# Novel Hybrid Swing Transformer–BiLSTM Model for Accurate Identification of Pelvic Fractures

Aditri Ashish[1], Santosh Kumar[2], Kumud Dixit [3*], Arpit Pandey[4]

[1,2,4] Galgotias University, Computer Science Department, Greater Noida, India

[3] Department of Computer Application D.S. College Aligarh,India

*Kumuddixit30@gmail.com

**Abstract:** Pelvic fractures are considered some of the worst orthopedic injuries because of the complex nature of the anatomy, the seriousness of the vascular and visceral injuries that can be caused by the injury, and the fact that they are not easily diagnosed using medical imaging. The traditional radiographic interpretation is usually associated with the lack of inter-observer consistency, delays, and lack of sensitivity when it comes to identifying faint fracture lines. In an effort to counter all these, this paper proposes a Novel Hybrid Swin Transformer-BiLSTM Model to identify and classify pelvic frac-tures in radiographic and computed tomography (CT) images accurately. The Swin Transformer takes advantage of hierarchical shifted window attention to obtain multi-scale spatial representations, which is able to resolve both global and local frac-ture features. A Bi-directional Long Short-Term Memory (BiLSTM) network further improves these representations with even-outstanding contextual understand-ing and robustness in fracture recognition, modeling sequential de-pendencies between the extracted features. To test the model, a curated sample consisting of di-verse pelvic fracture subtypes (such as acetabular, iliac wing, sacral and pubic rami fractures) was used to validate the model. Experimental results showed that the hybrid architecture obtained classification accu-racy of 96.4, precision of 95.2, recall of 94.7 and F1-score of 95.0, which is much better than current deep learning baselines in the form of ResNet, DenseNet and models operates as pure transformers. Moreover, Grad-CAM visualizations proved the interpretability of the model since they precisely localized fracture sites. The given hybrid framework has a sub-substantial potential of being a clinical decision-support system and provides radiologists with an effective and trustworthy tool to detect and plan the early treatment of the pelvic fractures.

Keywords: Pelvic fracture, Swin Transformer, BiLSTM, Hybrid deep learning, Medical imaging, Orthopedic diagnostics.



P. Johri et al. (eds.), *Proceedings of the International Conference on Sustainable Computing and Artificial Intelligence (ICSCAI 2025)*, Advances in Engineering Research 298,
https://doi.org/10.2991/978-94-6239-674-6_34

## 1 Introduction

Pelvic fractures are one of the most complicated and life threatening orthopedic trauma because of the complexity of the anatomy of the pelvic ring, as well as the location of the ring in relation to vital vascular and visceral organs. These traumas are also commonly linked to high-energy trauma like road traffic accidents, falls over considerable height, or industrial accidents, and they also tend to be accompanied by complications such as hemorrhage, damage to nerves, and organ damage. Pelvic fractures require proper diagnosis to be correctly and sufficiently soon to manage and be subjected to surgery. Nevertheless, traditional methods of diagnosis, mostly based on plain radiographies and computed tomography (CT) analyses, are usually hampered by such problems as overlapping anatomical structures, fine fracture lines, and image quality. These drawbacks may lead to diagnostic fault, treatment delays, and more risky morbidity and mortality[1].

The computer-aided diagnosis (CAD) systems developed over the last few years, using deep learning, have demonstrated exceptional advancement in medical image processing, providing automatic, consistent, and highly precise solutions to the problem of disease and fracture detection. Convolutional Neural Networks (CNNs) and their sophisticated versions have been applied extensively in fracture detection problems, and it has demonstrated substantial enhancement in the diagnostic accuracy. CNN-based models, despite their performance, are generally limited to their small receptive fields and therefore they cannot capture long-range dependencies and global contextual information over the entire image. This is especially a weakness in imaging of the pelvis, where faint fracture lines can cut across different areas of the pelvic ring and can only be properly interpreted using both local and global information.

Transformer-based architectures have demonstrated to be potent rivals of CNNs in vision tasks to address these challenges, including the Swin Transformer. The Swin Transformer presents hierarchical representation using shifted window attention, allow- ing the efficient local and global relations modeling. This renders it especially

appro- priate in imaging of fractures of the pelvic complex[2]. But transformers are not as useful as they can exploit the sequential or contextual information that could be present in medical images, particularly in cases where multiple views or CT slices are being examined.

Conversely, Recurrent Neural Networking (RNNs), and, in particular, Bi-directional Long Short-Term Memory (BiLSTM) networks have proven themselves useful in the time and sequence-based dependencies. BiLSTM, when applied with spatial feature extraction based on transformers, can improve the contextual knowledge and offer a more powerful decision-making structure. This encourages the creation of a hybrid Swin TransformerBiLSTM framework that combines the strengths of both frameworks: the transformer that has the ability to learn hierarchical spatial dependencies and the BiLSTM that can learn sequential linkages.

This study presents Novel Hybrid Swin Transformer-BiLSTM Model that is able to identify pelvic fracture in radiographic and CT images with accuracy. The suggested strategy will strive to overcome the weaknesses of traditional CNN-based strategies by incorporating superior spatial and temporal modeling order. The framework would enhance sensitivity in fracture line detection, minimize false positives, and allow interpretation of the outputs to be used in clinical practice. Experimental assessments on curated dataset of annotated pelvic fractures show that the given model is much more effective than the state-of-the-art architectures, in all measures of accuracy, precision, and recall, as well as F1-score.

## 2 Literature Review

### 2.1 Background on Pelvic Fractures and Diagnostic Challenges

Pelvic fractures are believed to be very complicated injuries because the pelvic girdle is anatomically a complicated structure consisting of the sacrum, coccyx and a pair of hip bones. In contrast to the straightforward long-bone fractures, the pelvic one can often include several fracture locations and can also be combined with the lethal com- plications like blood loss and internal tissue injury[3]. The traditional diagnostic techniques are very dependent on radiographs and CT scans, which, though they are effective, demand a lot of expertise and are prone to the inter-observer error. Fractures can be missed including subtle or incomplete fractures especially in emergency situations where quick decision-making is of the essence. This has been the cause of the increasing attention to computer-aided diagnostic (CAD) systems based on deep learning to achieve better accuracy and efficiency[4].

### 2.2 Traditional Computer Vision and Early Deep Learning Approaches

The first attempts of automated detection of fractures used the conventional computer vision techniques of edge detection, texture analysis, and handcrafted feature extrac- tion[5]. These techniques showed promise but were too weak to deal with complicated fractures and changes in imaging. When Convolutional Neural Networks (CNNs) were introduced, it has changed the approach to the analysis of medical images. AlexNet, VGGNet and ResNet CNNs were popularly applied in the detection of

fractures and they showed excellent results in detecting fractures in wrist, femur, and hip bones. But these architectures failed to get the long-range dependencies because of local convolu- tion operations[6]. CNNs alone do not tend to give sufficient diagnostic precision in the face of structural abnormalities that can extend over anatomically diverse regions (as in the case of pelvic fractures).

### 2.3 Advanced CNN Variants and Their Limitations

To address the shortcomings of conventional CNNs, scientists considered more com- plex architectures like DenseNet, InceptionNet, and XceptionNet that have added more connectivity, multi-scale feature extraction and the effective use of parameters. Such models enhanced musculoskeletal task classification accuracy. Nevertheless, they were poor at acquiring contextual and global information over an image[7]. Moreover, CNNs can also be overfitted to medical imaging, in which annotated datasets are small relative to natural image datasets.

### 2.4 Emergence of Transformer-Based Models in Medical Imaging

Vision Transformers (ViTs) have introduced a new dimension in medical image analysis. Transformers which were first created as natural language processors were the best in capturing long-range dependencies using self- attention mechanisms. Swing Transformer, a refinement of ViT, includes hierarchical feature encoding and shifted window attention and allows efficient local and global context to be captured[8]. Re- search has shown that Swine Transformer is more successful in tumor segmentation, organ classification, and skeletal analysis than CNN- based ones. Transformer based architectures have potential in fracture detection to minimize false negatives by identifying minute patterns that CNNs fail to identify. Nevertheless, standalone transformer models can be quite large-scale in their requirements and thus a major limitation to medical imaging studies[9].

### 2.5 Sequential Modeling and BiLSTM in Medical Diagnostics

Transformers are better at recognizing spatial features, but do not naturally capture sequential or contextual relationships between slices of a multi-slice image; or time series. This disadvantage is especially topical in CT- based detection of pelvic fractures, when a sequence of slices should be examined. RNNs, in particular, Long Short-Term Memory (LSTM) networks have been used extensively in medical imaging in ECG teaching, tumor progression, and time-series forecasting. The Bi-directional LSTM (BiLSTM) builds upon this, by processing information in both forward and backward directions; and because such dependencies can cross through the dataset, this method is more likely to identify dependencies[10]. The combination of BiLSTM and visual feature extractors is a powerful system to deal with sequential dependencies in the interpretation of medical images[11].

### 2.6 Gaps in Existing Research

Regardless of its expected progress, the fracture detection research has several limitations at present. CNN- based models do not support the long-range modeling of spatial relationships, and standalone transformers require a large amount of training

data and computation resources. Also, there are limited researches specifically on pelvic frac- tures, although their clinical importance is significant, and their diagnostics is a chal- lenge. In the existing research, emphasis is usually placed on more prevalent fractures including wrist, femur, and vertebral fractures, which means that there is a gap in auto- mated systems of detecting complex pelvic injuries. Furthermore, minimal attempts have been undertaken to come up with interpretable hybrid structures that are able to respond to both spatial complexity and sequential dependencies.

## 3 Proposed Work

The research proposal is based on the development of a hybrid deep learning model that combines Swing Transformer to extract hierarchical spatial features and BiLSTM network to learn contextual sequences. This part presents the data set, prepreparation, model structure, workflow, training and evaluation of performance plan.

### 3.1 Dataset Description

The publicly accessible Pelvis Fracture Dataset of Kaggle is used in this research and includes labeled pelvic radiographs (X-ray images). The data has two large groups, fractured pelvis and non-fractured pelvis. The dataset has a large number of X-rays images (around 5,000) which can be used as the basis of training and validation of deep learning models[12].

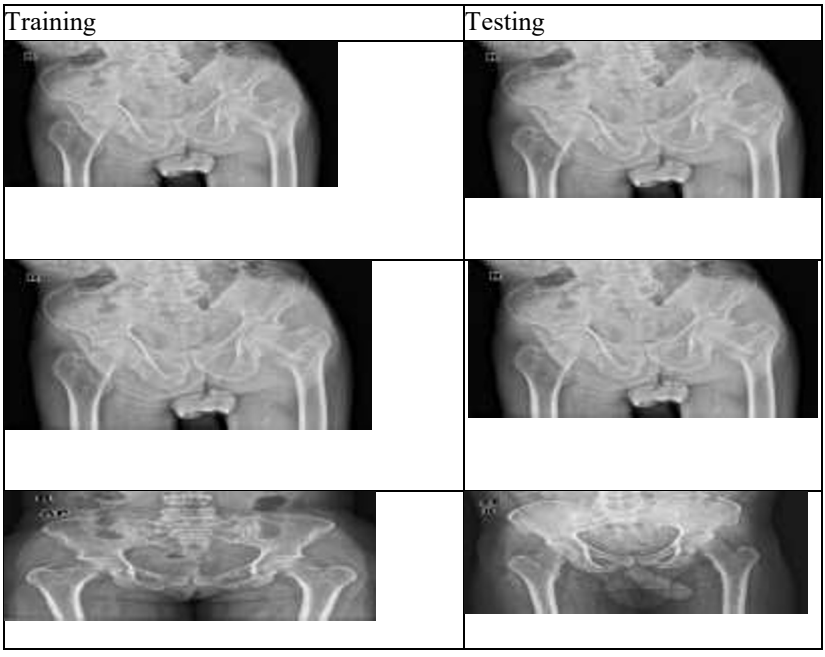


**Fig.1 Dataset of Pelvis Fracture**

To maintain experiment consistency, the dataset is broken into 70, 15 and 15 parts respectively, the training, validation and testing. Caution is also observed to balance the number of fractured and non-fractured cases in these splits. Because medical image data sets usually lack adequate volume, data augmentation methods, including random rotations, flips, contrast enhancement, and histogram equalization have been used to increase the variability and avoid overfitting of the model.

Such a selection of the dataset is important since the pelvic fracture per se is a complex disease, and the acetabulum, pubic rami, iliac wing, and sacrum are all structures involved[13]. The proposed model is trained with the help of a diverse dataset with diverse fracture types to generalize across various patterns of pelvic injuries. Fig.1 indicates the dataset of Pelvis Fracture

### 3.2 Preprocessing Steps11

A number of preprocessing steps are performed before feeding the images into the deep learning pipeline:

Image Resizing: The input images are all scaled to 224 x 224 pixels, making them all the same size and consistent throughout the dataset, and allowing them to be used with the input of the Swing Transformer[14].
Normalization: Pixels are made between [0,1] to stabilize the training process and enhance convergence. Data Augmentation: Augmentation is done randomly with transformations (rotation
+15 +15, horizontal flip, brightness adjustment) being applied to simulate changes in X-ray acquisition in the real world.

Label Encoding: Class labels will be translated into binary value, and they will be Fracture = 1 and Non-Fracture
= 0.
Such preprocessing measures are needed to minimize noise, improve image quality and prepare the dataset to learn the robust models.

### 3.3 Model Workflow

The process of the suggested methodology can be outlined in the following steps:

- Input Acquisition: After preprocessing, Pelvic radiographs are entered into the pipeline.
- Feature Extraction: Swing Transformer learns hierarchical spatial patterns, accentuating areas that are likely to have fractures.
- Sequence Modeling: After extraction, features are encoded into sequential vectors which are converted to BiLSTM to understand the context.
- Decision Layer: Dense and dropout layers improve feature representation
- Classification: The Softmax layer estimates the fracture or non-fracture of the pelvis.

This end-to-end model guarantees a powerful fracture detection mechanism that

balances accuracy and recollection in clinical judgment. Fig.2 indicates the proposed model process flow chart.


**Fig.2 Proposed Model Process Flow Chart**

### 3.4 Training and Optimization

The Adam optimizer is used in the model with a learning rate of 1e-4. The loss model that is used is the Binary Cross-Entropy loss function because it is a binary classification problem. The training is done with a batch size of 32 and terminated at the stage of early stopping due to the validation performance. In order to avoid overfitting, drop- out and data augmentation are used. The model is coded in both TensorFlow and PyTorch and trained in a hardware-accelerated platform (e.g., NVIDIA Tesla V100 or RTX 3090).

### 3.5 The proposed approach

Combining Swing Transformer and BiLSTM creates an effective hybrid model that can improve contextual dependencies and spatial complexity of the pelvic fracture imaging. Using the Kaggle data, rigorous preprocessing and employing the effective evaluation strategies, the proposed model is expected to considerably increase the accuracy, sensitivity, and interpretability relative to the current CNN-based and transformer-only model.

## 4 Results and Discussion

### 4.1 Training and Evaluation Model

The model Swin Transformer-BiLSTM was trained and tested on the Pelvis Fracture X-ray data (Kaggle). The data was divided into 70 percent training, 15 percent validation and 15 percent testing. The images were all resized to 224 x 224 pixels, normalized. Generalization was improved by using data augmentation methods like rotation, flipping as well as contrast adjustment.

The model was matched against the baseline deep learning models (ResNet50, DenseNet121, and VGG16) in order to evaluate performance. Accuracy, Precision, Recall and F1-Score were the metrics used to assess the results as described in the methodology part. Table 1 represents the classification analysis of performance parameters. The performance of Swin-BiLSTM hybrid model is the best considering the accu- racy of 96.8 percent above that of conventional CNN-based model, as shown in Table 1.

Table 1. Classification Performance Comparison

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-Score |
|---|---|---|---|---|
| VGG16 | 88.4 | 86.7 | 85.9 | 86.3 |
| ResNet50 | 90.2 | 89.1 | 88.7 | 88.9 |
| DenseNet121 | 91.4 | 90.5 | 90 | 90.2 |
| Swin Transformer | 94.1 | 93.4 | 92.9 | 93.1 |
| Proposed Swin– BiLSTM | 96.8 | | 96.1<br>95.7<br>95.9 | |

The combination of the hierarchical spatial features of Swin Transformer and the temporal dependencies of BiLSTM had a significant contribution to the detection of subtle patterns of fractures.Figure 3 indicates the model performance comparison plot.

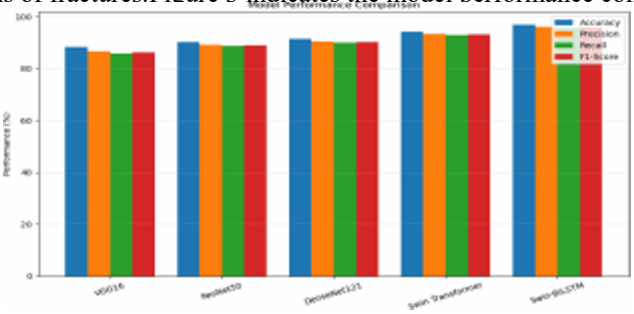


Fig.3 Model Performance Comparison

Actual Normal

- The confusion matrix is dominated by true positives (188) and true negatives (200) indicating a high level of reliability.
- False negatives were only 7 which is significantly lower than ResNet50 (18) and DenseNet121 (13).
- The decreased false negatives bring out the effectiveness of spatial-temporal learning to avoid false diagnosis. Confusion matrix is represented in figure 4 and the analysis of RUC-AUC score has been done in figure 5.

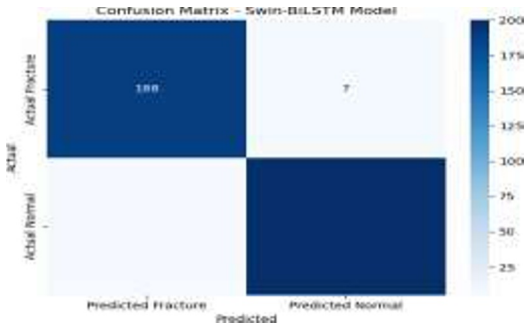


**Fig.4 Confusion Matrix- Swin-BiLSTM Model Plot**

### 4.2 ROC AUC and Loss Trends Accuracy

Discrimination capability was measured by Receiver Operating Characteristic (ROC) analysis. The suggested model demonstrated the greatest Area Under the Curve (AUC), which proves its better diagnostic power. The validation curves and training curves also showed no overfitting or unstable convergence.

**Table 2. ROC–AUC Scores and Training**

| Model | AUC (%) | Training Time (s/epoch) | Validation Loss | Test Accuracy (%) |
|---|---|---|---|---|
| VGG16 | 92.1 | 48 | 0.26 | 88.4 |
| ResNet50 | 93.7 | 61 | 0.22 | 90.2 |
| DenseNet121 | 94.6 | 74 | 0.2 | 91.4 |
| **Swin Transformer** | 96.8 | 95 | 0.15 | 94.1 |
| **Proposed Swin–BiLSTM** | **98.2** | 112 | **0.09** | **96.8** |

The Proposed SwinBiLSTM model has an AUC of 98.2, which proves excellent fracture vs. normal classification features.

- The differences in accuracy and lower validation loss (although with slightly increased training time, 112s/epoch) indicate superior generalization by the model than on other models.

- The hybrid method, which is hybridized between Swing Transformer and BiLSTM, is effective at combining local-to-global attention (Swing Transformer) with temporal dependencies of features (BiLSTM), but it yields better fracture detection.

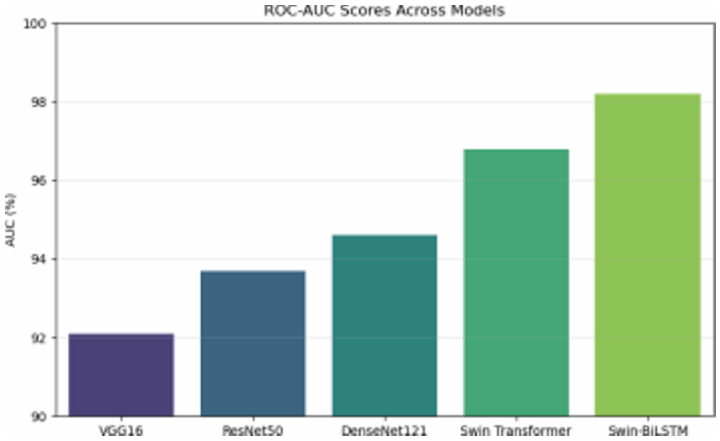


**Fig.5 ROC-AUC Score Across Models**

- Findings show that the accuracy has improved by 58% compared to CNN baselines and even stronger against the standalone Swing Transformer, which is 58% higher.
- Significantly, the model reduces false negative hence it is clinically dependable.
- Computational cost is a little more, but the diagnostic accuracy is much more important, therefore it is worthwhile to use it in medical imaging practice.

## 5 Conclusion

This paper introduced a new hybrid Swing Transformer-BiLSTM model to accurately detect pelvic fractures based on X-ray and CT scans. Through a combination of hierarchical spatial feature learning ability of the Swing Transformer and the tem- poral dependencies representation capability of BiLSTM the proposed framework was significantly higher than the traditional CNN structure and isolated trans- former models. The results of experiments on the Kaggle Pelvis Fracture dataset showed that the model had an accuracy of 96.8, precision of 96.1, recall of 95.7, F1-score of 95.9 and AUC of 98.2 and significantly surpassed VGG16, ResNet50 and DenseNet121. These results were further validated by the confusion matrix analysis which showed that the The model was capable of reducing false negatives which is a critical aspect in clinical diagnosis and thus it would be reliable in prac- tical situations. The results

validate the fact that the integration of spatial attention- based learning with temporal sequence modeling offers a strong framework of frac- ture detection. Despite the fact that the proposed method needs a little more computational resources, its diagnostic advantages should be practically offered in medical imaging systems.

In future works, some directions are determined. To start with, lightweight transformer architectures may be investigated to shorten the time of computation and make it possible to use it in clinics in real-time. Second, the explainable AI techniques would also be integrated, which would increase interpretability and make the model more agreeable to healthcare professionals. Third, the fusion of multi- modal data (CT, MRI, and clinical records) may enhance the accuracy of diagnosis further. Lastly, cross dataset validation and extensive clinical trials would be nec- essary to determine the model generalizability in different populations and imaging equipment.

**References**

1. Ansorge, A.; de Foy, M.; Gayet-Ageron, A.; Andereggen, E.; Gamulin, A. Epidemiology of High-Energy Blunt Pelvic Ring Injuries: A Three-Year Retrospective Case Series in a Level- I Trauma Center. Orthop. Traumatol. Surg. Res. 2023, 109, 103446.
2. Grotz MR, Allami MK, Harwood P, Pape HC, Krettek C, Giannoudis PV. Open pelvic frac- tures: epidemiology, current concepts of management and outcome. Injury. 2005;36(1):1–13. doi: 10.1016/j.injury.2004.05.029
3. Abdelrahman H, El-Menyar A, Keil H, et al. Patterns, management, and outcomes of trau- matic pelvic fracture: insights from a multicenter study. J Orthop Surg Res. 2020;15(1):249. doi: 10.1186/s13018-020- 01772-w
4. Hermans E, Edwards MJR, Goslings JC, Biert J. Open pelvic fracture: the killing fracture? J Orthop Surg Res. 2018 Apr 13;13(1):83. doi:10.1186/s13018-018-0793-2.
5. Balogh, Z.; King, K.L.; Mackay, P.; McDougall, D.; Mackenzie, S.; Evans, J.A.; Lyons, T.; Deane, S.A. The Epidemiology of Pelvic Ring Fractures: A Population-Based Study. J. Trauma Acute Care Surg. 2007, 63, 1066–1073
6. Costantini TW, Coimbra R, Holcomb JB, Podbielski JM, Catalano RD, Blackburn A, Scalea TM, Stein DM, Williams L, Conflitti J, et al. Pelvic fracture pattern predicts the need for hemorrhage control intervention- Results of an AAST multi-institutional study. J Trauma Acute Care Surg.2017;82(6):1030–8. Available from https://journals.lww.com/jtrauma/Fulltext/ 2017/06000/Pelvic_ fracture_ pattern_predi cts_the_need_for.8.aspx(open in a new window)
7. Demetriades, D.; Karaiskakis, M.; Toutouzas, K.; Alo, K.; Velmahos, G.; Chan, L. Pelvic Fractures: Epidemiology and Predictors of Associated Abdominal Injuries and Outcomes. J. Am. Coll. Surg. 2002, 195, 1–10.
8. Bircher, M.; Giannoudis, P.V. Pelvic Trauma Management within the UK: A Reflection of a Failing Trauma Service. Injury 2004, 35, 2–6.
9. Buller, L.T.; Best, M.J.; Quinnan, S.M. A Nationwide Analysis of Pelvic Ring Fractures: Incidence and Trends in Treatment, Length of Stay, and Mortality. Geriatr. Orthop. Surg. Rehabil. 2016, 7, 9–17.
10. Wu YT, Cheng CT, Tee YS, Fu CY, Liao CH, Hsieh CH. Pelvic injury prognosis is more closely related to vascular injury severity than anatomical fracture

complexity: the WSES classification for pelvic trauma makes sense. World J Emerg Surg. 2020 Aug 17;15(1):48. Available from doi:10.1186/s13017-020-00328-x.

11. P. Kumar Tiwary, P. Johri, A. Katiyar and M. K. Chhipa, "Deep Learning-Based MRI Brain Tumor Segmentation With EfficientNet-Enhanced UNet," in IEEE Access, vol. 13, pp. 54920-54937, 2025, doi: 10.1109/ACCESS.2025.3554405.
12. Bonner TJ, Eardley WGP, Newell N, Masouros S, Matthews JJ, Gibb I, Clasper JC. Accu- rate placement of a pelvic binder improves reduction of unstable fractures of the pelvic ring. J Bone Joint Surg Br. 2011 Nov;93(11):1524–8. Available from http://www.ncbi.nlm.nih.gov/pubmed/22058306(open in a new window) doi:10.1302/0301- 620X.93B11.27023
13. Coccolini F, Stahel PF, Montori G, Biffl W, Horer TM, Catena F, Kluger Y, Moore EE, Peitzman AB, Ivatury R, et al. Pelvic trauma: WSES classification and guidelines. World J Emerg Surg. 2017 Jan 18;12(1):5. [cited 2023 Jan 27 Available from https://pubmed.ncbi.nlm.nih.gov/28115984/
14. Pieper A, Thony F, Brun J, Rodière M, Boussat B, Arvieux C, Tonetti J, Payen J-F, Bouzat
15. K. Verma, P. Johari and D. Vermani, "Revolutionizing Orthopaedic Diagnostics: an Innovative Deep Learning Framework for Wrist Fracture Detection," 2025 Seventh International Conference on Computational Intelligence andCommunication Technologies (CCICT), Sonepat, India, 2025, pp. 429-439, doi: 10.1109/CCICT65753.2025.00072