2023 26th International Conference on Computer and Information Technology (ICCIT)
13-15 December, Cox's Bazar, Bangladesh

Pelvic bone region segmentation (PBRS) from X-ray image using convolutional neural network (CNN)

Syed Alif Ul Alam[1], Saadia Binte Alam[2*], Sourav Saha[3], Mahmudul Haque[4], Rashedur Rahman[5], and Syoji Kobashi[6]
[1,2,3,4]*Department of Computer Science and Engineering, Independent University, Bangladesh, Dhaka 1229, Bangladesh.*
[2*]*Center for Computational & Data Sciences, Independent University, Bangladesh, Dhaka 1229, Bangladesh.*
[5,6]*Graduate School of Engineering University of Hyogo, 2167 Shosha, Himeji, Hyogo, Japan.*
Email- [1]2120015@iub.edu.bd, [2*]saadiabinte@iub.edu.bd, [3]1931045@iub.edu.bd,
[4]mahmud.eece@gmail.com, [5]rashed.riyadh14@gmail.com, [6]kobashi@eng.u-hyogo.ac.jp
[*]Corresponding Author(s)

***Abstract*—Pelvic and hip fractures offer considerable public health risks with high morbidity and mortality rates. Because of the complicated bone structure of the pelvic bone region, detecting fractures is difficult. Though X-ray imaging is routinely utilised for detecting fractures, manual fracture diagnosis is prone to inaccuracies. This paper proposes the use of deep learning algorithms for automated segmentation of the pelvic bone region in X-ray images. In our work, we have investigated U-Net based pelvic area segmentation models with various convolutional neural network (CNN) backbones. The DenseNet121-based U-Net design emerged as the most optimal model, establishing a compromise between performance and computational efficiency. Although it had a modest loss in IoU and F1 scores when compared to InceptionNetV3, it had a remarkable 59.44% reduction in the number of parameters.**

***Index Terms*—pelvic bone, segmentation, deep learning, x-ray**

I. Introduction

Pelvic and hip fractures are a major public health concern around the world, posing a substantial risk of morbidity and mortality [1], [2]. Acting as a crucial link between the lower extremities and the spine, the pelvis encompasses an intricate arrangement of bones, including the pelvic ring and femur bones. Detecting fractures in this region can prove challenging due to the complex architecture of the pelvic bones. Research conducted by Balogh et al. [3] reveals that the occurrence of pelvic fractures amounts to 23 per 100,000 individuals. Disturbingly, patients afflicted with severe pelvic fractures face a mortality rate of 32%, according to the American Association for the Surgery of Trauma [4]. Furthermore, hip fractures have become a prevailing medical issue among the elderly population, with annual reports of approximately 250,000 cases in the United States alone [5]. Detecting fractures within the pelvic ring at an early stage and employing effective medical interventions can significantly reduce the mortality rate for patients [6], [7]. Consequently, a comprehensive diagnostic approach becomes imperative for identifying fractures within the pelvic ring, necessitating the exploration of various methods.

In the medical setting, X-ray imaging is used to assess the pelvic bone region in cases of pelvic injuries. Conventional radiography, however, may not always reveal hip fractures [8]. So, healthcare professionals frequently use Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) to gain a more detailed understanding of the pelvic bone region [9]. Nonetheless, while CT imaging provides a more thorough view of the pelvis, X-ray images are frequently sufficient for an accurate diagnosis in the majority of pelvic fracture patients [10]. CT scans are impractical in cases of severe injuries where rapid medical intervention and stabilisation are necessary due to their time-consuming nature. In such crucial situations, X-ray imaging becomes the preferred modality. Despite the benefits of X-ray imaging, such as its speed, availability, cost-effectiveness, and convenience of use, manually detecting pelvic fractures remains challenging and is susceptible to false-negative detection. As a result, the need for automated methods to address this issue becomes evident.

Deep learning has recently been found to be useful in the field of medical image diagnosis [11]. Olczak et al. [12] demonstrated the diagnostic performance of their deep learning model on X-ray radiographs, exhibiting comparable accuracy to orthopaedic doctors. Numerous studies support the hypothesis that deep learning systems can achieve high accuracy in fracture detection from radiographs [13]–[15]. Furthermore, deep learning has been effectively employed in a variety of diagnostic tasks, including kidney cancer segmentation from CT images [16], brain tumour segmentation from MRI [17], and retina image analysis [18]. Likewise, the application of such techniques is promising in pelvic fracture detection.

Recently, several studies have attempted to use deep learning to detect fractures in the pelvic bone region from X-ray images with great success. Even though some researchers conducted studies that yielded high accuracies in detecting hip fractures in the pelvic bone region [5], [19]–[21], considerably limited research has been done to detect fractures in the pelvic bone region [22]. Cheng et al.

979-8-3503-5901-5/23/$31.00 ©2023 IEEE

[23] initially developed a system that could detect a range of trauma-related fractures from the X-ray image of the pelvic bone region using PelvicXNet. It was able to detect the location of both pelvic and hip fractures from plain pelvic radiographs and achieve state-of-the-art performance.

However, the performance obtained through the current approaches is still not suitable for implementation. While performing fracture detection on human wrist radiographs, the deep learning-based approach by Kim et al. [24] also had similar limitations. To overcome this, Kalmet et al. [25] proposed segmenting the relevant region of interest before fracture detection, which could enhance the results obtained by Kim et al. by eliminating unnecessary data from the pixels outside the area of interest. Taking this into consideration, it is evident that the proposed models for fracture detection in the pelvic bone region could be further optimised if the pelvic bone region is extracted before the model attempts to detect fractures, which has the potential to reduce the impact of noise and false positive results.

Therefore, the objective of this study is to automate the segmentation of the pelvic bone region from X-ray images using deep learning. U-Net based architectures have recently emerged as a highly successful approach for biomedical image segmentation [26]. Furthermore, cutting-edge deep learning models pre-trained on non-medical images have demonstrated their efficacy in extracting complicated features from medical images with significant performance. Hence, this study aims to contribute to the utilisation of deep learning and investigate creative methodologies for automated segmentation of the pelvic bone region using an UNet based system with different variations of CNN-based backbones.

In this paper, we have discussed the development of the Pelvic Bone Region Segmentation (PBRS) system to enable the segmentation of the pelvic bone region. The dataset used for PBRS has been described in Sec. II. Sec. III discusses the development of the PRS model with appropriate settings and the approach to evaluate the performance of the proposed method. Following that, in Sec. IV, we present and discuss the findings of our studies. Finally, Sec. V summarises the findings and offers future study areas in the automated segmentation of the pelvic bone region.

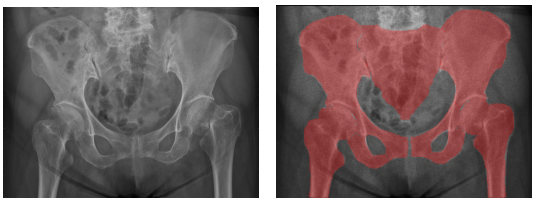
II. Dataset

For our segmentation task, we used 100 pelvic X-ray images from the PXR150 open-source dataset [25]. It contained a total of 150 X-ray images from the pelvic and hip regions. X-ray images from 1888 patients at first. These X-rays comprised 381 instances of only hip fractures, 135 instances of only pelvic fractures, 1334 instances without any acute findings, and 390 mixed instances. 150 X-rays were randomly selected from them to make the PXR150 dataset. The images were grayscale and in PNG format, with pixel values ranging from 0 to 255.

However, these images did not contain any masks for the pelvic bone regions, which are necessary for PBRS. Hence, 100 of these X-ray images were chosen randomly from the 150 X-ray images of the PXR150 dataset, comprising a mixture of instances with only hip fractures and only pelvic fractures and without any acute findings, to annotate and preprocess for training our PBRS model.

A. Annotation

Since the dataset of radiographs does not have any labels or annotations, manual annotation is necessary for segmentation. For this, the Medical Imaging Interaction Toolkit (MITK) [27] has been used to annotate the images. Since the dataset has been prepared for the segmentation task, the annotations, in this case, are the masks for the pelvic bone region. The mask is a binary image with pixel values of 1 or 0 corresponding to the target region or background. In our case, pixel values of 1 denote the pelvic bone region in an image. The masks are stored as Portable Network Graphics (PNG) files, which can be used for our use case after further processing. A sample of the X-ray image along with its mask is shown in Fig. 1.



(a) (b)

Fig. 1: Sample image of Pelvic bone X-ray dataset: (a) original X-ray and (b) PBRS mask.

B. Data Pre-processing

One of the prime considerations for the pre-processing of the data was the varying resolution of X-ray images in the dataset. So, the images have been resized to 224 x 224 for the convenience of training and to maintain constant shape. The 100 X-ray images and their corresponding masks are then separated into a train, validation and test dataset of 70, 10 and 20 images respectively. To prevent overfitting the model during training, we use an image augmentation strategy to increase the size of our training dataset. This method includes randomly applying horizontal flipping, scaling, rotating, shifting, adding Gaussian noise, and shearing, as well as random changes in brightness, contrast, and saturation.

III. Segmentation Model

After the dataset has been prepared for training, we have discussed the development and implementation of the segmentation model for extracting pelvic bone region from X-ray images. Our overall approach has been illustrated in Fig. 2. To develop an effective PBRS model, we have implemented a U-Net based approach followed by its effectiveness in biomedical image segmentation [26].

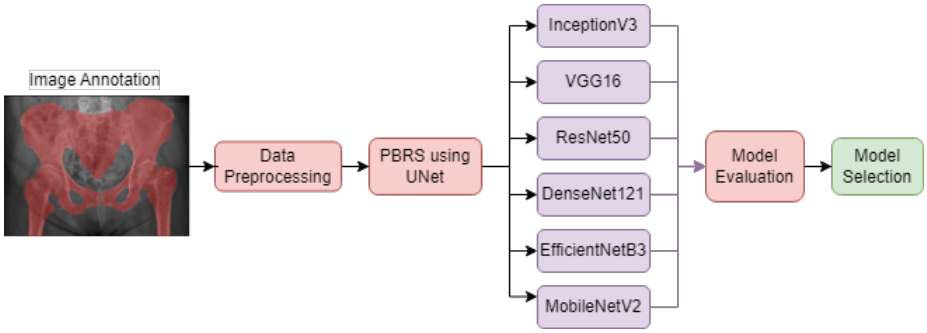


Fig. 2: Overview of the method

A. DCNN Backbones

U-Net is primarily a Deep Convolutional Neural Network (DCNN)-based architecture. With appropriate model settings, our work has explored a wide range of backbone architectures in implementing U-Net for PBRS. Our implementation of the U-Net architecture comprises several robust and popular DCNN architectures, namely: InceptionV3 [28], VGG16 [29], MobileNetV2 [30], DenseNet121 [31], ResNet50 [32], and EfficientNetB3 [33], as its backbone DCNN layers.

1) Inceptionv3: This neural network model was designed for high computational efficiency. To achieve this, its architecture uses both great height and great width. This model uses a module known as the Inception module, where there are parallel convolutional layers with filters of different sizes. The smaller filters help to extract smaller features in the image, and the larger filters help to extract larger features in the image. Besides, the model also uses max-pooling layers, which reduce the dimensions of the feature maps, reducing the computational cost further.

2) Resnet50: Resnet50 solves the problem of vanishing gradients in very deep CNNs, thereby increasing their accuracy. This model is made up of residual blocks from which the network can learn residual functions, which helps optimise the network. The residual blocks use "skip connections," where one or more layers are skipped and information from an earlier layer is added to a later layer.

3) VGG16: VGG16 is a simple deep CNN made up of blocks consisting of multiple 3 x 3 convolutional layers, with each block, followed by a max pool layer. It uses multiple 3 x 3 convolutional layers instead of a single larger filter layer to reduce the number of parameters needed. The number of filters (in a convolutional layer) in each block increases by a factor of 2 after each max pool layer until it reaches 512. The input is a fixed-size 224 x 224 RGB image. In this image, the mean RGB value, calculated from the entire training set, is subtracted from each pixel. This is done so that the pixel values centre around zero, which helps with convergence, thereby improving the model's performance.

4) MobileNetV2: The MobileNetV2 model was built to provide high accuracy on computer vision tasks with limited computational resources on mobile devices and embedded systems. It uses a technique called depth-wise separable convolution. Here, a separate filter is applied to all input channels instead of a single filter as in the standard convolution operation. A 1 x 1 filter is then applied to combine the output channels of this operation to produce a final set of output channels. Using this technique instead of standard convolution reduces the number of parameters drastically while only reducing the accuracy by a minuscule amount. Moreover, the model makes use of two hyperparameters, the width multiplier and the resolution multiplier, to decrease the computational cost further as needed with a small accuracy and latency trade-off. The model also uses inverted residual blocks, which improves accuracy and reduces the computational cost further.

5) DenseNet121: DenseNet is made up of dense blocks consisting of several convolutional layers, where each convolutional layer is connected in a feed-forward style. Here, each convolutional layer uses the feature maps of the previous layers as inputs, and the layer's feature maps are fed forward as inputs to the succeeding layers. In traditional deep CNN, the features learned in earlier layers may not get fully passed to the later layers. The feed-forward architecture in DenseNet solves this problem and prevents learned features from getting lost. Besides, the DenseNet architecture solves the vanishing gradient problem and reduces the number of parameters as well.

6) EfficientNetB3: Scaling the depth, width, and resolution of a CNN increases the accuracy until a certain point, after which the accuracy starts to deteriorate due to the vanishing gradient problem. To solve this problem, Tan et al. [33] proposed an efficient way of scaling a CNN by scaling its depth, width, and image resolution with a constant ratio. However, the results of such scaling are dependent upon the baseline network it is applied to; therefore, Tan et al. [33] used neural architecture search [34] to develop a set of models known as EfficientNets. EfficientNet-B0 can be scaled up several times until EfficientNet-B7, with each successive model having more parameters and giving a higher accuracy. For our purposes, we have found EfficientNet-B3 to be the perfect fit.

B. Model Settings

The pre-trained weights of these DCNNs available through ImageNet [35] have been used for more efficient optimisation during training. Incorporating these backbone architectures into the U-Net architecture, we have trained the

six segmentation models. Adam optimiser has been used with a learning rate of 0.0001 to help our models converge to the minima and a combination of binary focal loss with dice loss has been used as the loss function. Intersection over union (IoU) and F1 score metrics were monitored to identify the optimum model. The models were trained for 100 epochs with a batch size of 10 and the best weights were saved based on their minimum validation loss.

C. Evaluation

To evaluate model performance, we used a separate test set of 20 samples not used during training or validation. We measured the mean IoU-score, and mean F1 score for each model. The Intersection over Union (IoU) score gives a measure of the overlap between two sets of data. It is calculated by dividing the area of intersection between the predicted mask and the ground truth mask by the area of union between them using equation (1). The score ranges from 0 to 1, with 0 indicating no overlap and 1 indicating perfect overlap. The F1 score is obtained by multiplying the area of overlap by two and dividing the result by the combined area of the two images as per equation (2). The resulting score ranges from 0 to 1, where a higher score indicates a better performance of the model.

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{1}$$

$$F_1 = \frac{2 \times \text{Area of Overlap}}{\text{Total area of two images}} \tag{2}$$

IV. Results and Discussion

In this section, we discuss the significance of the results obtained by the six different variations of the U-Net model and the determination of the most optimum model for PBRS. The training and validation performance of the models has been evaluated after each epoch in terms of both IoU and F1 scores as illustrated in Fig. 5. For all models, even though a gap between the training and validation curves is noticed during the initial few epochs of training, the models eventually demonstrated convergence between the training and validation curves after a specific number of epochs. We subsequently selected the models with the minimum validation loss for each architecture to be the optimum models. Moreover, training and validation IoU and F1 scores for these models along with their total number of parameters have been shown in Table. I.

Now, these optimum versions of the PBRS models for each of the six backbone architectures have been further evaluated on test datasets comprising 20 samples. This evaluation

TABLE I: IoU and F1 scores for different CNN backbones.

Backbone	Total Parameters	Training		Validation	
		IoU	F1	IoU	F1
InceptionV3 [28]	29.93 M	0.940	0.969	0.926	0.961
VGG16 [29]	23.75 M	0.884	0.938	0.885	0.938
ResNet50 [32]	32.56 M	0.929	0.963	0.915	0.955
DenseNet121 [31]	12.14 M	0.931	0.964	0.909	0.950
EfficientNetB3 [33]	17.87 M	0.934	0.966	0.925	0.960
MobileNetV2 [30]	8.47 M	0.861	0.925	0.821	0.901

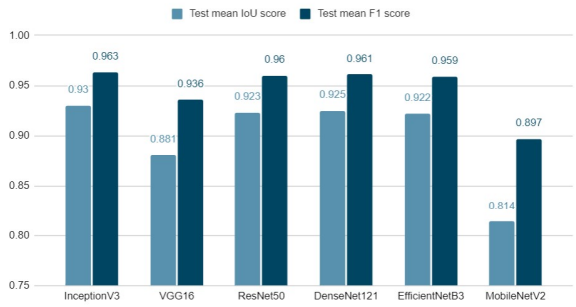
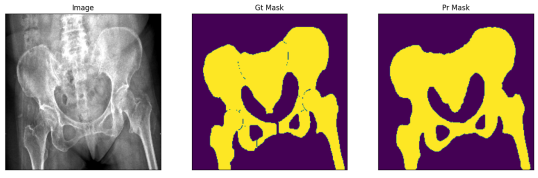


Fig. 3: Performance of U-Net on the test dataset.

demonstrates a significantly high performance of our proposed method in terms of mean IoU and F1 scores which have been visualised in Fig. 3. It is evident from the plot that the InceptionV3-based segmentation model achieved the best results across all evaluation metrics. Following InceptionV3, EfficientNetB3, DenseNet121, and ResNet50 exhibited comparable performances, with negligible discrepancies among their respective scores. A sample of such accurate segmentation of the pelvic bone region and its comparison with the ground truth has been shown in Fig. 4a.

Again, in terms of the number of parameters, the U-Net architectures based on MobileNetV2, DenseNet121 and EfficientNetB3 have the least number of parameters hence lightweight. Our proposed model utilises fewer parameters and is designed specifically for mobile devices with limited computational resources. The simplified and lightweight architecture of MobileNetV2 compromises its ability to capture complex details in the images, resulting in reduced accuracy. On the contrary, the InceptionV3-based PBRS model is the best-performing variation of U-Net which employs more advanced layer structures, such as Inception modules, residual blocks, and dense blocks, enabling them to effectively detect intricate features in the images and requires high computational resources to perform due to its higher number of parameters compared to the other models.



(a)

(b)

Fig. 4: Sample of PBRS Predictions: (a) good and (b) bad

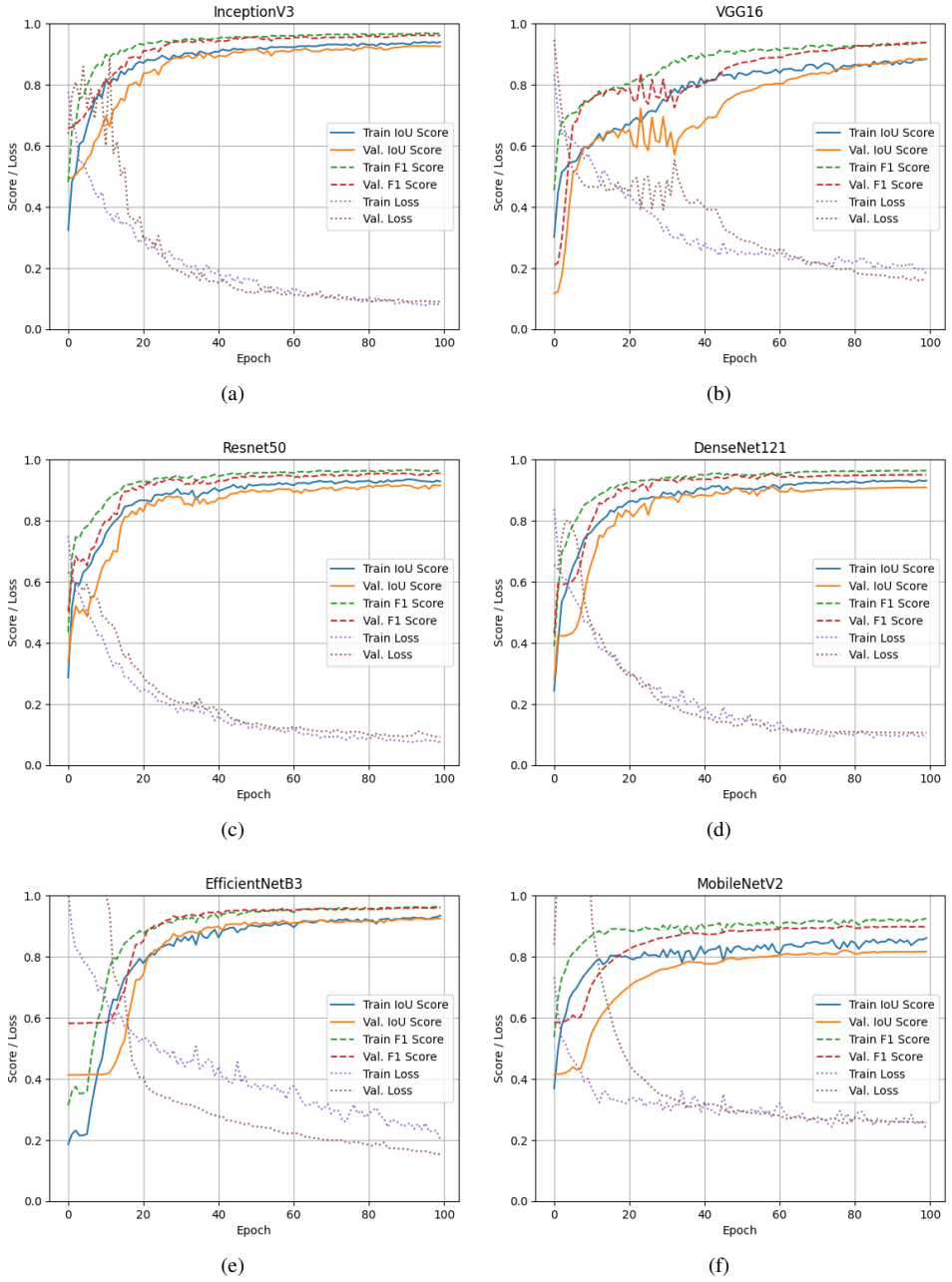


Fig. 5: Training and Validation Performance per epoch: (a) InceptionV3, (b) VGG16, (c) ResNet50, (d) DenseNet121, (e) EfficientNetB3 and (b) MobileNetV2.

Therefore, given the trade-off between performance and computational efficiency, the DenseNet121 architecture is the most optimal variant of our proposed PBRS model. This model demonstrates exceptional performance while maintaining a lightweight structure. Although it exhibits a slight reduction of 0.5% and 0.2% in IoU and F1 scores, respectively, compared to InceptionNetV3, this compromise is justified by the significant advantage of a 59.44% reduction in the number of parameters required by DenseNet121. However, it is important to note that the hazy nature of X-ray images presents challenges, resulting in instances where the segmentation may be noisy, as shown in Fig. 4b. As a result, additional improvements can be investigated to address these limitations and improve segmentation accuracy.

V. Conclusion

In conclusion, this study utilises deep learning to provide an automated approach for segmenting the pelvic bone region from X-ray pictures. The DenseNet121 architecture emerges as the most ideal option, providing exceptional performance

while being lightweight. Despite a minimal decrease in performance scores when compared to InceptionNetV3, the trade-off is justified by a substantial reduction in the number of parameters. The proposed method, which provides accurate segmentation results, meets the demand for automated fracture detection in the pelvic bone region. However, difficulties such as the haziness of images from X-rays continue to have an impact on segmentation accuracy, allowing an opportunity for further advancement. Overall, this study demonstrates the utility of deep learning in the identification of pelvic fractures and provides useful insights for enhancing medical radiographic image analysis.

REFERENCES

[1] Q. Guo, L. Zhang, S. Zhou, Z. Zhang, H. Liu, L. Zhang, T. Talmy, and Y. Li, "Clinical features and risk factors for mortality in patients with open pelvic fracture: A retrospective study of 46 cases," *Journal of Orthopaedic Surgery*, vol. 28, no. 2, p. 2309499020939830, 2020.

[2] M. Katsoulis, V. Benetou, T. Karapetyan, D. Feskanich, F. Grodstein, U. Pettersson-Kymmer, S. Eriksson, T. Wilsgaard, L. Jørgensen, L. Ahmed *et al.*, "Excess mortality after hip fracture in elderly persons from europe and the usa: the chances project," *Journal of internal medicine*, vol. 281, no. 3, pp. 300–310, 2017.

[3] Z. Balogh, K. L. King, P. Mackay, D. McDougall, S. Mackenzie, J. A. Evans, T. Lyons, and S. A. Deane, "The epidemiology of pelvic ring fractures: a population-based study," *Journal of Trauma and Acute Care Surgery*, vol. 63, no. 5, pp. 1066–1073, 2007.

[4] T. Costantini, R. Coimbra, J. Holcomb, J. Podbielski, R. Catalano, A. Blackburn, T. Scalea, D. Stein, L. Williams, J. Conflitti *et al.*, "Aast pelvic fracture study group current management of hemorrhage from severe pelvic fractures: results of an american association for the surgery of trauma multi-institutional trial," *J Trauma Acute Care Surg*, vol. 80, no. 5, pp. 717–723, 2016.

[5] C.-T. Cheng, T.-Y. Ho, T.-Y. Lee, C.-C. Chang, C.-C. Chou, C.-C. Chen, I. Chung, C.-H. Liao *et al.*, "Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs," *European radiology*, vol. 29, no. 10, pp. 5469–5477, 2019.

[6] B. L. Riemer, S. L. Butterfield, D. L. Diamond, J. C. Young, J. J. Raves, E. Cottington, and K. Kislan, "Acute mortality associated with injuries to the pelvic ring: the role of early patient mobilization and external fixation," *Journal of Trauma and Acute Care Surgery*, vol. 35, no. 5, pp. 671–677, 1993.

[7] Y. J. Seong, W. C. Shin, N. H. Moon, and K. T. Suh, "Timing of hip-fracture surgery in elderly patients: literature review and recommendations," *Hip & pelvis*, vol. 32, no. 1, p. 11, 2020.

[8] S. Dominguez, P. Liu, C. Roberts, M. Mandell, and P. B. Richman, "Prevalence of traumatic hip and pelvic fractures in patients with suspected hip fracture and negative initial standard radiographs—a study of emergency department patients," *Academic emergency medicine*, vol. 12, no. 4, pp. 366–369, 2005.

[9] D. K. Hakkarinen, K. V. Banh, and G. W. Hendey, "Magnetic resonance imaging identifies occult hip fractures missed by 64-slice computed tomography," *The Journal of emergency medicine*, vol. 43, no. 2, pp. 303–307, 2012.

[10] J. Young, A. Burgess, R. Brumback, and A. Poka, "Pelvic fractures: value of plain radiography in early assessment and management." *Radiology*, vol. 160, no. 2, pp. 445–451, 1986.

[11] J.-G. Lee, S. Jun, Y.-W. Cho, H. Lee, G. B. Kim, J. B. Seo, and N. Kim, "Deep learning in medical imaging: general overview," *Korean journal of radiology*, vol. 18, no. 4, pp. 570–584, 2017.

[12] J. Olczak, N. Fahlberg, A. Maki, A. S. Razavian, A. Jilert, A. Stark, O. Sköldenberg, and M. Gordon, "Artificial intelligence for analyzing orthopedic trauma radiographs: deep learning algorithms—are they on par with humans for diagnosing fractures?" *Acta orthopaedica*, vol. 88, no. 6, pp. 581–586, 2017.

[13] K. Üreten, H. F. Sevinç, U. İğdeli, A. Onay, and Y. Maraş, "Use of deep learning methods for hand fracture detection from plain hand radiographs," *Turkish Journal of Trauma and Emergency Surgery*, vol. 28, no. 2, p. 196, 2022.

[14] W. Wang, W. Huang, Q. Lu, J. Chen, M. Zhang, J. Qiao, and Y. Zhang, "Attention mechanism-based deep learning method for hairline fracture detection in hand x-rays," *Neural Computing and Applications*, vol. 34, no. 21, pp. 18 773–18 785, 2022.

[15] F. Hardalaç, F. Uysal, O. Peker, M. Çiçeklidağ, T. Tolunay, N. Tokgöz, U. Kutbay, B. Demirciler, and F. Mert, "Fracture detection in wrist x-ray images using deep learning-based object detection models," *Sensors*, vol. 22, no. 3, p. 1285, 2022.

[16] Z. Zhao, H. Chen, J. Li, and L. Wang, "Boundary attention u-net for kidney and kidney tumor segmentation," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2022, pp. 1540–1543.

[17] T. Magadza and S. Viriri, "Deep learning for brain tumor segmentation: a survey of state-of-the-art," *Journal of Imaging*, vol. 7, no. 2, p. 19, 2021.

[18] N. Asiri, M. Hussain, and H. Abualsamh, "Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey. arxiv 2018," *arXiv preprint arXiv:1811.01238*.

[19] J. D. Krogue, K. V. Cheng, K. M. Hwang, P. Toogood, E. G. Meinberg, E. J. Geiger, M. Zaid, K. C. McGill, R. Patel, J. H. Sohn *et al.*, "Automatic hip fracture identification and functional subclassification with deep learning," *Radiology: Artificial Intelligence*, vol. 2, no. 2, p. e190023, 2020.

[20] T. Urakawa, Y. Tanaka, S. Goto, H. Matsuzawa, K. Watanabe, and N. Endo, "Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network," *Skeletal radiology*, vol. 48, pp. 239–244, 2019.

[21] M. Adams, W. Chen, D. Holcdorf, M. W. McCusker, P. D. Howe, and F. Gaillard, "Computer vs human: deep learning versus perceptual training for the detection of neck of femur fractures," *Journal of medical imaging and radiation oncology*, vol. 63, no. 1, pp. 27–32, 2019.

[22] G. Kitamura, "Deep learning evaluation of pelvic radiographs for position, hardware presence, and fracture detection," *European journal of radiology*, vol. 130, p. 109139, 2020.

[23] C.-T. Cheng, Y. Wang, H.-W. Chen, P.-M. Hsiao, C.-N. Yeh, C.-H. Hsieh, S. Miao, J. Xiao, C.-H. Liao, and L. Lu, "A scalable physician-level deep learning algorithm detects universal trauma on pelvic radiographs," *Nature communications*, vol. 12, no. 1, p. 1066, 2021.

[24] D. Kim and T. MacKinnon, "Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks," *Clinical radiology*, vol. 73, no. 5, pp. 439–445, 2018.

[25] P. H. Kalmet, S. Sanduleanu, S. Primakov, G. Wu, A. Jochems, T. Refaee, A. Ibrahim, L. v. Hulst, P. Lambin, and M. Poeze, "Deep learning in fracture detection: a narrative review," *Acta orthopaedica*, vol. 91, no. 2, pp. 215–220, 2020.

[26] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[27] "The medical imaging interaction toolkit (mitk)," accessed on Sept 30, 2023. [Online]. Available: https://www.mitk.org

[28] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[30] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[31] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[33] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[34] B. Zoph and Q. V. Le, "Neural architecture search with reinforcement learning," *arXiv preprint arXiv:1611.01578*, 2016.

[35] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.