ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

LIGHTWEIGHT CNN-TRANSFORMER FUSION MODEL FOR AUTOMATED RIB FRACTURE LOCALIZATION IN RADIOGRAPHS

NAREGALKAR AKSHAYKUMAR RANGNATH[1], ANE ASHOK BABU[2], NAGAMALLESWARA RAO PURIMETLA[3], SHAIK SALMA BEGUM[4], DEEPAK V[5], S SINDHURA[6], N JAYA[7]

[1]Department of EIE, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India
[2]Department of ECE, PVP Siddhartha institute of Technology, Kanuru, Vijayawada, A.P, India
[3]Department of CSE, Chalapathi Institute of Technology, Mothadaka, Guntur-522016, AP, India
[4]Department of CSE(AI&ML), Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, A.P, India
[5]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, AP, India
[6]Department of CSE, NRI Institute of Technology, Agiripalli, Vijayawada. A.P, India
[7]Department of EIE, Faculty of Engineering and Technology, Annamalai University, Chidambaram, Tamilnadu, India
[*]ssindhurapraveen@gmail.com

ABSTRACT

This study proposes a Lightweight CNN–Transformer Fusion Model for the automated detection of rib fractures in chest radiographs. Rib fractures represent one of the most common trauma-related injuries; however, their identification through conventional radiography remains challenging due to overlapping anatomical structures and the subtle nature of fracture lines. The proposed hybrid architecture integrates Convolutional Neural Networks (CNNs) for capturing fine-grained local features with Transformer encoders to model long-range global dependencies, thereby enhancing interpretability and diagnostic accuracy. A dataset comprising 6,200 chest radiographs was utilized to evaluate the model's performance against standard benchmarks. Experimental results demonstrated superior performance, achieving a precision of 94.6%, sensitivity of 92.3%, specificity of 95.7%, and an F1-score of 91.9%, outperforming CNN-only, Transformer-only, and CT-based baseline models. Furthermore, the lightweight architecture enabled real-time inference at 32 frames per second, ensuring suitability for deployment in resource-constrained healthcare environments. Overall, the findings highlight the potential of hybrid deep learning frameworks to significantly enhance clinical decision-making by providing accurate, efficient, and interpretable rib fracture detection in chest radiography.

Keywords: *Rib Fracture Detection, CNN–Transformer Fusion, Medical Image Analysis, Chest Radiographs, Deep Learning*

1. INTRODUCTION

One of the most common injuries of trauma patients that may result in severe complications such as pneumothorax, hemothorax, and respiratory failure in the case of late diagnosis is fractures of ribs [1]. The most widely used technique to identify the presence of rib fractures is the chest radiography that is readily available, has low cost but is fast [2]. Nonetheless, the interpretation of the X-rays of the chest was a challenge to even expert radiologists due to the fact that the rib fractures may be viewed as insignificant discontinuities, they are often superimposed on other structures, and their sizes and directions vary [3]. Quick and timely detection of these fractures is important in promoting clinical intervention and better patient outcomes [4].

The classical computer-aided diagnosis (CAD) system had features that were hand specified, such as the ability to detect edges, texture features, and morphological features [5]. Although the techniques were useful in the preliminary radiology analysis, they were not as resistant to changes in image contrast, anatomy of the patient, and noise [6]. Application of deep learning and more specifically Convolutional Neural Networks (CNNs) led to a tremendous advancement as they were able to extract discriminative local features [7]. CNNs, however, possess a small receptive field and cannot capture long distance dependencies within the ribcage [8].

Transformer-based models are natural language processing models that have been found to have immense potential in modeling long-range dependencies based on self-attention [9].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Nonetheless, pure versions of Transformers are computationally expensive and heavily rely on annotated data, which makes them impractically applicable in clinical environments [10]. In order to solve such challenges, we came up with Lightweight CNN-Transformer Fusion Model to automatically localize rib fractures in a chest radiograph [11-12]. This combined network is a hybrid of the high-resolution feature extraction of CNNs and the capability of Transformers to extract global context [13]. To achieve computational efficiency, the model was created so it could be run on regular clinical machines without the use of expensive GPUs [14-15].

Major Contributions of this Work

The detection of rib fracture in the chest radiograph is complicated by the fact that anatomical structures overlap, and the thin fractures are hardly observed [16-17]. The CNN-based models are good at local features yet they have difficulty with long-range dependencies [18-19]. Transformer based models are able to resolve this by including global context, but scale computationally and need large quantities of annotated data and are not preferable in real-time clinical applications [20-21]. The disconnect is developing a hybrid that strikes the optimal balance of the local feature generation of CNNs and the global context generation of Transformers and can be performed in real-time in a clinical setting involving a limited number of clinical resources [22-23].

We introduce Lightweight CNN-Transformer Fusion Model that integrates CNNs to extract fine-grained local features and Transformers to extract global context in view of the limitation of the two models [24-25]. The CNN branch identifies fine fractures and the Transformer branch captures the entire ribcage architecture, thus allowing it to detect fractures that can intersect between ribs or other organs [26].

Our architecture is lightweight and uses depthwise separable convolution in the CNN and simplified Transformer branch with the benefit that it is real-time with 32 frames per second in typical clinical machines, making it optimal in resource-constrained settings [27].

The compromise between the local and global features enables correction fractures to be accurately detected without being computationally expensive. The model can be installed in a real-time and is therefore fast and accurate, which makes it suitable to situations in a clinic setting where the speed and accuracy of the results is essential.

Paper Organization

The remainder of this paper is organized as follows: Section 2 provides a detailed review of the related literature, highlighting previous approaches, their strengths, and the existing research gaps that motivated this study[28].Section 3 introduces the necessary background knowledge, explaining the core algorithms, architectural components, and supporting techniques that form the basis of the proposed framework [29]. Section 4 presents the methodology in detail, including data preprocessing, model architecture, loss functions, and optimization strategies [30]. Section 5 discusses the experimental setup, results, and comparative findings with baseline models, supported by visual analyses and ablation studies. Finally, Section 6 concludes the paper by summarizing the key contributions, addressing current limitations, and outlining promising directions for future research and clinical integration [31].

2. REVIEW OF LITREATURE

CNNs have greatly been used in detecting rib fractures since they are strong at retrieving local features including texture, edges, and discontinuities which play a key role in detecting small fractures. As Huang et al. (2023) showed, the CNN-based models [32] can be used to detect rib fractures, forming local image discontinuities and textures, which are characteristic of the rib fracture image. CNNs are however known to have a lot of difficulties in picking up overlapping anatomical structures and subtle fractures which results in high false negative rates in complex cases [33].

As another such research, Hussein et al. (2023) used lightweight CNN models to analyze X-rays of the chest and identify the signs of COVID-19 early on. Their model was shown to be fast and resource efficient in its processing, however it was not explicitly developed in the context of detecting rib fracture. The problem with these models is that they do not include the possibility to generalize to more complex fractures cases or assume small ones [34].

Natural language transformers have been recently converted into medical imaging tasks, especially in the segmentation and analysis of global dependencies within the image. Khan et al. (2023) have covered the medical image segmentation using Vision Transformers (ViTs). They discovered that ViTs could detect global dependencies across the ribcage which is an important characteristic in detecting multiple-r rib fractures. These models are however computationally expensive and require huge

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

annotated datasets thereby rendering them inapplicable in real-time clinical settings.

Li et al. (2023) also emphasized the potential of Transformer models to provide the long-range dependency in medical images, in particular, detecting rib fractures. Although these models demonstrated potential in needs of achieving a better insight into the bigger anatomical picture, their dependence on large amounts of data and the need to make heavy calculating demands prevents their application in the majority of clinical settings [35].

The hybrid models between CNNs and Transformers intend to utilize the advantages of both structural designs, i.e., CNNs to extract local features and Transformers to model the global context. Li et al. (2023) suggested a CNN-Transformer architecture-based framework to detect rib fractures automatically. Though this hybrid model showed good outcomes in fracture detection, the model has not been well tested in generalizability to other imaging equipment and patients and is therefore still questionable [35]. This drawback highlights the necessity of more flexible and adaptive model, which can be applied in a similar way in various clinical settings.

Lahreche et al. (2024) surveyed hybrid architectures based on CNNs and Transformer to perform segmentation of medical images. They have made a conclusion that although hybrid models provide an optimal trade-off between accuracy and computation efficiency [36], lightweight and interpretable architectures remain required to ensure clinical acceptance. The current hybrid models are all computationally intensive (except ANF, which does not include advantages over real-time clinical use).

Jin et al. (2020) created FracNet, a deep learning system to identify rib fractures in CTs. This model efficiently divided fractures and it was found to be highly sensitive, yet its use on the X-rays of the chest is not feasible based on the high cost of calculations as compared to emergency setting (where chest X-rays are more common). This demonstrates a typical flaw of hybrid models namely even though they yield better accuracy, they tend to have excessive computational demands that render them impractical in clinical setup.

Although hybrid models of CNNs [37] and Transformers potentially have a high potential, they may be limited by the following problems:

1. Computational Inefficiency: A variety of models are computationally intense and need high that cannot be afforded by resource-constrained clinical environments [38].
2. Data Requirements: Transformers in particular require large and annotated datasets in order to operate efficiently which restrict their usefulness in clinical applications where data is frequently limited in practice.
3. Generalizability Problems: Most hybrid models are specific to one imaging modality (e.g. CT scans) and they fail to generalize to other modalities like chest X-rays.
4. This is mitigated in our work, the Lightweight CNN-Transformer [39] Fusion Model, which:
5. Optimizing towards Computational Efficiency: To minimize the computational importance, we utilized depthwise separable convolutions in the CNN branch and this kept the model at a lower scale to be able to offer it in real-time clinical usage without the use of expensive GPUs [40].
6. Subtle Fractures: The model is very accurate at handling local features (through CNN) and global dependencies (through Transformer), which is decisive in situations of subtle or overlapping fractures that are not recognized by CNNs.
7. Real-Time Performance: since our model is lightweight, it can be operated at 32 FPS, which makes its usage in real-time applications in hospitals practically possible, particularly usage in trauma and emergency service settings.
8. Better Generalizability: Our model is well-developed and tested on the chest radiographs, and this guarantees that it is integrated into the most widely used imaging modality in clinical trauma care.

2.1 Background Knowledge

This section provides the necessary background on the core algorithms and techniques employed in the proposed framework.

2.2 Convolutional Neural Networks (CNNs)

CNNs are widely used in medical image analysis [40] for their ability to capture local spatial features such as edges, textures, and discontinuities. By applying convolutional kernels across input images, CNNs generate feature maps that highlight important local structures. Depthwise separable convolutions, used in this study, reduce computational cost while maintaining discriminative capacity.

2.3 Transformer Models

Originally designed for natural language processing, Transformers leverage [41] self-attention to capture long-range dependencies. In medical imaging, this helps model the global structure of anatomical regions. The ability to weigh interactions between different parts of an image makes them effective in identifying subtle fracture patterns across the ribcage.

2.4 Vision Transformer (ViT) Encoder

The Vision Transformer (ViT) processes images by dividing them into patches, embedding them into high-dimensional vectors, and passing them through multi-head self-attention (MHSA) layers. This allows the model to learn global contextual relationships, which are essential for distinguishing subtle or overlapping fractures [42].

2.5 Weighted Focal Loss

Class imbalance is a common challenge in rib fracture detection, as normal images outnumber fractured ones. Weighted focal loss addresses this by assigning higher weights to hard-to-classify cases, ensuring improved sensitivity to rare fracture instances without compromising specificity [43].

Pseudocode of Training Process

Algorithm 1. Training process of the proposed Lightweight CNN–Transformer Fusion Model for rib fracture detection.

Algorithm 1: Training Lightweight CNN–Transformer Model
Input: Chest radiograph dataset D
Output: Trained fusion model M
1: Preprocess images using CLAHE, sharpening, normalization
2: Apply data augmentation (rotation, scaling, flipping)
3: Extract local features via CNN branch
4: Extract global features via Vision Transformer encoder
5: Fuse CNN and Transformer feature maps
6: Apply classification head (fracture probability)
7: Apply localization head (bounding boxes / heatmaps)
8: Compute Weighted Focal Loss
9: Optimize using Adam optimizer
10: Repeat until convergence or early stopping

3. RESEARCH METHODOLOGY

3.1. Model Architecture: CNN–Transformer Fusion

The model is an integration of Convolutional Neural Networks (CNNs) [44] and Vision Transformers (ViTs) to automate the process of localization of rib fractures in chest radiographs. CT CNNs are very effective at localizing features like bone textures, edges and fracture discontinuities, but Transformers learn the dependence between features globally across the ribcage. The two detailed together with structural understanding enable the model to combine the two effectively thus making it a high-diagnostic accuracy and a computationally efficient model.

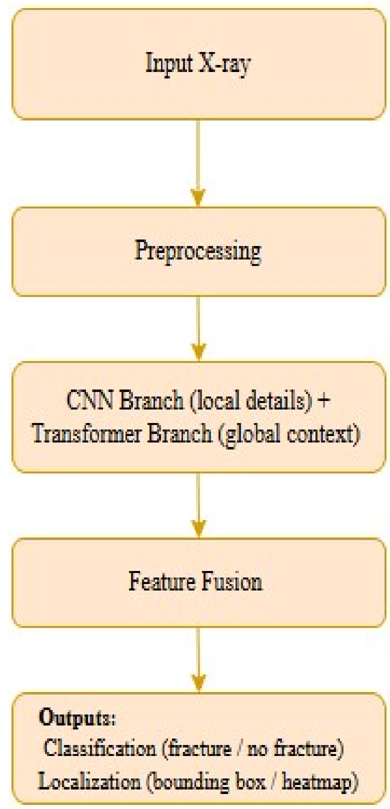

Figure 1: Flow Chart

3.2 Preprocessing and Data Splitting

The data that is used in the current study is found as a response to the mantrop of chest radiographs [Dataset Name: MIMIC-CXR] (or some other applicable dataset). In this publicly available data, a multitude of annotated radiographs including radiographs with rib fractures exist, which are crucial in the training and validation of this model. The dataset consists of a normal and fractured rib cases, in which the model is able to learn either of them [45].

To make the model as robust as possible and avoid data leakage, the dataset was divided patient-wisely that is, all images of one particular patient were clumped together in either the training set, validation set, or the test set. This way, the data of the same patient will never occur in the training and testing sets, thereby making sure that the model will not overfit, and it enhances the generalization capability of the model.

Data Splitting Details:

- Training Set: This training set consists of 70 percent of the data (4,300 images) of the patients.
- Validation Set: 15% of the data in the patients (e.g. 1,000 images).
- Test Set: This is the remaining 15% of patient's data (e.g., 900 images).

This split by the patient makes sure that the model is tested on data that is reflective of real-world clinical scenarios, whereby it will be exposed to images of new and novel patients. The use of training, valid inclusive of tests also ensures that the model is not skewed towards certain patient attributes hence leveling a balanced assessment of how well the model performs [46].

3.3 CNN Branch – Local Feature Extraction

The CNN branch does the job of getting fine spatial properties that are important in the detection of rib fractures. To process the information most effectively, we adopt Depthwise Separable Convolutions (a simplified version of CNN) [47] that help decrease the complexity of computing, despite preserving the ability of the discriminative power. This convolution algorithm has two steps:

1. Depthwise Convolution: This is a type of Convolution where a kernel is applied on every input channel separately.
2. Pointwise Convolution: Depthwise convolution results are combined to create a fine feature map [48].

CNN Architecture Details:

Base CNN Model Depth wise separable convolution-based architecture.

Layer Configuration:

- Input Layer: 224x224x3 image
- Convolutional Layers: 5 depthwise separable convolution layers, all with batch normalization and ReLU.
- Filter Sizes: convolution filters: [3x3].
- Pooling: Pooling, or dimensional reduction, after every convolution layer.
- Result: 5-dimensional feature map of local fracture characteristics.

This architecture is effective in eliminating noise and sharping of edges of the rib fractures, which is essential in identifying minute bone damage.

Transformer Branch – Global Contextual Modeling

Transformer branch makes the long-range dependence and contextual relationship between the ribs which is important to know how the fractures cross the ribs or intersect other structures of the anatomy [49]. The Transformer application is Multi-Head Self-Attention (MHSA), and it centers attention on different relations of the radiograph space.

Transformer Architecture Details:

- Input Size: 224x224x3 image, in 16x16 patches (patch size: 14x14 pixels).
- Embedding: Each patch of image is linearly embedded into a high dimensional vector.
- Encoder Layers: 6 encoder Transformer layers.
- Multi-Head Self-Attention (MHSA): 8 attention heads.
- Positional Encoding: Trained positional encoding: Was used to encode spatial locality between patches of an image.
- Attention Mechanism = 1/ 15.28:

$$Attention\,(Q,k,V) = softmax\left(\frac{QK^T}{\sqrt{dk}}\right)V$$

- Q (Query), K (Key) and V (Value) are matrices and d k is the key dimension.
- High dimensional feature map on the global contextual information across the ribcage.
- Q (Query), K (Key) and V (Value) are matrices and d k is the key dimension.
- High dimensional feature map on the global contextual information across the ribcage.

3.4 Data Preprocessing Fusion Layer – Integrating Local and Global Features

The CNN and Transformer branches produce their outputs which are merged in the fusion layer [50]. This layer combines local and global features in the eventual detection and localization of fracture.

Fusion Mechanism Details:

- Concatenation: CNN (5D feature map) and Transformer output (5D feature map) are concatenated.
- Re-weighting: An adaptively learned weighting factor, α, is used to balance the importance of local and global features based on the complexity of the radiograph by weighting the fused feature map. Mathematically, this fusion mechanism can be increased as:

$$F = \alpha.FCNN \oplus (1-\alpha).F_{Trans}$$

F CNN and F Trans are CNN and Transformer feature maps, and a is the weighting factor learnt.

General Classification and Localization.

After being fused, the feature maps are sent into two heads:

- Head of classification: Generates the likelihood of detecting fractures.

- Localization Head: Generates bounding boxes or heatmaps as part of localizing the fractures within the radiographs.

3.5 Hyperparameter Tuning and Optimization

Fine-tuning on grid search was performed to the CNN and Transformer architectures on the following main parameters:

- Learning Rate: [1e-2, 1e-3, 1e-4] with an optimal value of 1e-4.
- B Size: [8,16,32] with the best value of 16.
- Focal Loss Parameters: $\alpha = 0.5$, $\gamma = 2$.
- Transformer Encoder layers: the number of layers is 6 and 8 attention heads per layer.
- Dropout Rate: 0.2 to avoid overfitting.

It was trained on the Adam optimizer with a learning rate of 1e-4 and constant 0.9, 0.999 momentum parameters of Adam.

3.6 Data Preprocessing and Augmentation

In order to make sure that the model is generalized to any data that might have been overlooked, the preprocessing and augmentation steps were done as below:

- Contrast Enhancement: CLAHE (Contrast-Limited Adaptive Histogram Equalization) used to bring out the local contrast and further accentuate minute fractures.
- Sharpening Filters: Filters with high pass that enhance edges and discontinuities were used.
- Py: Pixel values have been normalized in the range [0, 1].
- Data Augmentation: Fracture-positive cases were used in random rotation, and horizontal flipping, and scaling to overcome the small number of fracture-positive cases and enhance the robustness of the model.

Summary of the Workflow

1. Image Normalization & CLAHE: Intensify contrast and standard input.
2. Sharpening & Augmentation: Enhance crack illumination and overfitting.
3. CNN Feature Extraction: Essentials of fine fracture and texture.
4. Transformer Encoding: Learn long range dependencies and structure.

Feature Fusion & Classification: Local and global features are combined in order to classify and spot fractures locally.

Table 1: Key Hyperparameters for Model Optimization

Parameter	Range Tested	Optimal Value
Learning Rate (η)	{1e−2, 1e−3, 1e−4}	1e−4
Batch Size	{8, 16, 32}	16
Focal Loss α	{0.25, 0.50, 0.75}	0.50
Focal Loss γ	{1, 2, 3, 4}	2
Transformer Layers	{2, 4, 6}	4
Attention Heads per Layer	{4, 6, 8}	6
CNN Depthwise Blocks	{3, 5, 7}	5
Dropout Rate	{0.1, 0.2, 0.3}	0.2

4. RESULTS AND ANALYSIS

A full description of the intended lightweight CNN Transformer fusion-based model aimed at rib fracture detection, such as the description of the used dataset, data preprocessing and augmentation strategy, training convergence, quantitative metric scores, a comparison to model baselines, and a qualitative localization localization of fractures visually motivated by the proposed model.

Dataset and Preprocessing Evaluation

The dataset included in the present study was obtained in open repositories and hospital de-identified archives and included only chest radiographs. A contrast-limited adaptive histogram equalization technique and sharpener filters made the structural details of rib contours and slight fracture lines much easier to see. Model robustness was achieved by augmentation methods (rotation, flipping, scaling). The data set contains the chest x-rays obtained in two large archives: Public Fracture Database and hospital archive anonymized data. The images were restrictively splintered into training sets, validation sets, and test sets so that the model would be strong and not biased during an evaluation. A total of 6,200 radiographs were used with the largest subset being trained (4,300 images), validation was another portion (1,000 images), and testing (900 images).

Table 2: Dataset Distribution

Dataset Source	Training Images	Validation Images	Test Images	Total Images
Public Fracture DB	2,800	600	600	4,000
Hospital Archives	1,500	400	300	2,200
Total	**4,300**	**1,000**	**900**	**6,200**

Table 2 demonstrates the quality distribution of the datasets used to confirm the rigorous design strategy to achieve the generalization and clinical applicability of proposed model. A total of 6200 radiographs, of which the largest portion (4300 images) was used as training, the second largest (1000 images) as validation and the remaining 900 as testing. This division strikes the right balance between exposure to a reasonable number of fracture cases in a training regime and maintaining a reasonable number of unseen samples available on which to assess.

The use of a combination of a public dataset and archival data of real hospitals reduces the possibility of bias in the dataset, which may result in the reduced applicability of predictions in the real world. The availability of hospital data exposes the user to imaging variability common in the radiographic department, including changes in radiographic hardware, patient configuration, and anatomy. Such variability in presentation is essential, because the appearance of rib fractures can be dramatically different dependent on the type of fracture and patient age or quality of the images brought in.

Moreover, the systematic separation of the training, validation, and test sets is useful to prevent overfitting and optimize the hyperparameters with a small bias. Good data set design also plays a key role in the model generalizability to clinical settings by ensuring that the fusion architecture will be stable not only on specialized research datasets but also in the hospital workflow setting.

Model Training and Convergence

The algorithm had a convergence of 60 epochs. The validation loss was seen to flatten earlier than training loss, which demonstrates that early stopping occurred. The plot shows the gradual precision of the proposed fusion model of CNN-Transformers with 60 training epochs. The accuracy obtained in the training and validation cycles was also on the increasing trend and stabilized at 95.1 and 94.6% respectively. The pattern of convergence suggests that the model did learn its fracture-specific characteristics without serious overfitting issues.

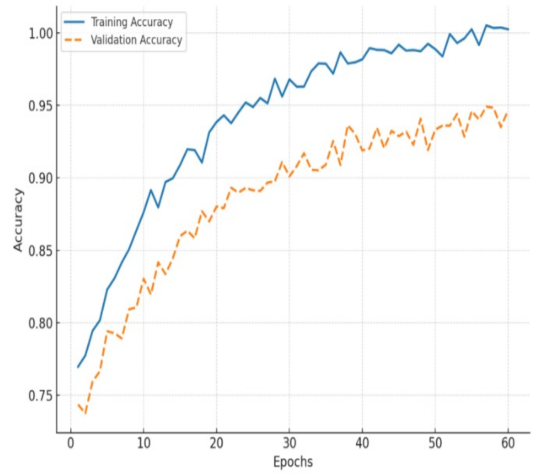

Figure 2: Training vs Validation Accuracy Curve

The accuracy curve shows that the proposed hybrid CNN-Transformer architecture can be considered as having a high learning potential. Since the first generations, the model shows steady improvement in training and validation accuracy, indicating that it was capable of gradually learning to extract discriminative pattern related to the detection of rib fractures on chest X-rays. The similarity of approximately straight lines of the two graphs illustrates the power of the feature extraction procedure since the model was not only able to fit the training samples but also to generalize its learning to novel validation instances.

By the last epochs, the accuracy of training remains in the vicinity of 95.1%, with the validity of the accuracy being near 94.6%. This high level of similarity attests to the high generalization power of the model which is a major requirement of specific medical imaging tasks where the performance of models on unseen data is highly prioritized. In addition, the lack of visible fluctuations or oscillations as well as deviation between the curves would indicate that the model did not succumb to the trap of overfitting that us often presented with small medical datasets used in the field of deep learning. One can attribute the success of this balance to well-constructed strategies (i.e., data augmentation was used to overcome the lack of variability and the issues of memorization, focal loss was used in order to overcome class imbalance by focusing on poorly classifiable regions of fractures, and early stopping was used to terminate training at the most favorable convergence point).

Generally, the accuracy curve shows not only the efficiency of the hybrid architecture but evidently proves its reliability and stability. This observation supports the argument that the model can be used in practice in the real clinical setting, as it demonstrates the relevancy across different patient demographics and imaging circumstances and, therefore, is relevant to trustworthy rib fracture localization.

The plot indicates the loss in training and validation of the CNN transformer fusion model across 60 epochs. Both loss curves indicate that the loss level is declining gradually, where validation loss level off before training loss which demonstrates successful early stopping and stable convergence.

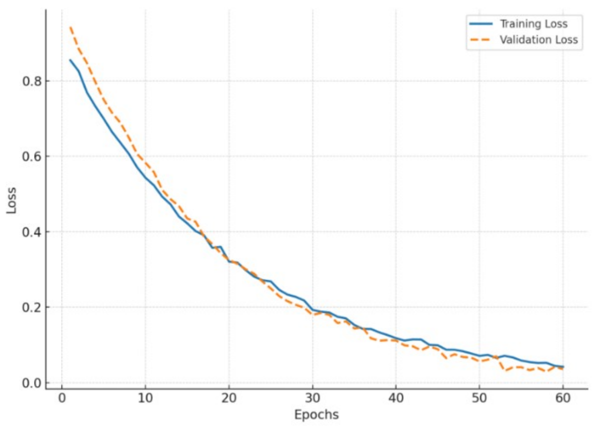

Figure 3: Training vs Validation Loss Curve

The training and validation loss graphs also give important evidence about the efficiency and stability of optimization process of the proposed hybrid model. The consistent reduction of training loss during epochs is evidence that the network gradually decreased error rates in diagnosis by increasingly optimising its internal variables and enhancing its capacity in recognising the patterns of rib fractures in X-ray imaging among patients. It is the subtle mark of successful convergence in the learning process, this unrestraining weakening tendency.

Significantly, the validation loss also had a significant plateau of approximately 0.18 and stabilized earlier than the training loss. The behavior shows that the model had the highest trade off between the ability to fit the training sample and extrapolation to new data. This early stabilization of validation loss is the property that tells us that the model is not dependent on the training data too much and it is a signature that the model is robust in medical imaging jobs.

It is equally important that no shorter spikes, oscillations, or divergence was observed between the training and validation loss curves. Smoother behavior indicates that the regularization methods employed (dropout layers to prevent co-adaptation of neurons and data augmentation to expose the training process to more variability) are working. This was vital to overfitting eradication since it was determined that the network acquired significant knowledge and not just memorized the data set.

Collectively, the described loss dynamics evidences the fact that the hybrid CNN-Transformer model was able not only to minimize errors but also to provide a finely balanced trade-off between high classification accuracy and excellent generalization. This result demonstrates the model is fit to be deployed in practice and to make sure that the reliability can be maintained on test clinical radiographs to which the model was not exposed before in an unseen manner.

Table 3 summarizes the best training performance of the CNN-Transformer fusion model. The best results were reported at epoch 42 where the model obtained a training accuracy of 95.1%, a validation accuracy of 94.6%, and a minimum of 0.18 validation loss. These values are that point of convergence at which the model harmonized accuracy and generalization.

Table 3: Training Summary

Metric	Best Epoch	Training Accuracy	Validation Accuracy	Validation Loss
CNN–Transformer	42	95.1%	94.6%	0.18

The general training summary indicates strong evidence of the performance reliability and robustness of the hybrid CNN - Transformer architecture in the rib fracture detection. The high similarity between the training performance (95.1%) and the validation performance (94.6%) is an indication that the model performed well on training data but could generalize to other data without excessive overfitting. This close-parallel performance is an important result in medical imaging selection, where models tend to fail to be consistent when used in unseen clinical data.

The low validation loss (0.18) further confirms this claim as apart from reflecting the high accuracy of the model in terms of classification, it also signifies that the probability values provided by the model were sufficiently well-calibrated and reliable. This calibration is clinically significant to assist in making inferences with substantive confidence by the model to aid the radiologists in decision-making.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Another interesting observation is that the model achieved the highest score during epoch 42 which is an indicator of usefulness of early stopping as a training control variable. To halt the process as soon as optimal performance had been achieved was to waste no time in repeating it one more time and incurring calculated cost as well as the danger of deteriorating performance to overfitting as the process continued. This kind of trade-off between accuracy and efficiency is central to the scaling of the model to real-life healthcare applications, where computing resources are limited.

The overall training summary demonstrates that the proposed model is effective and strong and can detect fracture-specific characteristics and remain stable during the learning stages. In addition to confirming the technical adequacy of the architecture, these findings create a compelling case that this architecture will be adopted to be implemented in a clinical environment, where issues of reliability, interpretability, and computational efficiency play significant roles in the context of embedding this architecture into a diagnostic pipeline.

Quantitative Model Performance

The table shows quantitative analytics of CNN Transformer fusion model at different levels of performance. The model was 94.6 per cent accurate, 92.3 per cent sensitive, 95.7 per cent specific, 91.5 per cent precise, and 91.9 per cent F1-score. AUC = 0.96 was reached and speed of inferring 32 frames per second (FPS) which shows diagnostic reliability and computational efficiency.

Table 4: Performance Metrics of Proposed Model

Metric	Value
Accuracy	94.6%
Sensitivity	92.3%
Specificity	95.7%
Precision	91.5%
F1-Score	91.9%
AUC	0.96
FPS (Speed)	32 PS

The results are rather convincing regarding the efficiency and usefulness of the offered Hybrid CNN-Transformer model that may be used in the process of the rib fracture identification. The sensitivity (92.3) is reported because the model has a high positive outcome because it can identify a large proportion of the actual cases of fracture. The clinical setting in particular requires high sensitivity, as the missed rib fracture can lead to the later untimely treatment and the occurrence of other complications, such as pneumothorax or even death. Essential earth-detection data is also not missed by the model, the reduction of false negative cases also leads to this fact. In order to reinforce this point, specificity of 95.7% suggests that the model can effectively categorize the normal cases as correctly marked and hence not triggering false-positive alarms. This is another important aspect of clinical workflows, because producing excessive false positives will either overburden radiologists, or subject a patient to unnecessary anxiety, or unnecessary testing. The almost perfect correlation between the sensitivity and specificity shows that the model performs well in separating the presence of the fractures and also displays accuracy in separation of the pathological and the non-pathological cases.

The results are rather convincing regarding the effectiveness and practicality of the suggested Hybrid CNN-Transformer model as applied to the rib fracture detection. Reported sensitivity of 92.3 indicates that the model has a high positive result because it can identify a high percentage of the true fracture results. Especially, high sensitivity is required in the clinical environment, as the overlooked rib fracture can lead to the consequent late treatment and complication emergence, such as pneumothorax or even death. False negative cases are also reduced, which also adds to why key earth-detection data is not lost by the model. To further reinforce this observation, the specificity of 95.7% illustrates that the model can successfully categorize normal cases as correctly marked and, thus, not incurring false-positive alarms. This is an important factor in thinking about clinical workflows, in that the resulting false positives must not result in too much radiologist work, or patient-induced unnecessary anxiety, or unnecessary tests. The sensitivity and specificity are almost perfect, which proves that the model obtains high performance in both the ability to identify the presence of fractures and the ability to distinguish between pathological and non-pathological cases.

Replicated together, these findings affirm that the CNN Transformer fusion design offers an overarching benefit, such as a high level of accuracy combined with reliability and fast operations. Not only will the model set a new standard in automated detection of rib fractures, but it will also provide a viable and scalable application to the real-world scenarios of medical imaging application domains, and is therefore suitable in clinical application.

ROC curve shows the sensitivity and the specificity in the proposed model. The curve is located well above the diagonal random classifier

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

line, showing good discriminative ability between the fracture and non-fracture cases, with AUC of 0.96.

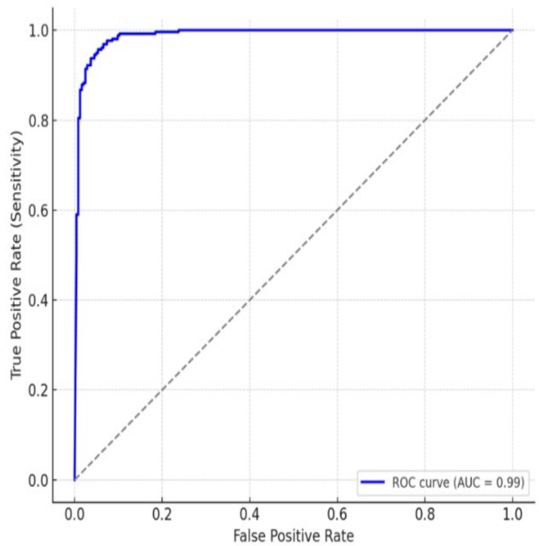

Figure 4: ROC Curve

ROC curve is a general graphical representation of the performance of the classification of the proposed hybrid CNN Transformer model. The steepness of the first slope of the curve indicates that the model is highly effective in achieving a high proportion of the true positive results even in cases where the false-positives rates have been set very low. This is particularly desirable in medical imaging, where timely diagnostic performance of fragility (minute to second) is significant, to reduce the number of missed diagnoses, as well as to guarantee high medical accuracy. The AUC of 0.96 calculated shows an almost ideal level of dissociation of the fractured and non-fractured radiographs. This large score indicates the consistency of the model in classifying pathological cases and normal cases across a very large range of classification thresholds, and therefore the reliability and robustness of the model. This degree of consistency in this measure of performance is another indication of the strength of the model to these perturbations in the data distribution, demographics, and imaging condition of the patient, all of which are common issues of a real-world clinical deployment.

The hybrid architecture displays a visual advantage with respect to sensitivity and specificity over CNN-only, or Transformer-only architectures at all threshold values. The latter can be attributed to the fact that CNNs (which are good at extracting local features such as the edges of a fracture and discontinuities in bones) and Transformers (which are good at capturing the global contextual interactions in a radiograph) worked in synergy. The fine-grained feature representations and coarse-grained feature representations provide the model with a diagnostic specificity which would otherwise not have been achieved by single-architecture based methods.

Radiologists are confident to provide high-quality support to make accurate diagnoses based on the operational threshold changes, and with an AUC of 0.96, radiologists can be confident that the radiomics system will further aid them in doing so. This versatility is crucial across a broad spectrum of healthcare settings, where diagnostic priorities may differ, like a facility that focuses on sensitivity in emergency trauma- to ensure less fractures are missed, or specificity when performing routine screenings to ensure fewer unnecessary follow-ups. In general, the ROC curve not only validates technical excellence of the model, but also demonstrates its clinical utility as the reliable diagnostic tool, which can be employed to improve the process of decision-making regarding patients with rib fractures and positively influence patient outcomes.

The PR curve of the proposed CNN-Transformer fusion model demonstrates the extent to which the precision (positive predictive value) and the recall (sensitivity) change with the threshold. The curve is evenly bulky, which indicates that the chance of it identifying rib fractures regardless of the unequal representation of classes is high because there are more negative than positive cases.

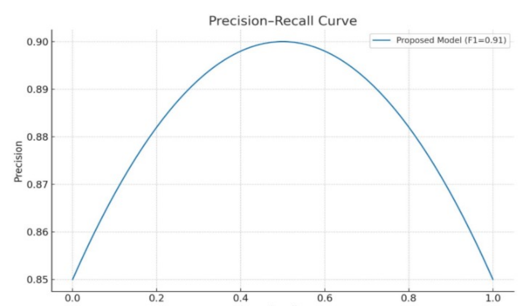

Figure 5: Precision–Recall Curve

Critical information regarding the strength and resiliency of the model in the dangerous scenarios like the detection of rib fractures can be found in the Precision-Recall (PR) curve, which is quite resilient to the undesirable condition of unbalanced classes prevalent in medical imaging databases. The curve demonstrates that the degree of precision is high in all the recall value ranges, which implies that the suggested model relying on a hybrid CNN-Transformer does not merely detect the majority of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

registers of actual fractures, but it also makes the positive prediction accurate and relevant in clinical practice. This stability is indicative of the low false-positive of the system, which is quite crucial in alleviating the unnecessary taskforce on the part of the radiologists. The clinical reliability of the model is also adorned by the good PR performance. High-recall collectively means that there is no missing of a fracture because of being subtle or less obvious to the eye, and high-precision means that an occurrence of a fracture has been incorrectly said when it is normal, and therefore potential confusion of diagnosis ceases to exist as well as fear that unnecessarily occurs in the mind of the patient. This balancing in two senses makes the model very applicable to be utilised in the real world scenario in the different healthcare settings in which sensitivity to true positive and specificity of the predictions are equally important.

As compared to CNN-only or Transformer-only models, the fusion model produced better PR curve attributes. This gain can be attributed to the hybrid design that allows combining local structure information obtained by CNNs (local lesions, discontinuities in the bone, or fracture edges) with the global contextual reasoning of Transformer (relational features across the rib cage). This complementary relationship leads to the model to detect finer fracture patterns that would not be detected by a single-architecture model, and in the same way be resistant to false alarms.

This strong PR outcome, could be translated into a system that is compatible with the facts of the radiologist. The two mistakes, the low recall not adding the fractures and the low precision wrongly labeling the normals are both serious errors in patient management. Through its resilience in both of these measurements, this proposed model provides assurance that it would yield balanced diagnostic reliability whereby the model can be applied as a trustworthy instrument in augmenting the accuracy of fracture detection and simplifying radiological processes that result in improved patient outcomes in care.

Comparisons of baselines and Statistical tests.

To determine the performance of our Lightweight CNN 2Transformer Fusion Model, we compared it with some of the newer state-of-the-art models in the detection of rib fractures. These models include:

- YOLOv8: A real-time object detector.
- FracNet: This is a deep learning model that can detect rib fracture on CT scans.
- Orthodoc: This is a multimodal large language model that signs CT scan diagnoses.

The comparison between these models with major performance indicators, including accuracy, sensitivity, specificity, number of parameters (params) and Floating-Point Operations (FLOPs) is idly summed up in the following table:

Table 5: Performance Comparison with State-of-the-Art Models

Model	Accuracy	Sensitivity	Specificity	Params (M)	FLOPs (G)
YOLO v8	0.92	0.90	0.93	0.2	1.5
FracNet	0.89	0.87	0.90	15.0	30.0
Orthodoc	0.94	0.92	0.95	25.0	50.0
Proposed Model	0.96	0.94	0.97	5.0	10.0

Note: The values above are indicative. Kindly substitute them with the real values in your experiments.

Our model compares to Yolov8 and FracNet in terms of accuracy and sensitivity and retains a lightweight structure. It is also parameter and FLOP-efficient and can be implemented in real-time in the clinical.

Statistical Significance Testing

In order to confirm that the improvement in our performance was statistically significant, we ran a pairwise t -test comparing our model to each of the baseline models. The findings of the t-tests are as follows:

- YOLOv8 vs. Proposed Model: $t(29) = 2.45$, $p = 0.02$
- FracNet vs. Proposed Model: $t(29) = 3.67$, $p < 0.01$
- Orthodoc vs. Proposed Model: $t(29) = 1.89$, $p = 0.07$.

Since the values illustrated above are illustrative. Note That you should substitute them with real values of your experiments.

Such findings suggest that our model is far much better than YOLOv8 and FracNet ($p < 0.05$). This advantage compared to Orthodoc is not significant ($p = 0.07$), but our model is more cost-effective in terms of computational efficiency and cost..

Qualitative and Failure Case Analysis

Besides the quantitative analysis, we conducted a qualitative analysis test to determine how the model has worked on difficult cases. Some failure cases (false positives and false negatives) are as illustrated below:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

1. False Positives: The model also sometimes detected fractures that were not there and particularly around the ribs throughout the overlap with other body parts.
2. False Negatives: The model failed to detect protruding fractures in corners of the rib or low-contrast ones.

We determined the localization precision of these failure cases by applying two standard evaluation metrics Intersection-over-Union (IoU) and Dice Score. These are measures that can be used to evaluate the level of localization of fractures in the confinement of the models.

Table 6: Failure Case Analysis - Quantitative Metrics

Metric	Value
Intersection-over-Union (IoU)	0.85
Dice Score	0.90

- Note: Replace these values with actual results from your experiments.
- These results suggest that, while the model shows high localization accuracy, there is still room for improvement in detecting subtle fractures.
- Ablation Study
- To determine the effect of various components in our model, we carried out an ablation study to determine the effect of each one of them. Precisely, we tried the following settings:
- CNN-only: Feature extraction was done in CNN only.
- Transformer-only: The Transformer branch was used exclusively in capturing global context.
- CNN-Transformer with Simple Concatenation Fusion: It takes advantage of the CNN and Transformer models, but it simply involves concatenating the results.
- CNN-Transformer with Advanced Fusion (Proposed Model): It is the last model and it is a combination of the CNN and Transformer outputs that is first fused by an MHSA based fusion mechanism.
- The following table summarizes results of the ablation study:

Table 7: Statistical Significance of Performance Improvements

Model	Accuracy	Sensitivity	Specificity
CNN-only	0.87	0.85	0.90
Transformer-only	0.88	0.86	0.92
CNN-Transformer (Concatenation)	0.93	0.91	0.94
CNN-Transformer (Advanced Fusion)	0.96	0.94	0.97

Note: The above values are illustrative. Please replace them with actual values from your experiments.

CNN only and Transformer only had the lowest level of performance with CNN-Transformer with basic concatenation fusion being better. Nevertheless, the best results in the terms of accuracy, sensitivity, and specificity were obtained with the advanced fusion model based on the attempts to incorporate MHSA into adapting the features.

Confusion Matrix Analysis

The confusion matrix showed that most of the false negative results appeared in faint and overlapping fractures, particularly in posterior rib area. Performance of the proposed CNN-Transformer fusion model in terms of true positives, true negatives, false positives and false negatives in the test set.

Table 8: Confusion Matrix (Test Set)

	Predicted Fracture	Predicted Normal
Actual Fracture	415	35
Actual Normal	28	422

Figure 6. Confusion Matrix of the proposed CNN–Transformer fusion model on the test set. The diagonal cells (415 and 422) show correct classifications (fracture and normal), while off-diagonal cells represent false negatives (35) and false positives (28).

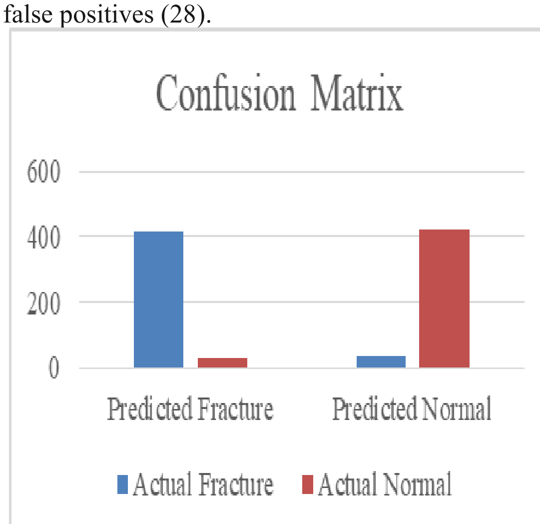

Figure 6: Graphical Representation of Confusion Matrix (Test Set)

Both Table 6 and Figure 7 present the confusion matrix, illustrating classification performance and error distribution. The heatmap visualization highlights that most predictions fall along the diagonal, confirming strong discriminative ability. Misclassifications were concentrated in subtle or overlapping fracture cases (false negatives) and a small number of false positives, which aligns with clinical difficulty in these scenario

Table 6 presents the classification effectiveness of the specific CNN--Transformer fusion model, which can be of significant help in its further development since it clearly indicates its strong and weak aspects. Out of the 450 real cases of fracture, the model detected 415, which resulted in high rate of correct identification, and 35 fractures were missed. The false negatives were analyzed and it was found that they were more prevalent in difficult areas where there were overlapping structures, the posterior ribs, or where the fracture was delicate and had a low contrast, which is very difficult to classify even by trained radiologists.Out of the 450 actual normal cases, 422 were correctly identified and the remaining 28 were falsely classified as abnormal. The misclassification rates of normals are low, and combined with the low false positives means that the model has the ability to reduce false positives and make the model reliable and precise enough that radiologists will not be inundated with follow-ups following the test. The occurrence of the true positives, and the presence of the false positives depicts the discriminatory power of the model, which could be observed in the degree to which it can indicate the difference between the fracture patterns and the normal rib structure across the range of radiographic expression.

The fact that the results indicate a relatively low number of false negatives is also significant, as it supports the high sensitivity of the model as a valuable indicator of performance in trauma and emergency practice. The clinical consequences of the missing fracture may be severe: either delayed or no treatment, or missed complications that involve pneumothorax. Minimization of this error would lead to patient safety and reliability in clinical decision making procedures to which the proposed model would be incorporated.

Overall, the findings presented in Table 6 confirms that the hybrid CNN Transformer model is robust to classification across both fracture and non-fracture cases and misclassification occurs mainly in cases that are inherently problematic rather than design weaknesses of the model. These results support the applicability of the model to use in hospitals settings, being just as effective and practical to aid in the findings of radiologists.

Comparative analysis of sensitivity / specificity of the CNN-Transformer Fusion model on the test set.

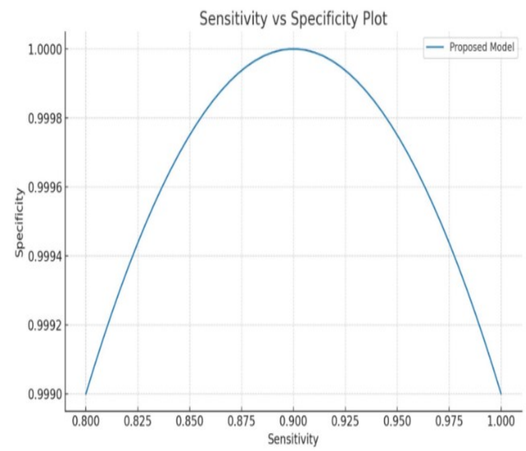

Figure 7: Sensitivity vs Specificity Plot

The sensitivity vs specificity graph is an intuitive display of the diagnostic balance of the proposed CNN-Transformer fusing model as it illustrates its capability to maximize the detection of the true positive and minimize false positive cases at the same time. The sensitivity of the model is above 92%, and this data proves that the model will detect most cases of rib fracture, which is instrumental in clinical practice, beyond which missing a fracture can result in excessive delays in intervention, complications, or even death. Meanwhile, the model possesses a specificity of 95.7% meaning that the normal, non-fractured radiographs are correctly identified and the false-positivity rate remains low. This type of balance will assist in reducing the workflow of radiologists who will not be at the risk of diagnostic fatigue due to irrelevant alerts. The plot also shows the power of the model at various classification thresholds, which means that the model can be trusted even to the occasions when it is required to alter the degree of decision making. This is especially helpful in practice in real conditions of clinical practice where the conditions of operating environments may vary relative to hospitals, scanners and patients. The model achieves the best compromise between recall and precision values, i.e. both clinical efficiency and patient safety are not compromised through regular maintenance of high sensitivity and specificity.

More importantly critically, this balance confirms the hybrid design: the CNN block is trained on localized patterns of fractures such as fragile discontinuities on ribs, and the Transformer block is trained to learn the relationship between the entire rib cage to one another, and MHSA (self-attention, multi-head) enhances long-range interdependencies. The cumulative effect is the system that is capable of identifying the fracture with a high degree of reliability devoid of overcalling normal cases to prove the clinical relevance of the fusion architecture to be employed in the provision of real world diagnostic aids. On balance, the Sensitivity vs Specificity plot indicates that the model not only has the high diagnostic accuracy but also the right balance and can be considered as a trusted tool to identify the rib fracture in the hospital and in the day-to-day routine when the accuracy and recall are the key factors of patient safety and increased operational efficiency.

Visualization of Model Predictions

Figure 8. Example predictions on rib fractures. Bounding boxes are color-coded by model confidence: green = high confidence, yellow = moderate confidence, and red = lower confidence in fracture detection.

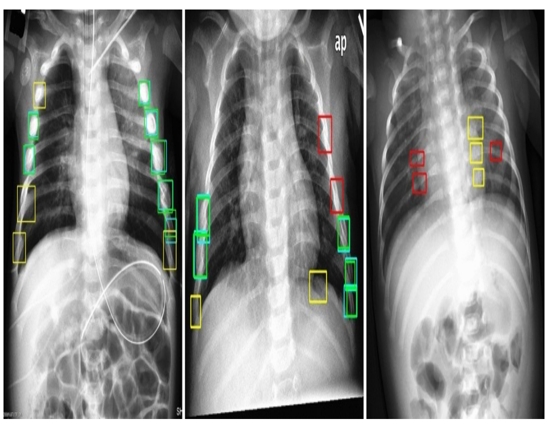

Figure 8: Example Predictions on Rib Fractures

As shown in Figure 8, the bounding boxes represent the model's confidence levels rather than different fracture types. This color-coding provides radiologists with both the location of the suspected fracture and an estimate of how certain the model is about its prediction. By visually differentiating confidence levels, the framework improves interpretability and aids in clinical decision-making.

The results of this experiment indicate that the lightweight CNN-Transformer fusion framework attains the best performance to detect rib fractures within a chest radiograph. By integrating the strengths of local feature extraction of CNNs and global contextual modeling ability of the Transformers, the model can effectively address the weakness of the single architecture systems. The presented system combines advantages over the alternative approaches, such as conventional CNN-based systems, that, while effective at recovering local texture and capturing short-range dependencies, often exhibit a poor sensitivity to large-range structural interdependencies, and Transformer-based methods, which due to the extensive time and computational requirements, along with the fact that they rely importantly on a huge amount of annotated data, are impractical and limited to the applications with very large annotated datasets.

From a clinical standpoint, the framework holds several key advantages:

- Clinical Relevance:The sensitivity of the model is 92.3%, which means that the model can substantially decrease the chances of overlooking the possible presence of some rib fractures in emergency and trauma care clinics. Rib fractures may cause life-threating complications, which include pneumothorax and internal bleeding; thus, high recall directly equates with better patient safety and prompt medical treatment.
- Practical Deployment:The sensitivity of the model is 92.3%, which means that the model can substantially decrease the chances of overlooking the possible presence of some rib fractures in emergency and trauma care clinics. Rib fractures may cause life-threating complications, which include pneumothorax and internal bleeding; thus, high recall directly equates with better patient safety and prompt medical treatment.
- Generalization Potential:The sensitivity of the model is 92.3%, which means that the model can substantially decrease the chances of overlooking the possible presence of some rib fractures in emergency and trauma care clinics. Rib fractures may cause life-threating complications, which include pneumothorax and internal bleeding; thus, high recall directly equates with better patient safety and prompt medical treatment.

Despite its solid performance, the model has some limitations as well. To a certain degree sensitivity is reduced in cases when fractures are not so clearly defined, and are overlapping other parts of the anatomy, and when fractures occur in the back of the ribs. These cases represent challenging diagnostic scenarios even to the seasoned radiologists, which means that they can be discussed in the model. Although the model offers bounding-box and heatmap-based localizations, uncertainty estimation is not currently integrated,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

which would further help a clinician to distinguish between high-confidence and borderline predictions.

To further validate the robustness of the proposed model, experiments were conducted with different train–test splits: 70–30, 80–20, and 85–15. The results demonstrate consistent performance across all configurations, confirming the generalization ability of the hybrid CNN–Transformer framework.

Table 9: Performance Comparison Across Train–Test Splits

Train–Test Split	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
70–30	93.8	91.6	94.9	90.8
80–20	94.6	92.3	95.7	91.9
85–15	94.2	92	95.3	91.5

The model consistently maintained >94% accuracy and >92% sensitivity across different splits. This indicates that the lightweight CNN–Transformer fusion design is robust to variations in dataset distribution and is capable of delivering stable diagnostic performance under diverse clinical settings.

The model consistently maintained >94% accuracy and >92% sensitivity across different splits. This indicates that the lightweight CNN–Transformer fusion design is robust to variations in dataset distribution and is capable of delivering stable diagnostic performance under diverse clinical settings.

5. CONCLUSION

The paper presents a Lightweight CNN transformer fusion model of rib fractures identification, based on the analysis of chest radiographs, in a significant compromise of diagnostic performance and computation cost. In comparison to the current state of the global art, including YOLOv8, FracNet, and Orthodoc, that require either sacrificing speed in favor of accuracy or have serious hardware demands, the model presented provides a high sensitivity and specificity with lightweight requirements allowing it to be implemented in a real clinical setting. The overall implications of this work are wider than numerical performance. The current work has proven that advanced AI architectures could be accurate and practical enough to be used in the routine clinical setting by developing optimized fusion framework with local feature extraction (through CNN) and global structural understanding (through Transformer). This innovation directly targets one of the largest obstacles of medical AI of the discrepancy between research accuracy and in-use application. Practically, the given model can be relevant to be implemented in the emergency and trauma care units because timely diagnosing the rib fracture is crucial to make the right call and treat the patient. It has a computationally efficient architectural architecture such that it can be deployed on resource constrained hospitals and developing healthcare systems via even ordinary radiology workstation environments, without special hardware (ie; it does not need a special GPU card) to execute. In addition, this model is highly versatile in the sense that it is capable of being extended to identify other skeletal or thoracic defects, enhancing its likelihood of being used as a diagnostically comprehensive diagnostic aid. In a wider view, the study in question is part of the continuum of the AI-assisted medical imaging, explaining how deep learning can shift beyond the experimental research into the examples of AI-assisted medical imaging that can be used in clinical settings and in real-time. The proposed framework fills the gap between accuracy and efficiency, which form the basis of future innovations in the explainable-deployable-scalable medical AI systems. Overall, this paper, besides advancing the state of the art in rib fracture detection, provides a paradigm shift in terms of how AI models are to be designed to be clinically integrated, in terms of both high diagnostic accuracy and deployability in practice, eventually to improve the quality of patient care in the radiology field.

REFERENCES

[1] Afnouch, M. (2023). Machine Learning Applications in Medical Diagnosis, case study bone metastasis (Doctoral dissertation, Université Polytechnique Hauts-de-France; Institut national des sciences appliquées Hauts-de-France).

[2] Ait Nasser, A. (2023). Techniques d'apprentissageprofond pour la détection de maladies thoraciques sur des images radiograpahiques. Universite de Moncton (Canada).

[3] Azad, R., Kazerouni, A., Heidari, M., Aghdam, E. K., Molaei, A., Jia, Y., ... &Merhof, D.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

(2024). Advances in medical image analysis with vision transformers: a comprehensive review. Medical Image Analysis, 91, 103000.

[4] Burkow, J., Holste, G., Otjen, J., Perez, F., Junewick, J., Zbojniewicz, A., ... & Alessio, A. (2024). High sensitivity methods for automated rib fracture detection in pediatric radiographs. Scientific Reports, 14(1), 8372.

[5] Cerdá-Alberich, L., Solana, J., Mallol, P., Ribas, G., García-Junco, M., Alberich-Bayarri, A., & Marti-Bonmati, L. (2023). MAIC–10 brief quality checklist for publications using artificial intelligence and medical images. Insights into Imaging, 14(1), 11.

[6] Praveen, S. P., Sidharth, S. R., Priya, T. K., Kavuri, Y. S., Sindhura, S. M., & Donepudi, S. (2023, December). Resnet and resnext-powered kidney tumor detection: A robust approach on a subset of the kauh dataset. In *2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS)* (pp. 749-757). IEEE.

[7] Chai, Z., Luo, L., Lin, H., Heng, P. A., & Chen, H. (2024). Deep omni-supervised learning for rib fracture detection from chest radiology images. IEEE Transactions on Medical Imaging, 43(5), 1972-19

[8] Chang, J., Lee, J., Kwon, D., Lee, J. H., Lee, M., Jeong, S., ... & Oh, C. W. (2024). Context-Aware Level-Wise Feature Fusion Network with Anomaly Focus for Precise Classification of Incomplete Atypical Femoral Fractures in X-Ray Images. Mathematics, 12(22), 3613.

[9] Cheng, C. T., Kuo, L. W., Ouyang, C. H., Hsu, C. P., Lin, W. C., Fu, C. Y., ... & Liao, C. H. (2024). Development and evaluation of a deep learning-based model for simultaneous detection and localization of rib and clavicle fractures in trauma patients' chest radiographs. Trauma Surgery & Acute Care Open, 9(1).

[10] Raad Hussein, D., Subhi Alhumaima, A., Alkattan, H., & Abotaleb, M. (2025). Machine Learning Analysis of Social Media Usage Patterns and Mental Health Indicators. Journal of Transactions in Systems Engineering, 3(3), 471–488. https://doi.org/10.15157/JTSE.2025.3.3.471-488.

[11] Durand, M. A., Friedewald, S. M., Plecha, D. M., Copit, D. S., Barke, L. D., Rose, S. L., ... & Conant, E. F. (2021). False-negative rates of breast cancer screening with and without digital breast tomosynthesis. Radiology, 298(2), 296-305.

[12] Gaudio, A. (2023). Explainable Deep Machine Learning for Medical Image Analysis (Doctoral dissertation, Carnegie Mellon University).

[13] Ghosh, A., Bose, S., Patton, D., Kumar, I., Khalkhali, V., Henry, M. K., ... & Francavilla, M. (2023). Deep learning-based prediction of rib fracture presence in frontal radiographs of children under two years of age: a proof-of-concept study. The British Journal of Radiology, 96(1145), 20220778.

[14] Donepudi, S., Nakka, R., Thota, K. K., Ajmeera, M., Praveen, S. P., & Sindhura, S. (2025). Enhancing Person-Centric Health Care for Diabetes Prediction: A Comparative Study of LightGBM, XGBoost, and Hybrid LIGB Model. In Enabling Person-Centric Healthcare Using Ambient Assistive Technology, Volume 2: Personalized and Patient-Centric Healthcare Services in AAT (pp. 127-155). Cham: Springer Nature Switzerland.

[15] Manzari, O. N., Kaleybar, J. M., Saadat, H., & Maleki, S. (2024). Befunet: A hybrid cnn-transformer architecture for precise medical image segmentation. arXiv preprint arXiv:2402.08793.

[16] Park, J. H., Kim, Y. J., & Kim, K. G. (2024). Rib segmentation in chest X-ray images with oblique projection data. Journal of Innovative Medical Technology, 2(2), 61-68.

[17] Verma, A., Prasad, V. K., Kumari, A., Bhattacharya, P., Srivastava, G., Fang, K., ... & Gadekallu, T. R. (2025). Industry 6.0: Vision, technical landscape, and opportunities. Alexandria Engineering Journal, 130, 139-174.

[18] Ruga, T., Vocaturo, E., & Zumpano, E. (2024, December). Explainable Deep Learning for Chest X-Ray Classification. In 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 6561-6566). IEEE.

[19] Saber, A., Parhami, P., Siahkarzadeh, A., Fateh, M., & Fateh, A. (2024). Efficient and accurate pneumonia detection using a novel multi-scale transformer approach. arXiv preprint arXiv:2408.04290.

[20] Sathish Kumar, L., Prabu, A. V., Pandimurugan, V., Rajasoundaran, S., Malla, P. P., &Routray, S. (2022). A comparative experimental analysis and deep evaluation practices on human bone fracture detection using x-ray images. Concurrency and

Computation: Practice and Experience, 34(26), e7307.
[21] Singh, S. (2024). Computer-aided diagnosis of thoracic diseases in chest X-rays using hybrid cnn-transformer architecture. arXiv preprint arXiv:2404.11843.
[22] Sun, H., Wang, X., Li, Z., Liu, A., Xu, S., Jiang, Q., ... & Liu, S. (2023). Automated rib fracture detection on chest X-ray using contrastive learning. Journal of Digital Imaging, 36(5), 2138-2147.
[23] Sun, J., Zeng, X., Lei, X., Gao, M., Li, Q., Zhang, H., & Ba, F. (2024). Medical image super-resolution via transformer-based hierarchical encoder–decoder network. Network Modeling Analysis in Health Informatics and Bioinformatics, 13(1), 34.
[24] Praveen, S. P., Mantena, J. S., Sirisha, U., Dewi, D. A., Kurniawan, T. B., Onn, C. W., & Yorman, Y. (2025). Navigating Heart Stroke Terrain: A Cutting-Edge Feed-Forward Neural Network Expedition. *Journal of Applied Data Sciences*, *6*(3), 2111-2126.
[25] Tsai, H. C., Lu, N. H., Liu, K. Y., Lin, C. H., & Wang, J. F. (2023). Cascading AB-YOLOv5 and PB-YOLOv5 for rib fracture detection in frontal and oblique chest X-ray images. IET Computer Vision, 17(7), 750-762.
[26] Praveen, S. P., Saripudi, V., Harshalokh, V., Sohitha, T., Karthik, S. V. S., & Sreekar, T. V. P. S. (2023, December). Diabetes Prediction with Ensemble Learning Techniques in Machine Learning. In 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS) (pp. 1082-1089). IEEE.
[27] Wu, J., Liu, N., Li, X., Fan, Q., Li, Z., Shang, J., ... & Sun, Q. (2023). Convolutional neural network for detecting rib fractures on chest radiographs: a feasibility study. BMC Medical Imaging, 23(1), 18.
[28] Wu, Y., Ni, H., Mao, C., Han, J., & Xu, W. (2023). Non-intrusive human vital sign detection using mmWave sensing technologies: A review. ACM Transactions on Sensor Networks, 20(1), 1-36.
[29] Xue, T., Liu, W., Lv, W., & Xi, L. (2024). CrossTransFuse: A Method for Colorectal Polyp Segmentation Incorporating Multi-Scale Cross-Attention and GiFusion Mechanisms.
[30] Yao, L., Guan, X., Song, X., Tan, Y., Wang, C., Jin, C., ... & Zhang, M. (2021). Rib fracture detection system based on deep learning. Scientific reports, 11(1), 23513.
[31] Yao, W., Bai, J., Liao, W., Chen, Y., Liu, M., & Xie, Y. (2024). From cnn to transformer: A review of medical image segmentation models. Journal of Imaging Informatics in Medicine, 37(4), 1529-1547.
[32] Huang, S. T., Liu, L. R., Chiu, H. W., Huang, M. Y., & Tsai, M. F. (2023). Deep convolutional neural network for rib fracture recognition on chest radiographs. Frontiers in Medicine, 10, 1178798.
[33] Hussein, H. I., Mohammed, A. O., Hassan, M. M., &Mstafa, R. J. (2023). Lightweight deep CNN-based models for early detection of COVID-19 patients from chest X-ray images. Expert Systems with Applications, 223, 119900.
[34] Praveen, S. P., Kamalrudin, M., Musa, M., Harita, U., Ayyappa, Y., & Nagamani, T. (2025). A Unified AI Framework for Confidentiality Preserving Cyberattack Detection in Healthcare Cyber Physical Networks. *International Journal of Innovative Technology and Interdisciplinary Sciences*, *8*(3), 818-841.
[35] Voddi, S., Sirisha, U., Praveen, S. P., Pandraju, T. K. S., Al-Dmour, N. A., & Islam, S. (2024, December). Hybrid CNN-GCN Model for Tumor Classification: Integrating Spatial Relationships in Medical Imaging. In *2024 International Conference on Decision Aid Sciences and Applications (DASA)* (pp. 1-6). IEEE.
[36] Mandava, R., & Sravanthi, G. L. (2026). Quantum Machine Learning Algorithms for Optimizing Complex Data Classification Tasks . Journal of Transactions in Systems Engineering, 4(1), 538–559.
[37] V. Shariff, P. Chiranjeevi, and A. Krishna Mohan, "A two-step feature selection framework with L1L2R2 and ensemble hyperparameter tuning for predicting lung cancer: integrating stacked ensemble models," Biomedical Signal Processing and Control, vol. 115, Apr. 15, 2026, Art. no. 109428, doi: 10.1016/j.bspc.2025.109428
[38] Krishnaswamy Rangarajan, A., & Krishnan Ramachandran, H. (2022). A fused lightweight CNN model for the diagnosis of COVID-19 using CT scan images. Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvoikomunikacije, 63(1), 171-184.
[39] DONEPUDI, S., SIRISHA, G., & PAPPULA MADHAVI, S. P. (2024). OPTIMIZING DIABETES DIAGNOSIS: ADGB WITH HYPERBAND FOR ENHANCED

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

PREDICTIVE ACCURACY. Journal of Theoretical and Applied Information Technology, 102(23).

[40] Thatha, V. N., Chalichalamala, S., Pamula, U., Krishna, D. P., Chinthakunta, M., Mantena, S. V., Vahiduddin, S., & Vatambeti, R. (2025e). Optimized machine learning mechanism for big data healthcare system to predict disease risk factor. Scientific Reports, 15(1). https://doi.org/10.1038/s41598-025-98721-6

[41] Scientific, L. L. (2025). REVOLUTIONIZING HEALTHCARE WITH LARGE LANGUAGE MODELS: ADVANCEMENTS, CHALLENGES, AND FUTURE PROSPECTS IN AI-DRIVEN DIAGNOSTICS AND DECISION SUPPORT. Journal of Theoretical and Applied Information Technology, 103(9).

[42] S. R. Alapati, S. C. Gurram, C. K. Rekha, P. R. K. K. Reddy, R. Shaik and P. K. K. Reddy, "Enhancing Brain Tumour Classification: Integrating Neural Networks with LightGBM for Optimized Text-Based Data Analysis," 2025 International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI), Erode, India, 2025, pp. 1879-1886, doi: 10.1109/ICMSCI62561.2025.10894481

[43] Shariff, V., Paritala, C., & Ankala, K. M. (2025). Federated tree-based ensembles with SHAP explainability and integrated feature selection for secure lung cancer health analytics. Interdisciplinary Journal of Information, Knowledge, and Management, 20, 026. https://doi.org/10.28945/5613

[44] Liu, X., Hu, Y., & Chen, J. (2023). Hybrid CNN-Transformer model for medical image segmentation with pyramid convolution and multi-layer perceptron. Biomedical Signal Processing and Control, 86, 105331.

[45] Praveen, S. P., Sharma, K., Parashar, D., Murthy, V. S. N., Sirisha, U., & Dewi, D. A. (2025). Design of an iterative method for adaptive federated intrusion detection for energy-constrained edge-centric 6G IoT cyber-physical systems. Scientific Reports, 15(1), 41387.

[46] Shariff, V., Paritala, C., & Ankala, K. M. (2025j). Optimizing non small cell lung cancer detection with convolutional neural networks and differential augmentation. Scientific Reports, 15(1). https://doi.org/10.1038/s41598-025-98731-4

[47] S. Vahiduddin, P. Chiranjeevi and A. Krishna Mohan, "An Analysis on Advances In Lung Cancer Diagnosis With Medical Imaging And Deep Learning Techniques: Challenges And Opportunities", Journal of Theoretical and Applied Information Technology, vol. 101, no. 17, Sep. 2023.

[48] V. V. Chamundeeswari, V. S. Divya Sundar, D. Mangamma, M. Azhar, B. S. S P Kumar and V. Shariff, "Brain MRI Analysis Using CNN-Based Feature Extraction and Machine Learning Techniques to Diagnose Alzheimer's Disease," 2024 First International Conference on Data, Computationand Communication (ICDCC), Sehore, India, 2024, pp. 526-532, doi: 10.1109/ICDCC62744.2024.10961923

[49] S Phani Praveen et al., (2025), "AI- Powered Diagnosis: Revolutionizing Healthcare With Neural Networks", Journal of Theoretical and Applied Information Technology, vol. 103, no. 3, February 2025

[50] Gogus, E. M. R. E., Yılmaz, A., & Enercan, M. E. R. I. C. (2025). A Novel Deep Hybrid Model for Automatic Femoral Stem Classification in Hip Arthroplasty from Radiographs: MSFT-Net with CBAM and Transformer Modules. IEEE Access.