed no indication of overfitting. Looking forward, future research will expand to include osteoarthritis as a research objective. The focus will shift to analyzing radiographs of the entire knee area, extending study from the characteristics of osteoporosis to those of arthritis, thereby broadening the scope and potential impact of work.

References

1. Sukegawa, S., Fujimura, A., Taguchi, A.: Identification of osteoporosis using ensemble deep learning model with panoramic radiographs and clinical covariates. 6088 (2022).
2. Pingjun, C.: Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss. Computerized Medical Imaging and Graphics, 75, 84-92 (2019).
3. Kanis., John, A.: A brief history of FRAX. Archives of osteoporosis 13, 1-16 (2018).
4. Hippisley, C., Julia., Carol, C.: Derivation and validation of updated QFracture algorithm to predict risk of osteoporotic fracture in primary care in the United Kingdom: prospective open cohort study. Bmj 344 (2012).
5. Jang, R., Choi, J.H., Kim, N.: Prediction of osteoporosis from simple hip radiography using deep learning algorithm. 19997 (2021).
6. Osteoporosis Knee X-ray Dataset, https://www.kaggle.com/datasets/stevepython/osteoporosis-knee-xray-dataset, last accessed 2021/5/20.
7. Simonyan., Karen., Andrew, Z.: Very deep convolutional networks for large-scale image recognition. arXiv preprint:1409.1556 (2014).
8. Ponnada., Venkata, T., Naga, Srinivasu, S.: Efficient CNN for lung cancer detection. Int J Recent Technol Eng 8(2), 3499-505 (2019).
9. Wager., Stefan., Sida, W., Percy, S.L.: Dropout training as adaptive regularization. Advances in neural information processing systems, 26 (2013).
10. Chaithanya, B.N.: An approach to categorize chest X-ray images using sparse categorical cross entropy. Indonesian Journal of Electrical Engineering and Computer Science, 1700-1710 (2021).