WILEY | Hindawi

Research Article

Explainable Transfer Learning-Based Deep Learning Model for Pelvis Fracture Detection

Mohamed A. Kassem [1], **Soaad M. Naguib** [2], **Hanaa M. Hamza** [3], **Mostafa M. Fouda** [4], **Mohamed K. Saleh** [5], and **Khalid M. Hosny** [3]

[1]*Department of Robotics and Intelligent Machines, Faculty of Artificial Intelligence, Kafrelsheikh University, Kafr el-Sheikh 33516, Egypt*

[2]*Department of Information Systems, Faculty of Computers and Informatics, Zagazig University, Zagazig 44519, Egypt*

[3]*Department of Information Technology, Faculty of Computers and Informatics, Zagazig University, Zagazig 44519, Egypt*

[4]*Department of Electrical and Computer Engineering, Idaho State University, Pocatello, ID, USA*

[5]*Department of Orthopedic Surgery, Faculty of Medicine, Zagazig University, Zagazig 44519, Egypt*

Correspondence should be addressed to Khalid M. Hosny; k_hosny@yahoo.com

Pelvis fracture detection is vital for diagnosing patients and making treatment decisions for traumatic pelvis injuries. Computer-aided diagnostic approaches have recently become popular for assisting doctors in disease diagnosis, making their conclusions more trustworthy and error-free. Inspecting X-ray images with fractures needs a lot of time from experienced physicians. However, there is a lack of inexperienced radiologists in many hospitals to deal with these images. Therefore, this study presents an accurate computer-aided-diagnosing system based on deep learning for detecting pelvis fractures. In this research, we construct an explainable artificial intelligence (XAI) framework for pelvis fracture classification. We used a dataset containing 876 X-ray images (472 pelvis fractures and 404 normal images) to train the model. The obtained results are 98.5%, 98.5%, 98.5%, and 98.5% for accuracy, sensitivity, specificity, and precision.

1. Introduction

The pelvis is a complex and functionally informative bone that contributes directly to human movement and child delivery [1]. The human pelvis is located in the lower abdomen, between the spine and the lower limbs [2]. It comprises the right and left innominate bones, the sacrum, and the coccyx. The innominate bones consist of the pubis, ischium, and ilium. The sacrum and coccyx are part of the axial skeleton and are variably fused vertebrae [3–6], as shown in Figure 1.

Managing patients with pelvis fractures is one of the most complex aspects of trauma care, which occurs in 3% of skeletal injuries [8]. The function of the pelvis is as follows: (1) to protect and support the abdominal and pelvis organs [6], (2) to provide attachment points for muscles, (3) to transmit weight from the upper body to the lower limbs [3], (4) locomotion, and (5) childbirth. As a result, pelvis has great clinical significance to humans [1].

Pelvis fractures, such as osteoporosis, can occur mainly in motor vehicle accidents, sports, or after minor falls in people with fragile bones. Often, pelvis fractures are associated with other severe injuries, which can lead to acute bleeding and damage to the surrounding internal organs and soft tissues [5, 9]. Pelvis fractures are considered a major cause of mortality. According to a study of patients with pelvis fractures in the United States [10], pelvis and abdominal bleeding are mortality's main causes in the first six hours. In addition, the mortality rate in injured patients with pelvis fractures is 5-20% in all emergency cases [11]. Given these factors, diagnosing pelvis injuries should be done urgently without any delay. X-ray imaging is the most
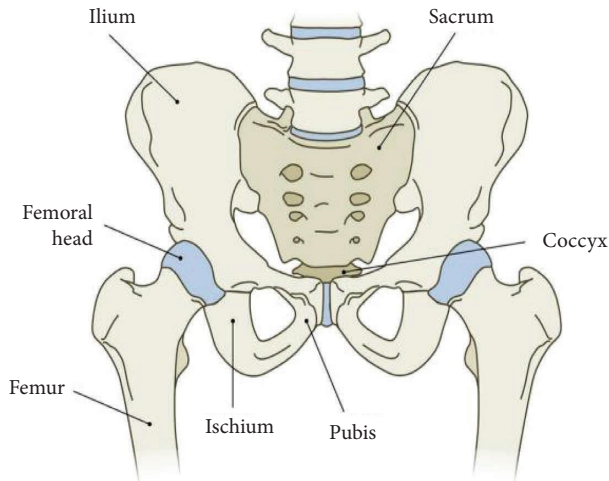
Figure 1: The anatomy of the pelvis [7].

common routine and cheap modality used in emergency units for the early detection of injuries. X-ray imaging should be carefully evaluated to detect any fractures in the pelvis [12]. Inspecting X-ray images with fractures needs a lot of time from experienced physicians. However, there is a lack of experienced radiologists in many hospitals to deal with these images [13]. AI-based systems are widely used to help radiologists and physicians detect fractures. In a recent study, Rainey et al. [14] showed that using an AI-based system reduces 20% of the time radiologists spend reviewing medical images. Therefore, building an AI model to support physicians in interpreting pelvis X-ray images can reduce radiologists' stress, decrease errors, and improve patient care. Despite advances in AI, very limited methods have been proposed for detecting pelvis fractures.

Kitamura [15] used a deep learning technique to identify pelvis fractures on X-ray images, where the results were 0.70 and 0.85 for the posterior pelvis and acetabular categories, respectively. Yamamoto et al. [16] proposed a method for detecting pelvis fractures using 3D CNN on CT images. The test data's accuracy, specificity, recall, and precision are 69.5%, 77.7%, 56.4%, and 61.1%, respectively. Ukai et al. [17] used DCNNs to detect pelvis fractures in CT images, where the AUC was 0.824 with 0.805 recall and 0.907 precision. The AUC with a single orientation was 0.652.

Deep learning has recently become a potential approach for feature extraction from input images using various models. Several neuron layers are utilized to create various layers that extract small information from input images while combining the features of earlier layers. Convolutional neural networks, or CNNs, are these models. We aim to support physicians in diagnosing pelvis injuries, especially in emergencies. Additionally, deep learning models cannot be used for high-risk judgments such as automated pelvis fracture diagnosis because of their black-box nature. An explainable artificial intelligence (XAI) framework with deep learning models is necessary to support deep learning models. A set of procedures and methods known as XAI makes it possible for different users to comprehend and have faith in the outcomes and output produced by machine learning algorithms. XAI could be pre- or post-hoc. Additionally, "model-agnostic" refers to a class of explainers with broad use that are not explicitly created for a particular ML technique.

The main contributions of this paper can be summarized as follows:

(1) Building a new deep learning model based on ResNet50 for detecting pelvis fractures

(2) An XAI model is created using the Grad-CAM framework to explain why a deep learning model predicts pelvic fractures and improves model accuracy, which can raise user confidence and boost the diagnostic system's safety

(3) Validating and evaluating the performance of the proposed model on real-case X-ray images.

This paper is organized as follows: Section 2 provides the methods used for pelvis fracture detection. Section 3 presents the proposed algorithm. Section 4 discusses the results obtained and case study. Section 5 presents the conclusion and future work.

2. Methods

Deep learning-based methods are widely used in medical computer-aided diagnosis systems [18]. ResNet [19], Inception [20], Exception [21], and EfficientNet [22] networks have gained popularity in classifying medical images. Transfer learning improves classification capabilities, especially with small-size datasets [23, 24]. Transfer learning is utilizing a previously learned model to solve a new problem. Furthermore, deep learning models' problematic "black-box" nature necessitates the development of AI that can be explained (XAI). The neural network is known for its categorization task for users and subject-matter experts to examine the many elements.

Additionally, we provide an XAI framework for the pelvis fracture classification problem in the current study employing class activation maps. Ensuring the neural network has acquired the correct characteristics of the many illnesses considered rather than certain local noises in the dataset is crucial. When tested with pelvis fracture other than those present in the dataset, the neural network would fatally misidentify some cases due to its erroneous learning of the local noises in the dataset.

Thus, the suggested XAI framework verifies that the neural network has learned the correct characteristics and increases confidence in its predictions. AlexNet, GoogleNet, and ResNet50 are black-box deep learning models. The transfer learning was applied to these models, and the XAI was used to introduce the trusted model for medical purposes [25, 26].

2.1. AlexNet. AlexNet was the first convolutional network developed by Krizhevsky [27]. AlexNet contains several layers, such as five convolutional layers, two normalization layers, three max-pooling layers, two fully connected layers, and a SoftMax layer. The concept of spatial correlation in an

image frame was investigated using convolutional layers and receptive fields. To increase performance, a GPU was used.

2.2. ResNet50. ResNet is also defined as residual mapping. There are 48 convolution layers, 1 max pool layer, and 1 average pool layer in a ResNet model version called ResNet50 [19]. Shortcut connections are used in ResNet's architecture to address the vanishing gradient issue, as shown in Figure 2.

A residual block, used repeatedly throughout the network, serves as the fundamental ResNet building block. The network learns the mapping from $x\ F(x) + G(x)$, as opposed to $x \longrightarrow F(x)$ alone. The function $G(x) = x$ is an identity function, and the shortcut connection is known as an identity connection when the dimensions of the input x and output $F(x)$ are the same. Since it is simpler to zero out the weights in the intermediate layer during training than to push them to one, identical mapping is learned by doing so. In ResNet, two mapping types were taken into consideration. The input x is padded with zeros to make the dimension match that of $F(x)$, which is the first nontrainable mapping (padding). Trainable Mapping (Conv Layer) is the second way, while $G(x)$ is mapped from x using the 1×1 Conv Layer. The spatial dimensions are maintained or decreased throughout the network, the depth is maintained or doubled, and the product of width and depth after each convolutional layer is maintained.

2.3. GoogleNet. Google's research team proposed Google-Net, also known as Inception V1 [20]. The goal behind the GoogleNet architecture is to have filters of various sizes that may function at the same level. The network gets bigger rather than deeper with this concept. Each inception module can capture different levels of salient features. The 5×5 conv layer captures global features, but the 3×3 conv layer is more likely to capture scattered (distributed) features. The max-pooling operation captures the low-level features that are distinctive in a neighborhood. All these features are retrieved and concatenated at a certain level before being passed to the following layer.

2.4. XAI-Based Methods. Despite the challenge of identifying which features of a model's input drive its decisions, deep neural networks (DNNs) are an essential machine learning technique. Such diagnosis is crucial in various real-world areas, from law enforcement to healthcare, to ensure that appropriate factors for the usage environment influence DNN decisions. As a result, research on the methods and studies that explain a DNN's judgments have grown into a vibrant and expansive field. Competing definitions of what it means to "explain" a DNN's activities and to evaluate an approach's "ability to explain" add to the field's complexity purpose [26].

In deep neural networks, gradients are vectors whose magnitude is the partial derivative of the function $f(x)$ and points in the direction of that function's greatest rate of
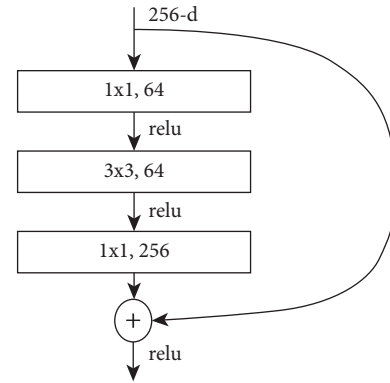

FIGURE 2: Residual learning in ResNet50 [19].

increase. Grad-CAM uses class specifics to produce localization maps of the significant regions of the image based on this information that flows through a generic convolutional network. By displaying visualizations that support output predictions, Grad-CAM makes black box models more transparent. In other meaning, Grad-CAM combines class discriminative capabilities with pixel-space gradient visualization. Grad-CAM can be used with a wide range of CNN architectures, including structured output, multimodel output CNNs, and fully connected layers, such as the AlexNet, ResNet, GoogleNet, VGGNet, and reinforcement learning. So, we used Grad-CAM to explain and visualize the ability of the proposed method to localize the significant region.

3. The Proposed Algorithm

In this research, feature extraction is carried out using the GoogleNet, ResNet50, and AlexNet networks. The ImageNet dataset is used to train these networks. The network layers' filters are used to identify input features, such as colors and shapes.

The pre-trained network is then used to classify various pelvis in a new dataset into fractions and normal. Except for the final three layers (fully connected layer (FCL), SoftMax (SM), and classification), the training parameters from the original pre-trained model are frozen [28].

The network's recently added layers are then trained using the images from the new dataset. In addition, these layers are integrated with the previously trained layers in the pretrained network to classify the new classification classes. Therefore, there are not many newly trained dense layers.

As a result, compared to CNN training from scratch, the training process may be established relatively quickly, and very little training data are required. The new FCL, SM, and classification output layers are subsequently trained using the extracted features [29].

The stochastic gradient-descent method with momentum (SGDM), essentially an enhanced form of SGD with fixed learning parameters, is used for fine-tuning. The SGDM aims to boost velocity across all dimensions, even those with
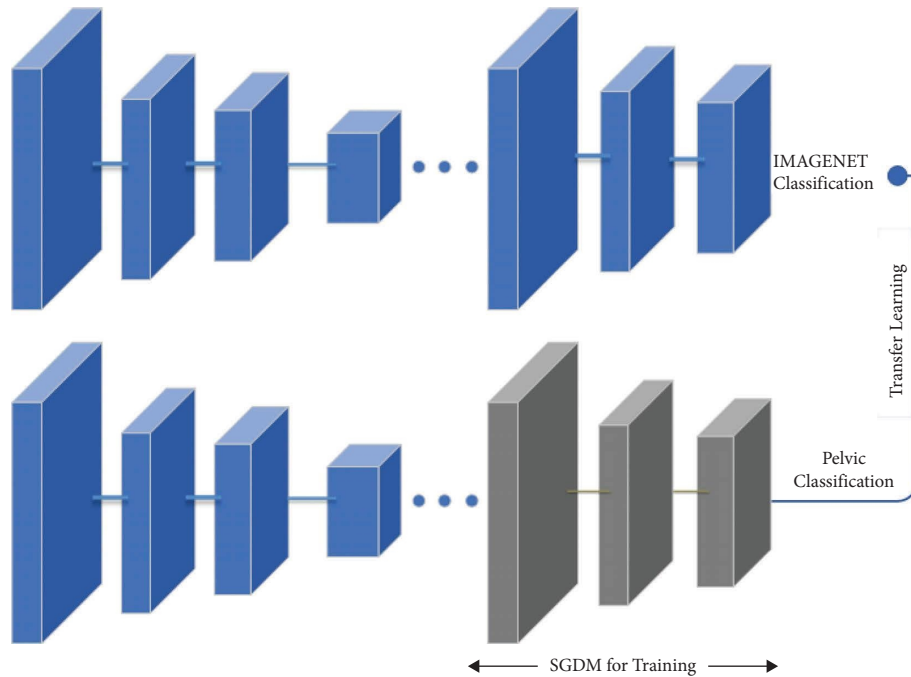
FIGURE 3: The proposed transfer learning model.

constant gradients [30, 31]. All these experiments use the same hyperparameter setting. Figure 3 shows the transfer learning process of GoogleNet, AlexNet, and ResNet50.

3.1. Grad-CAM-Based Method for XAI. Using class activation maps, we create an XAI framework for the pelvis classification problem. Employing the gradient-weighted class activation mapping (Grad-CAM) approach to validate that the proper input pelvic segments are becoming activated while classifying them to their related label. When the network net analyses the classification score for the class indicated by the label, Grad-CAM delivers the gradient-weighted class activation mapping of the change in the classification score of an image *X*. We use this function to validate that your network is focusing on the appropriate areas of a picture and to explain network predictions. The Grad-CAM interpretability technique uses the gradients of the classification score concerning the finished convolutional feature map. The portions of an image that have a significant value on the Grad-CAM map have the most effect on the network score for that class.

4. Experimental Results

An IBM-compatible computer with a Core i7 CPU, 16 GB of DDRAM, and an NVIDIA GeForce MX150 graphics card was used for the research. The application was executed on a MATLAB 2022 (x64-bit). The performance of three distinct transfer learning models, AlexNet, GoogleNet, and RestNet50 was compared to the dataset. The experimental findings and analysis of our models' use of Kaggle-sourced data are presented in this section.

A batch size of 32 models was trained across 10 epochs. Training accuracy, training error, validation error, and validation error were calculated for each epoch. We used a categorical cross entropy loss function and a stochastic gradient-descent technique with a momentum (SGDM) optimizer with a learning rate (LR) of 0.001. A learn rate drop factor approach using was utilized by LR to speed up and bring the optimizer closer to the global minimum.

We dynamically reduced the LR every four epochs based on the validation accuracy to maintain the benefit of a high LR's faster computation time. If the validation loss did not decrease after four epochs and the data was shuffled between each epoch, we decided to cut the LR by 0.1 using the "LearnRateDropFactor" function.

4.1. Datasets. The dataset was obtained from Kaggle [32]. The dataset's name in Kaggle is "ChestPelvisCSpineScans." It contains 876 images and 501 MB in size. The images are organized into two groups. The first group includes 404 normal images (Figure 4). The second group includes 472 pelvis fracture images (Figure 5).

4.2. Results and Discussion. The performance of the proposed method was computed using quantitative and qualitative. Accuracy, sensitivity, specificity, and precision were computed as quantitative measures, while the ROC curve was used as a qualitative measure [33]. Values of false positives (f_p), false negative (f_n), true positive (t_p), and true negative (t_n) are used to compute the following measures from the confusion matrix:
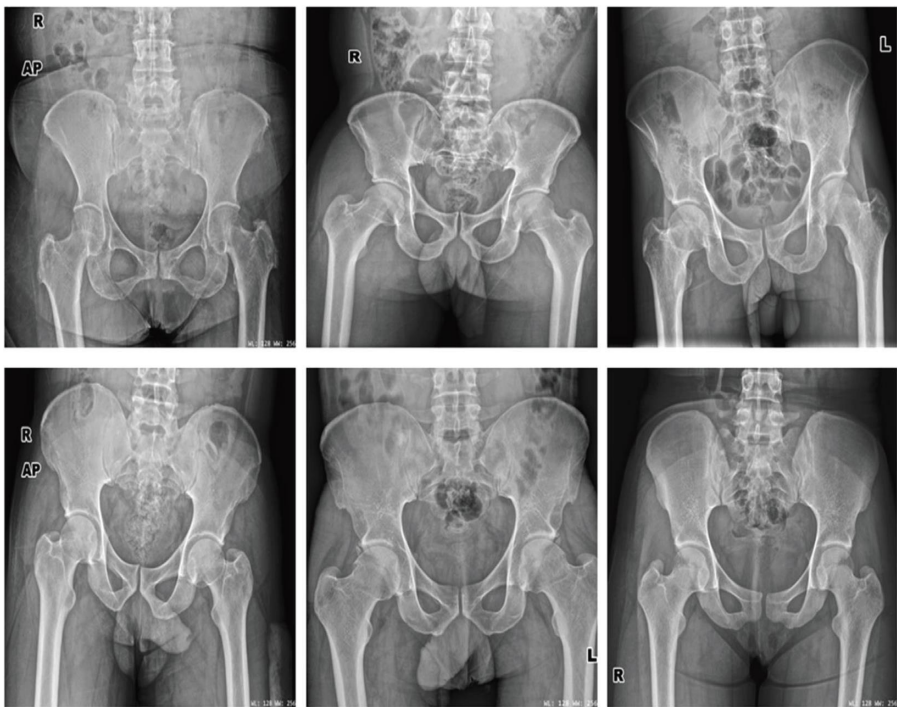
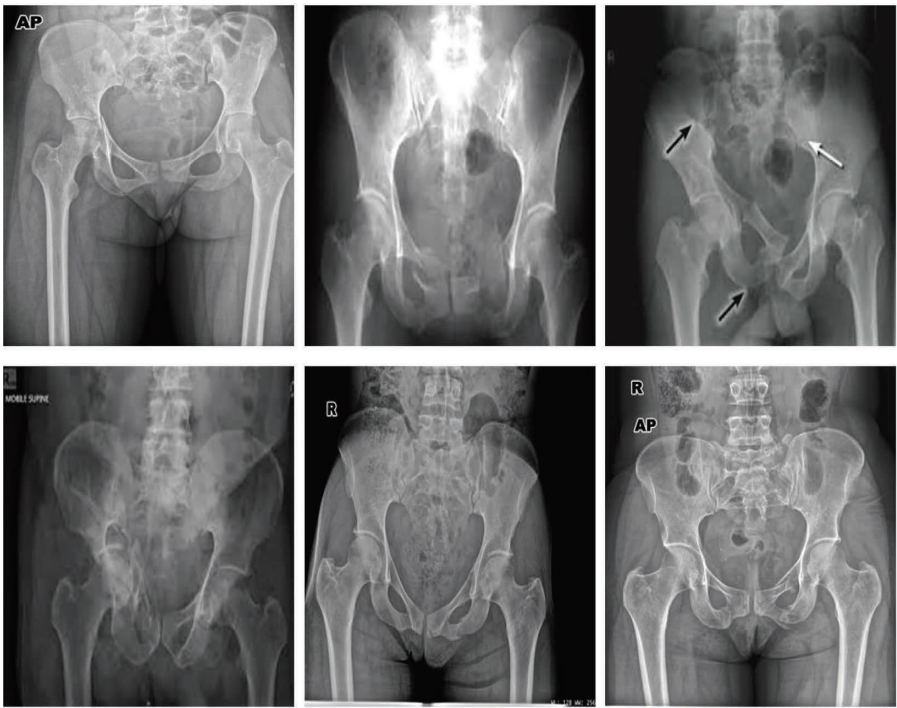FIGURE 4: Samples from the first group (normal images).

FIGURE 5: Samples from the second group (fracture images).

$$\text{Accuracy} = \frac{t_p + t_n}{t_p + f_p + f_n + t_n},$$

$$\text{Sensitivity} = \frac{t_p}{t_p + f_n},$$

$$\text{Specificity} = \frac{t_n}{f_p + t_n}, \qquad (1)$$

$$\text{Precision} = \frac{t_p}{t_p + f_p}.$$

In the first experiment, we removed the last three layers from AlexNet and added new ones for classifying the pelvis into fracture and normal. We resized all images in the dataset to $227 \times 227 \times 3$ to match the width and height of the AlexNet input layer. The dataset is divided into 70%, 15%, and 15% for training, validating, and testing the refined AlexNet.

Due to the class imbalance, each class's performance metrics are calculated separately. The average of these measurements is then determined. Figure 6 shows the confusion matrix for training and testing the refined AlexNet using the pelvis dataset, while Table 1 provides an overview of the average accuracy, sensitivity, specificity, and precision values.

In the second experiment, we adopted ResNet50 by removing the last three layers and adding three layers for the pelvis fracture and normal classification. All images' width, height, and channels have been resized to $224 \times 224 \times 3$ to match the input layer of ResNet50.

Figure 7 shows the confusion matrix for training and testing the refined AlexNet using the pelvis dataset. Because of the imbalance between the class's images, Table 1 provides an overview of the average accuracy, sensitivity, specificity, and precision values. Figure 8 indicates the receiver operating characteristic (ROC) curve for the refined AlexNet and ResNet50.

Figure 9 contains three curves for the proposed models AlexNet, ResNet50, and GoogleNet. These curves visualize the performance measures for the three proposed methods. These curves were plotted by the true positive rate (sensitivity) against the false positive rate (1-specificity). As shown, the performance measures of AlexNet were the lowest, proving the values obtained from the confusion matrix. The performance measures of ResNet50 enhanced more than AlexNet. The ResNet50 curve indicates that the sensitivity and 1-specificity are increased compared to the values obtained from the AlexNet confusion matrix. The final proposal for GoogleNet obtained the best measures, as indicated in the ROC curve. This curve visualizes the obtained values in the confusion matrix.

4.3. XAI Framework. In AI and machine learning, XAI is a new and developing field. Constructing trust among humans about the choices made by artificial intelligence models is vital. It can only be performed by making ML models' black boxes more transparent. Explainable AI frameworks are tools that

FIGURE 6: Confusion matrix of the pelvis classification using AlexNet and transfer learning.

attempt to explain how the model works. These tools generate reports about how the model works.

Deep learning networks are frequently referred to as "black boxes" because they do not provide any means of determining which component of an input to the network was responsible for the prediction made by the network or what it has learned. These models frequently fail spectacularly without warning or explanation when they make incorrect predictions. Class activation mapping is one method for obtaining visual explanations of the predictions made by convolutional neural networks. Mistaken, apparently nonsensical forecasts can frequently have sensible clarifications. We utilized the class activation mapping to see if a certain part of an input image confused the network and caused it to make an inaccurate prediction. Therefore, we utilized Grad-CAM.

The Grad-CAM method, which yields class activation maps, is used to create the XAI framework for the pelvis classification job [26]. Grad-CAM creates a map of weights, highlighting the key areas in the input that the CNN utilized to predict its class label. Grad-CAM leverages the gradient values flowing into the final convolutional layer to create these class activation maps.

Selvaraju et al. [26] describe Grad-CAM in depth. We chose a few samples of pelvis that our CNN properly identified, and then we used Grad-CAM to obtain their class activation maps. Grad-CAM can be slightly modified to produce explanations that indicate support for locations where the network might revise its forecast.

Therefore, removing concepts from those areas would increase the model's confidence in its forecast. This type of explanation is referred to as a counterfactual explanation. Concerning feature maps A of a convolutional layer, we specifically negate the gradient of y^c (score for class c). As a result, the significance weights $\propto_k^c$ are now [26] as follows:

TABLE 1: Quantitative values of the performance measure.

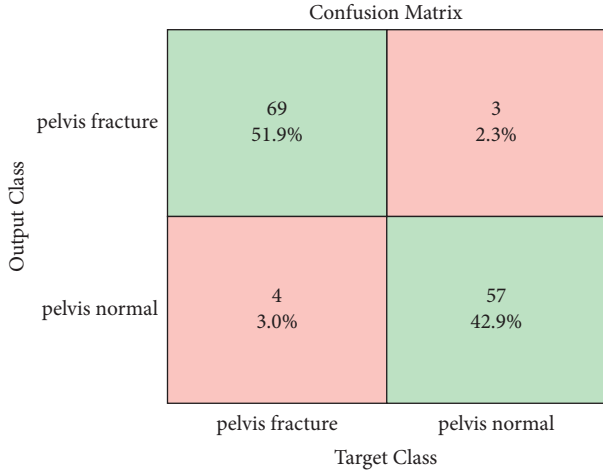| Refined models	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)
AlexNet	54.19	50	50	54.19
ResNet50	94.7	94.5	94.5	95
GoogleNet	98.5	98.5	98.5	98.5

FIGURE 7: Confusion matrix of the pelvis classification using ResNet50 and transfer learning.
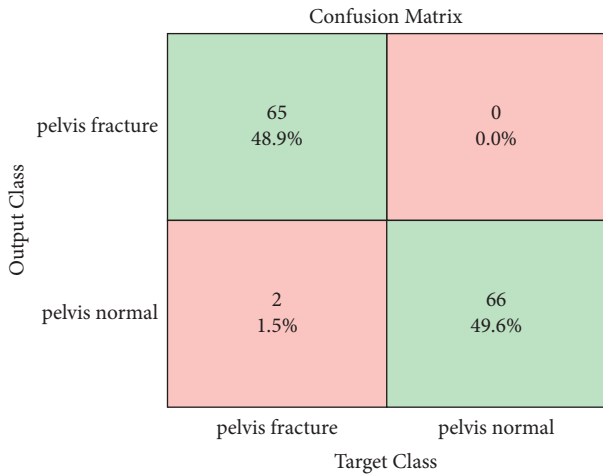

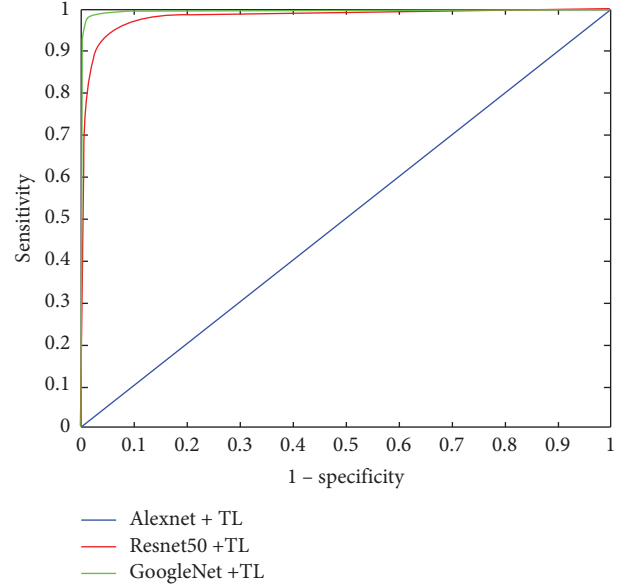FIGURE 8: Confusion matrix of the pelvis classification using GoogleNet and transfer learning.

$$\propto_{k}^{c} = \frac{1}{z} \sum_{i} \sum_{j} -\frac{\partial y^{c}}{\partial A_{ij}^{k}} \qquad (2)$$

Pooling layer

Gradients

FIGURE 9: ROC curve for the three proposed methods.
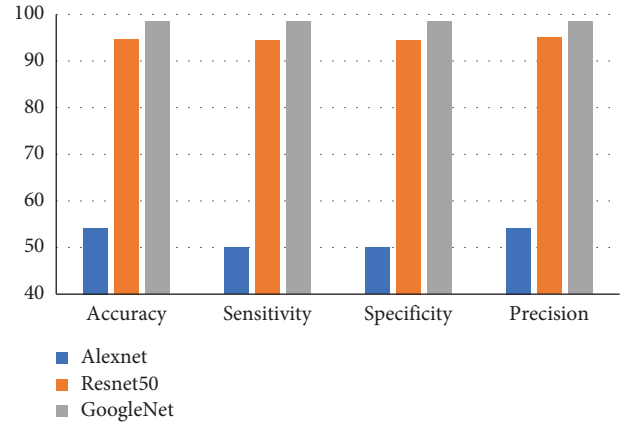
FIGURE 10: Visualization of quantitative measures.

This procedure is shown in Figure 10. The pelvis regions in the image that CNN used to forecast that specific disease successfully were indicated by the average activation obtained. The analysis is done on the average activation along the acquired pelvis segments. Creating an XAI framework for the pelvis classification task is important for medical purposes. Figure 11 shows that the proposed method can detect the pelvis part from the whole image. To evaluate and show how to classify the normal and fractured pelvis using the provided method.

Pelvis normal
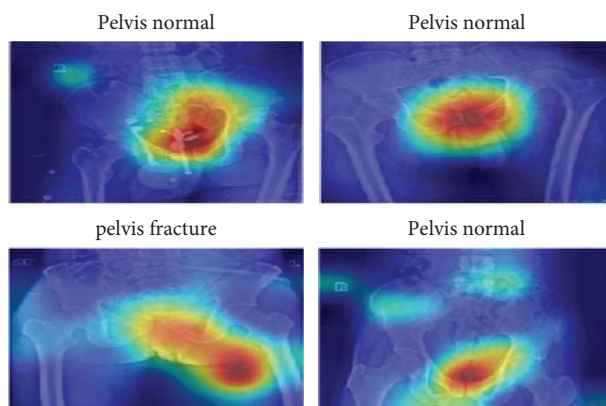
Pelvis normal

pelvis fracture

Pelvis normal

FIGURE 11: XAI Framework for the proposed method using GoogleNet.
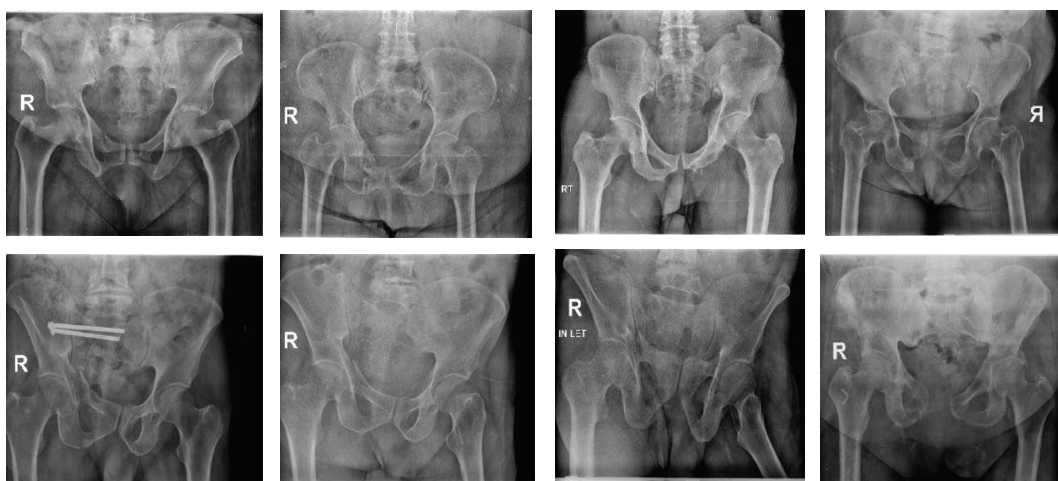

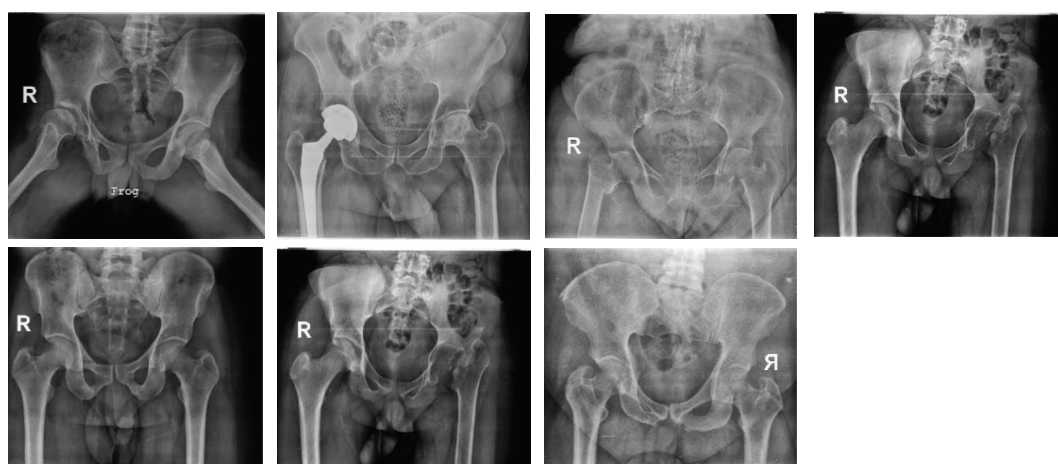FIGURE 12: Normal pelvis X-ray.

FIGURE 13: Pelvic fracture X-ray.

From the XAI of the proposed method, the proposed method can detect the pelvis part correctly and classify the new image.

4.4. Case Study. To evaluate and prove the proposed method's ability to detect pelvis fracture, we collected the X-ray images used in the experiment from the radiology center for 15 cases. It contains 15 X-ray images. Figure 12 shows anteroposterior images of a normal pelvis. The images are considered normal due to the absence of fracture features, loss of bone continuity, fissure lines, or dislocation in the form of loss of pelvic alignment or sacroiliac joint separation.

In conclusion, all the images present a normal pelvis regarding fractures or dislocations. Figure 13 shows anteroposterior views of fracture pelvises of different types. The images considered fracture due to fracture criteria in the form of complete loss of bone continuity and separation of bone ends. We used ResNet50 and GoogleNet to classify these cases. We observed that the proposed method based on ResNet50 and transfer learning fails to classify three classes of pelvis fracture. In contrast, the proposed method using GoogleNet and transfer learning fails only in one normal and fracture case.

5. Conclusions and Future Work

In this study, we proposed an explainable artificial intelligence (XAI) framework for pelvis fracture detection. It has been shown that the proposed technique can be used and provides fast and accurate solutions to the detection of pelvis image (X-ray) fractures. The proposed system aims to support physicians in diagnosing pelvis fractures, especially in emergencies where inexperienced radiologists and physicians cannot deal with these images. We used a dataset containing 876 X-ray images (472 pelvis fractures and 404 normal images) to train the model. The results show an accuracy of 98.5%, a sensitivity of 98.5%, a specificity of 98.5%, and a precision of 98.5%. In the future, in addition to pelvis fracture detection, a system can be developed that can perform classification for major fracture types of the pelvis, such as fracture of the iliac bone, fracture of the sacrum, and fracture symphysis.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] J. DeSilva and K. Rosenberg, "Anatomy, development, and function of the human pelvis," *The Anatomical Record*, vol. 300, no. 4, pp. 628–632, 2017.

[2] G. Gillen, "Trunk control: supporting functional independence," *Stroke Rehabilitation*, pp. 360–393, Mosby, Maryland Heights, MO, USA, 2016.

[3] S. B. Choi and A. A. Cwinn, "Pelvic trauma," *Rosen's Emergency Medicine: Concepts and Clinical Practice*, Elsevier, Amsterdam, Netherlands, 7th edition, 2009.

[4] C. L. Lewis, N. M. Laudicina, A. Khuu, and K. L. Loverro, "The human pelvis: variation in structure and function during gait," *The Anatomical Record*, vol. 300, no. 4, pp. 633–642, 2017.

[5] S. Skitch and P. T. Engels, "Acute management of the traumatically injured pelvis," *Emergency Medicine Clinics of North America*, vol. 36, no. 1, pp. 161–179, 2018.

[6] T. D. White, M. T. Black, and P. A. Folkens, "Pelvis: sacrum, coccyx, and Os coxae," *Human Osteology*, pp. 219–240, Academic Press, Cambridge, MA, USA, 2012.

[7] H. M. De Bakker, M. Tijsterman, B. Kubat, V. Soerdjbalie-Maikoe, R. R. van Rijn, and B. S. de Bakker, "Postmortem radiological case series of acetabular fractures after fatal aviation accidents," *Forensic Science, Medicine and Pathology*, vol. 14, no. 1, pp. 62–69, 2018.

[8] F. Coccolini, P. F. Stahel, G. Montori et al., "Pelvic trauma: WSES classification and guidelines," *World Journal of Emergency Surgery*, vol. 12, no. 1, pp. 5–18, 2017.

[9] W. T. Gordon, M. E. Fleming, A. E. Johnson, J. Gurney, S. Shackelford, and Z. T. Stockinger, "Pelvic fracture care," *Military Medicine*, vol. 183, pp. 115–117, 2018.

[10] R. Vaidya, A. N. Scott, F. Tonnos, I. Hudson, A. J. Martin, and A. Sethi, "Patients with pelvic fractures from blunt trauma. What is the cause of mortality, and when?" *The American Journal of Surgery*, vol. 211, no. 3, pp. 495–500, 2016.

[11] M. J. Weaver and M. Heng, "Orthopedic approach to the early management of pelvic injuries," *Current Trauma Reports*, vol. 1, no. 1, pp. 16–25, 2015.

[12] A. M. Durso, F. M. Paes, K. Caban et al., "Evaluation of penetrating abdominal and pelvic trauma," *European Journal of Radiology*, vol. 130, Article ID 109187, 2020.

[13] Y. Ma and Y. Luo, "Bone fracture detection through the two-stage system of crack-sensitive convolutional neural network," *Informatics in Medicine Unlocked*, vol. 22, Article ID 100452, 2021.

[14] C. Rainey, J. McConnell, C. Hughes, R. Bond, and S. McFadden, "Artificial intelligence for diagnosis of fractures on plain radiographs: a scoping review of current literature," *Intelligence-Based Medicine*, vol. 5, Article ID 100033, 2021.

[15] G. Kitamura, "Deep learning evaluation of pelvic radiographs for position, hardware presence, and fracture detection," *European Journal of Radiology*, vol. 130, Article ID 109139, 2020.

[16] N. Yamamoto, R. Rahman, N. Yagi et al., "An automated fracture detection from pelvic CT images with 3-D convolutional neural networks," in *Proceedings of the 2020 International Symposium on Community-centric Systems (CcS)*, pp. 1–6, Tokyo, Japan, September 2020.

[17] K. Ukai, R. Rahman, N. Yagi et al., "Detecting pelvic fracture on 3D-CT using deep convolutional neural networks with multi-orientated slab images," *Scientific Reports*, vol. 11, no. 1, pp. 11716–11811, 2021.

[18] S. Atasever, N. Azginoglu, D. S. Terzi, and R. Terzi, "A comprehensive survey of deep learning research on medical image analysis with focus on transfer learning," *Clinical Imaging*, vol. 94, pp. 18–41, 2023.

[19] K. He, Z. Xiangyu, R. Shaoqing, and S. Jian, "Deep residual learning for image recognition," in *Proceedings of the IEEE*

Conference on computer vision and pattern recognition, pp. 770–778, Las Vegas, NV, USA, June 2016.

[20] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, Boston, MA, USA, June 2015.

[21] F. Chollet, "Xception: deep learning with depthwise separable convolutions," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1800–1807, Piscataway, NJ, USA, July 2017.

[22] T. Mingxing and L. Quoc, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the International conference on machine learning*, pp. 6105–6114, Long Beach, CA, USA, March 2019.

[23] K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Skin cancer classification using deep learning and transfer learning," in *Proceedings of the 9th Cairo International Biomedical Engineering Conference (CIBEC)*, pp. 90–93, Cairo, Egypt, December 2018.

[24] K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Classification of skin lesions using transfer learning and augmentation with Alex-net," *PLoS One*, vol. 14, no. 5, 2019.

[25] M. A. Kassem, K. M. Hosny, and M. M. Fouad, "Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning," *IEEE Access*, vol. 8, pp. 114822–114832, 2020.

[26] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: visual explanations from deep networks via gradient-based localization," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 618–626, Venice, Italy, October 2017.

[27] A. Krizhevsky, S. Ilya, and H. E. Geoffrey, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 60, no. 6, 2012.

[28] K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Skin melanoma classification using ROI and data augmentation with deep convolutional neural networks," *Multimedia Tools and Applications*, vol. 79, no. 33-34, pp. 24029–24055, 2020.

[29] K. M. Hosny, M. A. Kassem, and M. M. Fouad, "Classification of skin lesions into seven classes using transfer learning with AlexNet," *Journal of Digital Imaging*, vol. 33, no. 5, pp. 1325–1334, 2020.

[30] K. M. Hosny and M. A. Kassem, "Refined residual deep convolutional network for skin lesion classification," *Journal of Digital Imaging*, vol. 35, no. 2, pp. 258–280, 2022.

[31] M. A. Kassem, K. M. Hosny, R. Damaševičius, and M. M. Eltoukhy, "Machine learning and deep learning methods for skin lesion classification and diagnosis: a systematic review," *Diagnostics*, vol. 11, no. 8, p. 1390, 2021.

[32] Kaggle, "Chest, pelvic and C-spine fractures," 2022, https://www.kaggle.com/datasets/pardonndlovu/chestpelviscspinescans.

[33] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.