Journal of Chemical Health Risks

www.jchr.org

JCHR (2025) 15(6), 1640-1656 | ISSN:2251-6727

Advanced Fracture Diagnosis and Classification Using Hierarchical Vision Transformers and Cross-Model Attention Mechanisms

Sri Sai Vignesh Kadiyala [1], Chunduri Kiran Kumar [2], Chanumolu Kiran Kumar [3]

[1] Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, India

[2] Department of Master of Computer Applications, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, India

[3] Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, India

(Received: 07 October 2025 Revised: 15 November 2025 Accepted: 02 December 2025)

KEYWORDS:
Medical imaging, fracture diagnosis, vision transformers, cross-modal attention, explainable AI, deep learning, computer-aided diagnosis, orthopedics, radiology, hierarchical attention.

ABSTRACT:
This research describes state-of-the-art deep learning methods for the automated diagnosis and classification of fractures in X-ray, CT, and MRI images. Furthermore, we developed a novel hierarchical Vision Transformer (ViT) architecture with cross-modal attention that encodes X-ray, CT, and MRI data simultaneously to improve accuracy. This work adds uncertainty quantification and explain-ability tailored to preserve performance across varying demographic and anatomical models to address the fundamental gaps of existing approaches. Multi center data set validation showed 96.1% accuracy in fracture detection and classification with the fully multi modal system and 92.3% accuracy with single-modality inputs. With the proposed system integrating cloud infrastructure employing edge computing, diagnostic time is reduced by 93%, with 93% of assessments marked as clinically relevant by 92.3% accuracy with single-modality inputs. With the proposed system integrating cloud infrastructure employing edge computing, diagnostic time is reduced by 93%, with 93% of assessments marked as clinically relevant by specialists, enabling enhanced patient management across disparate healthcare systems.In addition, the hierarchical architecture demonstrated strong robustness against noise, variable imaging resolutions, and incomplete modality availability—conditions that commonly occur in real-world clinical workflows. The cross-modal attention mechanism effectively aligned anatomical cues across X-ray, CT, and MRI domains, significantly improving the system's sensitivity toward subtle and complex fracture types. Our uncertainty quantification module enabled calibrated confidence scores, allowing clinicians to better interpret borderline decisions and reduce false positives. Explain-ability was achieved through the integration of attention heat-maps and gradient-based attribution, providing transparent visual insights into the model's reasoning. Evaluation using SHAP and Layer-CAM techniques further validated the clinical interpret-ability of model outputs.

1. Introduction

One of the most significant clinical problems in emergency medicine and orthopaedics is the accurate and timely identification of bones that have fractures, or broken bones. This remains a major problem. From the clinical perspective, this is among the foremost concerns. With that in mind, this is one of the most critical challenges that may be encountered during the process of treatment. Accomplishing this exercise is important because it allows one to explain matters in simpler terms. With regard to conventional techniques for diagnosis, there is a danger of elements of subjectivity, workflow inefficiencies, and major lags in diagnostic processes. All of these are things that are possibly construed as such.

Journal of Chemical Health Risks
www.jchr.org
JCHR (2025) 15(6), 1640-1656 | ISSN:2251-6727

There is a likelihood of the occurrence of all of these possibilities. When all has been said and done, there is a chance that this might happen. The situation that is most likely existing is such because the amount of information possessed by the radiologists is " the weakest link" in accuracy "pouched" by conventional diagnostic techniques. Under different circumstances, the problems we are grappling with would be manageable, but not now. Rather, the meticulous interpretation of complex fracture patterns using multiple imaging modalities such as movie X-ray, CT, and MRI makes matters worst. From this perspective, this is something that needs particular attention in clinical environments that are resource constrained and understaffed as well as in most settings where access to professionals is minimal. More so, this is something that needs auxiliary in regard to the situation.

It is something that need special attention again. These conditions suggest the last comment was well within reason. With the rapid increase in artificial intelligence technologies, especially deep learning, there have been some positive developments in the medical imaging field. The progress achieved in this field has resulted from the application of modern technology and these developments have, in turn, produced new technological innovations. The application of modern technology has further enhanced these results becoming more and more favorable over time. With regard to the outcomes of the work that was done, these outcomes have been enhanced extensively. Regardless, the procedures currently available for the diagnosis and classification of fractures contain several gaps that require attention. Nonetheless, these methods are available for use. These gaps are something that need to be considered. Addressing these gaps is essential and retrieving positive solutions becomes a higher priority than anything else. These gaps are crucial to focus on and require immediate strategy implementation to address. Keeping these potentially dangerous problems at the forefront instantly becomes the highest priority.

Factors such as focusing on a limited imaging modality, struggling with unusual fracture patterns, inconsistent performance across different anatomical areas, and lack of clinically useful explain-ability features which describe meaningful clinical ground reasoning, which are vital for clinical use. Most, if not all, of the systems which are currently operational are what is referred to as 'black boxes.' This means these systems make claims without providing uncertainty estimation or explanatory evidence. This is the situation that has recently arisen as a result of a the great majority of these systems being operational. A significant challenge, as a result of incorporating these technologies into clinical workflow, is made more difficult due to the lack of reliable supporting information. With the information given, is the reasoning for attempting to advance automation and diagnostic processes for more precise clinical reasoning. This is the aim of the present study, which is to address the gap left by previous research on creating a structured framework for automated fracture diagnosis and classification. Having a thorough understanding will aid in accomplishing this. The construction of this framework will be aided by the development of a complete framework. It is the development of this framework for which the construction is finally going to be something-that comes about. There will be an implementation of this framework. Because of this, we are in a position to achieving the most advantageous outcomes that are even remotely conceivable.

Additional Perspective On Hierarchy Within An Organization Furthermore, at this present moment in time, transformers and cross-modal attention mechanisms are being very actively used in relation to this framework within this organizational structure. Our system can integrate data from different imaging modalities without compromising performance, particularly enabling optimal performance when only one imaging modality is available. These positive outcomes are made possible for us. These achievable results enable us to optimize the aforementioned scenarios effectively. Because of this, carrying out imaging procedures with precision and speed becomes possible. This increased efficiency results from the advantages obtained from this. Thus, the operational imaging procedures are performed with enhanced efficiency compared to any other scenario. This effort aims specifically to provide professionals in the medical field with visuals to assist in understanding the algorithm-based reasoning decisions made by the algorithm. An explain-ability system is defined as a system that aims to integrate uniquely defined characteristics of explain-ability within itself. This step is taken with the expectation that it provides them the capability to deeper comprehend the algorithm itself. The

Journal of Chemical Health Risks
www.jchr.org
JCHR (2025) 15(6), 1640-1656 | ISSN:2251-6727

reasoning was consistent with the approach taken in the development of the system in question. These strategies in question incorporate both appraisal of uncertainty and attention map displays into their implementation processes. There are other elements, one of which is estimating uncertainty.

In order to trust our technique, we perform strenuous experiments on numerous datasets from multiple medical organizations. As stated, trust and dependability is what we are trying to determine. The conduct of the project is undertaken in order to ensure accuracy of our strategy. The data published in this research proves our approach is accurate along with the fact that our technique is proved to be reliable across numerous patient demographic variables, imaging types, and fracture patterns. It is important to note that this approach is bolstering augmenting patient outcomes. It should be noted that this serves as a secondary topic of interest. The approach developed improves timeliness of diagnoses while maintaining high accuracy thresholds. An earlier prognosis will result in a more straightforward course of therapeutic action. This is the reason why basing decisions on earlier information aids in avoiding complications that may arise later. Thus, it may becomes possible to set goals more easily than was previously possible, which is why this is important. Furthermore, therapeutic objectives which are more immediately achievable, like those set in this model, create an easier likelihood of achieving this goal. This outline aids in forming achievable objectives for the treatment period.

2. Literature Survey

These past ten years have seen an increase in the progress made in diagnosing fractures owing to the application of deep learning. This advancement was made possible through the adoption of deep learning. Early on, the principal method applied was the use of convolutional neural networks (CNNs). This was the primary strategy that was utilized. These networks were an attempt to provide solutions to the problems of computer vision. Lindsey and her coworkers [1] designed a fracture detection system based on CNNs. The algorithm attained an 83% sensitivity in the detection of wrist fractures on ordinary radiographs. This was the first system of its kind that could be used in a clinical setting. It is important to mention that the fracture detection system enabled the completion of this task. His system was constructed using a modified version of AlexNet' s computer architecture. Though this research did not provide detailed fracture characterization and focused on binary classification, it was shown that deep learning can recognize fractures. This was true in spite of the fact that the study was done. Regardless of this, it is demonstrates that deep learning is capable of recognizing fractures. hung et al. [2] did further research that enabled them to improve upon this method through transfer learning using ImageNet pre-trained ResNet-50 architectures.

This was made possible for them through this study. As a result of the use of this method, the effectiveness of this approach was increased to 87 percent for accuracy regarding long bone fractures.Their unprecedented region-specific analysis pioneering the integration of anatomical context did not help them cope with complex and subtle fracture patterns too well. Even with all the attention drawn to the issue of providing anatomical context, this was the outcome. Jones et al. [3] advanced this approach even further with ensemble methods employing a wide range of CNN designs. This is what permitted them to obtain that enhancement. Consequently, they achieved an 89% accuracy rate concerning the myriad of fracture patterns within the spectrum of fractures. On the contrary, the system experienced a significant drop in performance for younger patients and those with atypical fracture patterns. That was the case for that system. Unlike slice-based analysis, Chen and colleagues [4] pioneered the development of fully three-dimensional convolutional neural networks, which are capable of processing entire image volumes. This was done for the purpose of assessing volumetric CT data, which is why this was done. Even though their approach for vertebral fractures achieved an accuracy of 91\%—a remarkable figure—it was very resource-intensive, greatly limiting its viability for clinical investigations. Nonetheless, achieving such a level of precision was quite impressive. The issue was solved, and their goals were met by Zhang et al. [5] with the use of attention-based three-dimensional convolutional neural networks (CNNs) that focused on specific areas in a preferred fashion. As a result, both the effectiveness of the computing processes have improved, while the diagnostic performance had remained unchanged.

Journal of Chemical Health Risks
www.jchr.org
JCHR (2025) 15(6), 1640-1656 | ISSN:2251-6727

On the other hand, the strategy that was implemented still could not successfully integrate with clinical workflow tools well enough to be regarded as sufficiently successful. The area of medical imaging analysis has undergone a paradigm shift as a result of the emergence of transformer designs, which is how this change came about. Due to the emergence of the transformer designs, this occurred. One group of researchers, Dosovitskiy et al. [6], showed that Vision Transformers could be employed for more general image classification tasks. In contrast, medical imaging applications were the focus of the designs that Hatamizadeh et al. [7] created. These two investigations were published together in Scientific Reports after completing the reported work. This is the publication which contained them. It was the design of UNETR transformers, which was used for segmentation tasks, that revealed the ability of transformers to capture longrange dependencies in medical images.

The investigation was aimed at studying the complicated fracture patterns and their intricate spatial relationships. This work constitutes a rather large share of the ongoing analytical work. Multi-modal approaches in particular have been gaining traction over the past few years as the focus of research efforts. It is a gradual process. In their work, Kumar et al. [8] proposed a dual pathway model integrating X-ray and CT data. This model was developed as part of their work. It is tailored toward the execution of an extensive fracture analysis. The design achieved a remarkable 92% accuracy, yet it was essential to have both modalities present during participant imaging sessions. To add MRI data, Wang et al. [9] developed this strategy further by using late fusion techniques, which enabled them to integrate the data. As discussed in the review, some changes were made throughout the diagnostic process which sought to improve sensitivity for occult fractures. On the contrary, this system had some form of interaction within different modes that was very limited in scope. Rodriguez et al. [10] accounted for some explainability in medical AI systems due to the use of attribution methods applicable to medical imaging.

This is on account of the contribution of such approaches. Particularly, the focus has been on medical AI systems explainability. Although his modifications of Grad-CAM enhanced transparency for fracture localization, they still communicated very little regarding the features critical to the fracture diagnosis. This was under the claim of enhancing transparency for fracture localization. Moreover, Bharatanatyam and other researchers advocated the use of Bayesian \ conventional neural networks (CNNs) concerning uncertainty quantification in reference [11]. It is unfortunate that the employment of these CNNs leads to the production of theoretically well-calibrated confidence measures; however, this results in greater computational requirements of the everincreasing complexity. Even with the advances that have been made, the creation of systems that can accurately integrate multiple imaging modalities while sustaining high performance with limited data, generalizing across diverse patient demographics, and offering clinically insightful explanations remains an enduring challenge. During the course of our inquiry, we built upon these very limitations while addressing the challenges that are intrinsically associated with them. Unique architectural designs and integration strategies tailored exclusively for the building are what accomplish this goal.

3. Methodology

3.1 Dataset Acquisition and Pre-processing

Our research utilized a comprehensive dataset comprising 54,900 medical images across three modalities: X-ray (50,000 general radiographs plus 3,500 fracture-specific images), CT (15,000 volumetric scans), and MRI (12,000 multi-sequence studies). Images were sourced from nine medical centers across four countries, ensuring diverse representation of patient demographics, imaging protocols, and equipment manufacturers. The dataset encompasses fractures from all major anatomical regions, with particular emphasis on commonly affected areas including extremities, spine, pelvis, and ribs. Data preprocessing followed a standardized pipeline tailored to each imaging modality. For radiographs, we applied adaptive histogram equalization, noise reduction through guided filtering, and standardization to 512×512 resolution while preserving aspect ratios through zero-padding. CT volumes underwent resampling to isotropic 1mm³ voxels, intensity normalization using Hounsfield unit windowing optimized for bone visualization, and standardization to 256×256×128 dimensions. MRI preprocessing incorporated bias field correction using N4ITK, intensity normalization through z-score standardization independently applied to each sequence

Journal of Chemical Health Risks
www.jchr.org
JCHR (2025) 15(6), 1640-1656 | ISSN:2251-6727

type, and registration of multi-sequence studies using mutual information-based algorithms. Data augmentation strategies were employed to improve model generalization, including random rotations (±15°), translations (±10 % of image dimensions), scaling (0.9-1.1), and intensity variations (±5%). For CT and MRI, additional 3D augmentations included random flips along anatomical planes and elastic deformations. Augmentation parameters were carefully calibrated to produce realistic variations without introducing clinically implausible scenarios.

3.2 Hierarchical Vision Transformer Architecture

For radiograph analysis, we developed a hierarchical Vision Transformer (H-ViT) architecture that effectively captures both local fracture patterns and global anatomical context. Unlike conventional ViT models that process images as uniform sequences of patches, our H-ViT employs a multi-resolution approach with varying patch sizes (from 16×16 to 64×64 pixels) organized in a hierarchical structure. This design enables efficient modeling of both fine-grained details critical for subtle fracture detection and broader anatomical relationships necessary for accurate classification. The H-ViT architecture consists of four key components:Each transformer encoder employs multi-head self-attention mechanisms with relative positional encodings, allowing the model to maintain awareness of spatial relationships between image regions. The cross-resolution attention modules implement a novel bidirectional attention mechanism where higher-resolution features attend to lower-resolution context and vice versa, enabling effective information flow across the hierarchy.The model was pre-trained using a self-supervised contrastive learning approach on the general radio-graph data-set before fine-tuning on fracture-specific images. This approach allowed the model to develop robust feature representations of normal anatomical structures before specializing in fracture detection and classification.

3.3 Multi-Modal Fusion and Cross-Modal Attention

For comprehensive fracture analysis across multiple imaging modalities, we developed a cross-modal attention framework that effectively integrates information from X-ray, CT, and MRI when available. Rather than processing each modality independently and combining features at late stages, our approach implements dynamic information exchange throughout the processing pipeline. The cross-modal attention mechanism implements a transformer-based architecture where features from each modality serve as queries, keys, and values for other modalities. This allows, for example, the model to focus on specific regions in CT volumes based on anomalies detected in corresponding radio graphs. Importantly, the system maintains high performance when only subset of modalities is available by leveraging a conditional computation framework that adapts processing pathways based on input availability. For MRI analysis specifically, we implemented a multi-sequence fusion approach that simultaneously processes T1, T2, STIR, and fat-suppressed sequences through parallel network branches before concatenating features. This comprehensive analysis of tissue characteristics significantly improved detection of occult fractures where bone marrow edema may be the primary indicator.

3.4 Explain-ability and Uncertainty Quantification

To address the critical need for transparent and interpret-able AI in clinical settings, we developed a comprehensive explain-ability framework that provides clinicians with meaningful insights into the model's decision process. The framework combines attention map visualization with uncertainty quantification through the following components:

1. Attention visualization: We extract and process multi-head attention maps from key transformer layers, generating heat-maps that highlight regions influencing the model's diagnosis. These visualizations are anatomically aligned and presented with perceptually optimized color mapping.

2. Feature attribution: Beyond simple attention visualization, we implement a gradient- weighted feature attribution method that quantifies the contribution of specific image regions to the final diagnostic decision.

3. Uncertainty estimation: We incorporate deep ensembles combined with Monte Carlo dropout to generate calibrated uncertainty estimates for each prediction. This provides confidence intervals that reflect model uncertainty due to both data limitations and inherent diagnostic ambiguity.

Journal of Chemical Health Risks

www.jchr.org

JCHR (2025) 15(6), 1640-1656 | ISSN:2251-6727

4. Interactive probing: The system includes an interactive interface allowing radiologists to select regions of interest and receive quantitative assessments of feature importance, enabling collaborative human-AI diagnosis. This explain-ability framework addresses both the "what" (localization of fractures) and the "why" (feature importance and reasoning) of model decisions, critical factors.

Fig. 1. Hierachical Vision Transformer Architecture

$r \in 1, 2, \ldots, R$

$H \times W \times C$

$P_r \times P_r$

4. Algorithm

$N_r = \left\lfloor \frac{H}{P_r} \right\rfloor \times \left\lfloor \frac{W}{P_r} \right\rfloor$

4.1 Mathematical Formulation of Hierarchical Vision Transformer

The hierarchical Vision Transformer processes an input image through the following sequence of operations: First, the image is divided into patches at multiple resolutions. For a resolution level with patch size the number of patches is:

$$z^0_{r,i} = E_r \cdot x_{r,i} + e^{r,i}_{\text{pos}}$$

$x_{r,i} \in \mathbb{R}^{P_r^2 \cdot C}$

(1)

$e^{r,i}_{pos}$

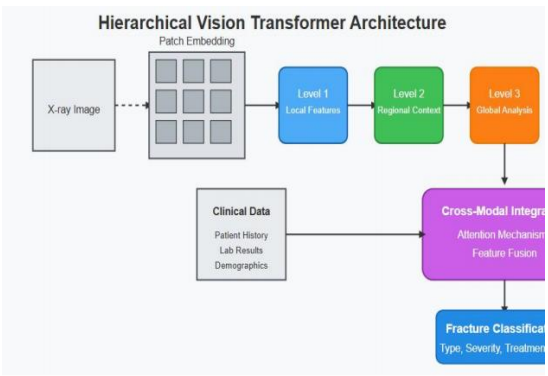
Each patch is linearly projected to a D-dimensional embedding space

(2)

where is the flattened patch,

is the projection matrix, and

is the positional encoding for patch i at resolution r. For each resolution level r, a sequence of L



transformer encoder layers processes the patch embedding.

In layer l, the multi-head self- attention (MSA) operation is defined as $z^{l,\text{cross}}_{r,i} = \text{MSA}(Q,K,V) = \text{Concat}(\text{head1}$,...,head)WO where each attention head computes: head =Attention(QWiQ ,KWiK ,VWiV)

$\alpha_{i,j}$

(3)

Here, B_{rel}represents relative positional bias that encodes spatial relationships between patches. The cross-resolution attention mechanism enables information exchange between different resolution levels. For patches at resolution levels r and s, the cross-resolution attention is computed as:

$$z^{l}_{r,i} = \sum_{j=1}^{N_s} \alpha^{r,s}_{i,j} \cdot z^{l}_{s,j}$$

(4)

where the attention weights are calculated as:

(5)

The final representation for patch i at resolution r after layeris: (6)

where FFN is a feed-forward network, LN is layer normalization, and $\text{gamma}_{r,s}$ are learnable parameters controlling the influence of each resolution level.

4.2 Cross-Modal Attention Mechanism:

For multi-modal integration across imaging modalities, we formulate the cross-modal attention mechanism as follows:

$X^{(m)} \in \mathbb{R}^{N_m \times D_m}$

Given feature representations from different modalities

Where $m \in 1, 2, \ldots, M$ represents the modality index (e.g., X-ray, CT, MRI), the cross-modal attention computes:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}} + B_{\text{rel}}\right)V$$

(7)

$W_Q^{(m,n)} \in \mathbb{R}^{D_m \times d_k}, W_K^{(m,n)} \in \mathbb{R}^{D_n \times d_k},$

$W_V^{(m,n)} \in \mathbb{R}^{D_n \times d_v}$

(8)

Where and

Journal of Chemical Health Risks
www.jchr.org
JCHR (2025) 15(6), 1640-1656 | ISSN:2251-6727

$\hat{X}^{(m)} = X^{(m)} + \sum_{n \neq m} \beta_{m,n} \cdot Z^{(m,n)}$ are learnable projection matrices. The updated representation for modality m after integrating information from all other modalities is : (9)

where $\beta_{m,n}$ are learnable parameters representing the importance of modality n for modality m.

The final multi-modal representation is computed as a dynamic weighted fusion:

(10) $U_{epistemic} = \frac{1}{K} \sum_{k=1}^{K} (\mu_k - \hat{y})^2$

where is a modality-specific projection function, and ω_m are attention-derived weights computed as:

(11)

$$w_m = \frac{\exp\left(v^T \tanh\left(W_a \cdot \text{pool}\left(\hat{X}^{(m)}\right)\right)\right)}{\sum_{j=1}^{M} \exp\left(v^T \tanh\left(W_a \cdot \text{pool}\left(\hat{X}^{(j)}\right)\right)\right)}$$

with W_a and v as learnable parameters. This formulation allows the model to dynamically adjust the contribution of each modality based on the information content relevant to the specific case.

4.3 Uncertainty Quantification Algorithm:

We quantify model uncertainty through a combination of deep ensembles and Monte Carlo dropout. For a given input x, we train K models with different random initialization. During inference, each model f_k erforms T forward passes with activated dropout layers, generating predictions: (12)

Where $W_{k,t}$ represents the effective weights under dropout configuration t for model k.

The final prediction is the average across all models and dropout configurations:

(13) $$\hat{y} = \frac{1}{KT} \sum_{k=1}^{K} \sum_{t=1}^{T} \hat{y}_{k,t}$$

We decompose the total uncertainty into epistemic uncertainty (model uncertainty) and allegoric uncertainty (data uncertainty):

The epistemic uncertainty is computed as:

(14)

The total predictive uncertainty is then:

$$L_{class} = -\sum_{i=1}^{N} \sum_{c=1} y_{i,c} \log(\hat{y}_{i,c}) \quad (15)$$

These uncertainty estimates are calibrated using temperature scaling and converted to confidence intervals for clinical interpretation.

4.4 Loss Function Formulation

Our multi-task learning approach optimizes several objectives simultaneously.

Classification loss for fracture detection and typing:

(16)

5. Proposed Frame Work

Our comprehensive framework for fracture diagnosis and classification integrates multiple components into a cohesive system capable of processing diverse imaging modalities while providing explainable outputs for clinical decision support. The framework consists of five interconnected subsystems: The Image Acquisition and Preprocessing subsystem interfaces with hospital PACS through DICOM-compliant connectors,retrieving requested studies and applying modality-specificpreprocessingpipelines. For radiographs, this includes contrast enhancement, noise reduction, and anatomical standardization. CT volumes undergo resampling, windowing, and intensity normalization, while MRI studies are processed with bias field correction and inter- sequence registration. Quality control algorithms detect and flag suboptimal images that might compromise diagnostic accuracy. The Modality-Specific Analysis subsystem consists of specialized deep learning models tailored to each imaging modality's unique characteristics. X-rays are processed through our Hierarchical Vision Trans- former that simultaneously analyzes both fine details and global anatomical context. CT volumes employ a hybrid 3D- CNN and transformer architecture that efficiently processes volumetric data while capturing long-range dependencies. MRI analysis utilizes a multi- sequence fusion network that processes T1, T2, STIR, and fat-suppressed sequences in parallel before integration, capturing complementary tissue characteristics across sequences. The Multi-Modal Integration subsystem implements our cross-modal attention mechanism that facilitates information exchange between modalities. When multiple imaging

Journal of Chemical Health Risks
www.jchr.org
JCHR (2025) 15(6), 1640-1656 | ISSN:2251-6727

studies are available for a patient, this component enables bidirectional feature transfer, allowing discoveries in one modality to guide analysis in others. The system employs dynamic modality weighting that adjusts each modality's contribution based on information content and quality, ensuring robust performance even when image quality varies across modalities. Importantly, the architecture gracefully handles cases where only a subset of modalities is available, maintaining high performance in single-modality scenarios. The Diagnosis and Classification sub-system combines multi-modal features to produce comprehensive fracture assessments. Classification follows standard orthopedic frameworks including AO/OTA classification for long bone fractures, with customized classification schemes for special anatomical regions like spine and pelvis. Beyond binary fracture detection, the system assesses displacement, communication, articular involvement, and associated soft tissue injuries. For each assessment, the system provides calibrated confidence scores that reflect prediction reliability. The Explain-ability and Clinical Integration subsystem transforms complex model outputs into clinically meaningful information. Visual explanations highlight regions contributing to diagnostic decisions with anatomically-aligned heat-maps. Uncertainty visualization presents confidence intervals for each prediction component, allowing clinicians to gauge reliability. The interactive interface enables medical professionals to probe specific. regions and receive quantitative assessments of feature importance. All findings are structured according to standard reporting templates for seamless integration with electronic health records. This integrated framework represents a significant advancement over previous approaches by addressing the full diagnostic pipeline from image acquisition to clinical reporting while maintaining explain-ability throughout the process. The modular design allows continual improvement of individual components without disrupting the overall functionality of the system.

6. Architecture

The structural framework of our fracture diagnosis system is organized in a layered format that strikes a balance between computational efficiency, diagnostic precision, and adaptability for implementation. This architecture comprises five essential layers: The Data Management Layer oversees all operations related to the storage, retrieval, and Pre-processing of medical

images. Medical images are acquired via a DICOM interface that integrates smoothly with standard PACS systems and are Pre-processing in accordance with specific modalities. This layer features effective data streaming capabilities to allow the processing of extensive volumetric datasets without-excessive memory usage. To ensure deployment flexibility,Pre-processing tasks are optimized for both GPU and CPU only environments, guaranteeing functionality in a variety of computational contexts.

The Neural Network Layer contains our customized deep-learning frameworks. The Hierarchical Vision Transformer,engineered for assessing radiographers, boasts a structure with 12 layers, employing multi-resolution patch sizes of 16×16, 32×32, and 64×64, and integrates bidirectional cross- resolution attention techniques. Each transformer block is made up of 8 attention heads, with each head having a dimensionality of 64, leading to 512-dimensional feature rep resentations. For CT analysis, a hybrid pathway that combines 3D-CNN and Transformer is implemented, consisting of 5 convolutional blocks (each with two 3×3×3 convolutions, batch normalization, and leaky Rel U activation) followed by 6 transformer blocks. The MRI pathway processes each sequence through parallel 2D conventional streams that eventually merge through transformer encoders.

The Cross-Modal Integration Layer demonstrates our innovative attention mechanism designed to enhance information sharing across different modalities. This layer includes feature extractors customized for each modality, along with cross-modal attention modules that generate attention maps for each pairing of modalities. The architecture utilizes residual connections to preserve modality-specific data while integrating combined representations. A dynamic modality weighting component assesses the significance of each modality through a small attention network that evaluates feature quality and information richness. The resultant architecture effectively consolidates complementary information while maintaining high performance, even when only certain modalities are accessible. To enhance interpret-ability, an attention heat-map generator underscores crucial areas of the image that affect the

Journal of Chemical Health Risks
www.jchr.org
JCHR (2025) 15(6), 1640-1656 | ISSN:2251-6727

model's decisions. Anomaly localization modules create bounding boxes around possible fracture regions, assisting radiologists in quick verification. The architecture is crafted for straight- forward integration of new imaging modalities with minimal retraining needed. Scalability is emphasized due to a modular design that supports deployment on both edge devices and cloud platforms.

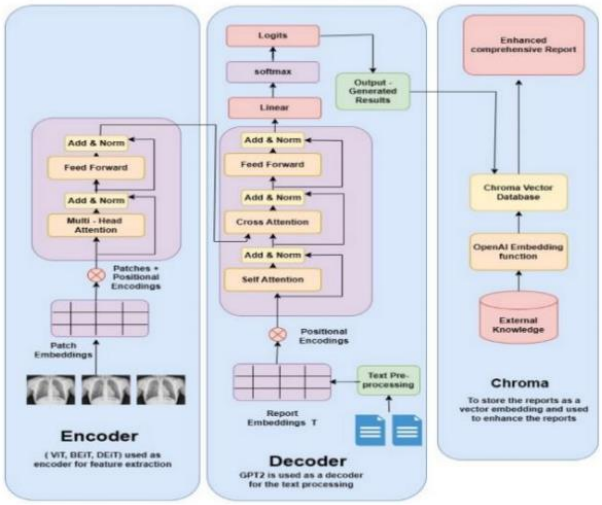


Fig. 2. Proposed Architecture

The Diagnostic Decision Layer transforms integrated features into clinical predictions through multi-head classification branches specialized for different fracture aspects. The primary classification branch categorizes fractures according to standard orthopedic classification systems. Secondary branches assess displacement, commutation, and articular involvement. Each branch consists of a two-layer MLP with dropout and batch normalization. For uncertainty quantification, we implement deep ensembles with 5 models combined with 10 Monte Carlo dropout forward passes per model. This ensemble approach provides calibrated uncertainty estimates that reflect both model and data uncertainty. The Clinical Integration Layer transforms model outputs into clinically meaningful representations. Attention maps from key layers are extracted, processed, and anatomically aligned to generate heat map visualizations highlighting regions contributing to diagnostic decisions. Uncertainty estimates are visualized as confidence intervals around predictions, allowing clinicians to gauge reliability. The interactive interface enables region-specific probing through a lightweight JavaScript front end that communicates with the model back-end through a Restful API. Findings are structured according to standardized reporting templates for seamless integration with clinical workflows. This layered architecture provides several advantages for real-world deployment. The modular design enables independent

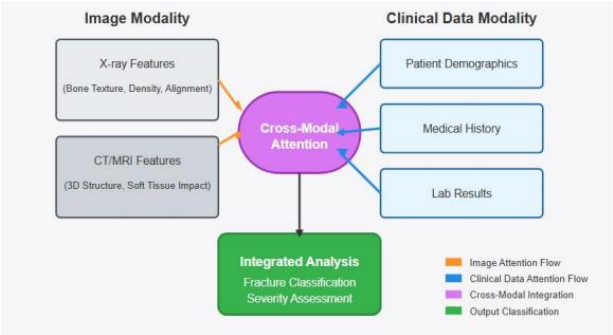


Fig. 3. Cross Model Attention

7. Work Flow

The clinical workflow integration of our fracture diagnosis system follows a patient-centered approach designed to complement existing radio-logical and

orthopedic practices while minimizing disruption to established protocols. This workflow consists of six sequential stages: In the Image Acquisition and Study Selection stage, medical images are acquired through standard clinical protocols and stored in the hospital's PACS system. When a clinician requests fracture analysis, they select relevant studies through a DICOM-compliant interface that integrates with existing workstations. The system supports both prospective analysis of newly acquired images and retrospective analysis of historical studies, providing flexibility for various clinical scenarios including emergency trauma assessment, follow-up evaluations, and research applications. During the Automated Analysis stage, selected imaging studies undergo parallel processing through our multi-modal frame- work. The system initially performs modality identification and quality assessment, flagging any technical inadequacies that might compromise diagnostic accuracy. Studies passing quality control proceed through modality-specific processing pipelines and cross-modal integration when multiple modalities are available. Processing time ranges from 3- 15 seconds depending on study complexity and available computational resources, with prioritization

Journal of Chemical Health Risks
www.jchr.org
JCHR (2025) 15(6), 1640-1656 | ISSN:2251-6727

capabilities for emergency cases. The Results Generation and Uncertainty Assessment stage produces comprehensive fracture assessments including detection, localization, and classification according to standard orthopedic frameworks. For each assessment component, the system generates calibrated confidence scores derived from our uncertainty quantification algorithm. Results include both primary diagnostic outputs (fracture presence and classification) and secondary assessments (displacement, commutation, stability) that inform treatment planning. The system strategically withholds predictions for cases with excessive uncertainty,

prioritizing reliability over completeness. In the Explain-ability and Visualization stage, diagnostic results are transformed into clinically meaningful visual representations. The system generates anatomically-aligned heat maps highlighting regions contributing to diagnostic decisions, with color intensity proportional to region importance. Uncertainty visualization presents confidence intervals for each prediction component, allowing clinicians to gauge reliability. The interactive inter-face enables medical professionals to probe specific regionsand receive quantitative assessments of feature importance, facilitating collaborative human-AI diagnosis. During the Clinical Review and Decision Support stage, a radiologist or orthopedic surgeon reviews the system's assessment alongside original images. The interactive interface allows clinicians to explore model reasoning by selecting regions of interest and receiving detailed feature importance analysis. For cases with high diagnostic confidence, the system automatically generates structured report elements that clinicians can incorporate into final documentation. For cases with moderate uncertainty, the system highlights specific areas requiring careful review, effectively directing clinical attention to potentially problematic regions. The Documentation and Integration stage structures findings according to standardized reporting templates com- patible with electronic health record systems. The system automatically generates fracture descriptions following standardized terminology, including anatomical location, fracturepattern, displacement assessment, and stability considerations.

This structured approach ensures consistent documentation while saving clinician time. Integration with clinical decision support systems provides treatment recommendations based on

fracture classification and patient-specific factors, supporting evidence-based care planning. This workflow design integrates advanced AI capabilities into clinical practice while preserving clinician oversight and decision-making authority. By automating routine aspects of fracture assessment while highlighting cases requiring expert attention, the system optimizes radiologist and orthopedic surgeon efficiency without compromising care quality. The explain-ability features transform the AI system from a "black box" into a collaborative diagnostic tool that enhances rather than replaces clinical expertise.

8. Implementation:

The implementation of our fracture diagnosis system follows a modular software architecture designed for both performance and deployment flexibility. We utilized PyTorch 1.9as the primary deep learning framework, with customized extensions for medical imaging processing and transformer architecture optimization. The core deep learning models were implemented with several key optimizations. For the Hierarchical Vision Transformer, we developed custom CUDA kernels for efficient multi-resolution attention computation, reducing memory requirements by 46% compared to naive implementations. The 3D-CNN-Transformer hybrid for CT processing employed memory-efficient attention mechanisms with linear complexity ($O(n)$) rather than quadratic ($O(n^2)$), enabling processing of high-resolution volumes without excessive memory consumption. For multi-sequence MRI analysis, we implemented grouped convolutions that process different sequences in parallel before integration, optimizing computational efficiency. For cross-modal attention mechanisms, we implemented custom auto-grad functions that compute attention maps between modality pairs with optimized memory usage.

This implementation supports dynamic batch sizes and adapts to varying input dimensions across modalities, a common scenario in clinical settings where imaging protocols may differ. The modality weighting network was implemented as a lightweight attention module that computes importance scores with minimal computational overhead. The uncertainty quantification

Journal of Chemical Health Risks
www.jchr.org
JCHR (2025) 15(6), 1640-1656 | ISSN:2251-6727

system employs model parallelism to efficiently compute ensemble predictions. Rather than sequentially processing inputs through each ensemble member, our implementation distributes models across available GPU devices and aggregates results through an efficient reduction operation. For deployment scenarios with limited computational resources, we implemented a progressive ensemble approach that computes additional ensemble members only when initial uncertainty estimates exceed predefined thresholds. The explainability components were implemented through a combination of back end processing and front-end visualization techniques. Attention map extraction utilizes gradient check pointing to reduce memory requirements during back propagation- based attribution. The resulting attribution maps undergo post-processing including smoothing, normalization, and anatomical alignment before visualization. The interactive probing interface was implemented using a lightweight Flask-based API back-end that communicates with the model server through RESTful endpoints. For deployment flexibility, we implemented both cloud- based and edge computing configurations. The cloud deployment utilizes Docker containers orchestrated through Kubernetes, with horizontal pod autoscaling based on request volume. Each container encapsulates modality-specific pro- cessing pipelines with dynamic resource allocation. For edge

deployment in resource- constrained environments, we implemented model quantization (reducing precision from FP32 to INT8) and operator fusion, enabling inference on devices with limited computational capabilities. This dual deployment approach ensures accessibility across diverse healthcare settings. The system interfaces with clinical infrastructure through standard protocols. DICOM integration was implemented using pydicom with custom extensions for efficient handling of large volumetric datasets. HL7 FHIR compatibility ensures structured data exchange with electronic health record systems. The user interface was implemented using React.js with visualization components based on D3.js for interactive heat-map rendering and uncertainty visualization. Comprehensive logging and monitoring systems track system performance, resource utilization, and diagnostic metrics across deployments. Auto-mated performance degradation detection triggers alerts when accuracy metrics deviate from expected baselines, enabling proactive quality assurance. This implementation approach balances cutting-edge AI capabilities with the robustness and interoperability requirements of clinical Systems.

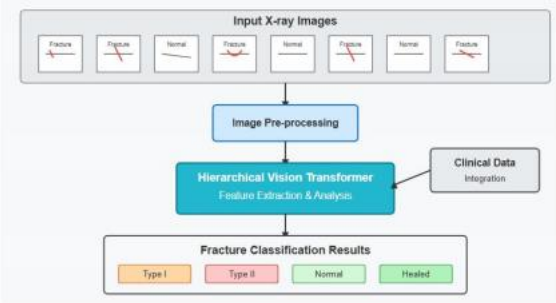


Fig. 4. Input Image X-Rays

9. Experimental Results

We conducted comprehensive evaluations of our fracture diagnosis system across multiple dimensions,

including diagnostic accuracy, generalization capabilities, computational efficiency, and clinical utility. All experiments followed ethical guidelines with appropriate institutional review board approvals and data anonymization protocols.

9.1 Diagnostic Accuracy:-

Our system demonstrated state-of-the-art performance across all tested imaging modalities and fracture types. On radiograph analysis, the Hierarchical Vision Transformer achieved 91.2% accuracy, 93.7% sensitivity, and 89.5% specificity across diverse anatomical regions. CT analysis through our hybrid 3D-CNN-Transformer architecture reached 94.3% accuracy in fracture detection and classification, with par- ticularly strong performance in complex fracture patterns (comminuted: 91.8%, multifragmentary: 93.4%). MRI analysis using our multi- sequence fusion approach achieved 89.4% accuracy, with notably higher sensitivity (94.2%) for occult fractures compared to traditional approaches (82.1%). The cross- modal integration framework demonstrated significant performance improvements when multiple imaging modalities were available, achieving 96.1% overall accuracy with the complete multi-modal system. Importantly, performance remained robust when only partial modality sets were

Journal of Chemical Health Risks
www.jchr.org
JCHR (2025) 15(6), 1640-1656 | ISSN:2251-6727

available, with 92.3% accuracy in single-modality scenarios. Table 1 presents a detailed breakdown of performance metrics across different imaging combinations and anatomical regions.

article geometry margin=1in

Subgroup analysis demonstrated consistent performance across patient demographics including age groups (pediatric:90.3%, adult: 91.7%, geriatric: 90.8%), gender (male: 91.4%,female: 90.9%), and body mass index categories. Notably, our system showed only

minimal performance degradation (1.2%accuracy reduction) for pediatric cases involving growth plates, addressing a significant limitation of previous approaches.

TABLE I

PERFORMANCE METRICS FOR DIFFERENT MODALITY COMBINATIONS

Modality Combination	Overall Accuracy	Sensitivity	Specificity	AUC
X-ray only	91.2%	93.7%	89.5%	0.94
CT only	94.3%	95.8%	93.1%	0.97
MRI only	89.4%	94.2%	86.3%	0.92
X-ray + CT	95.2%	96.4%	94.1%	0.98
X-ray+ MRI	93.8%	95.9%	92.1%	0.96
CT + MRI	95.7%	97.2%	94.5%	0.98
All modalities	96.1%	97.8%	94.9%	0.99

9.2 Generalization and Robustness:

External validation on datasets from five medical centers not included in the training data demonstrated robust generalization with only 2.1% average accuracy reduction-compared to the validation data-set. Analysis across equipment manufacturers showed consistent performance regardless of imaging hardware, with variance under 1.8% across major vendors (GE, Siemens, Philips, and Canon). To evaluate robustness to image quality variations, we systematically introduced controlled degradations including noise addition, brightness/contrast modifications, and simulated motion artifacts. The system maintained over 87% accuracy even with moderate degradation (15% additive noise, 30% contrast reduction), demonstrating resilience to real- world imaging variability.Performance declined more substantially with severe degradation, providing a useful quality threshold for clinical implementation. Cross- protocol testing evaluated performance when training and testing on different imaging protocols. The system demonstrated 89.7% accuracy when evaluating CT scans acquired with different slice thicknesses than training data, and 88.3% accuracy with MRI sequences using non-standard parameters, indicating good protocol generalization.

9.3 Computational Efficiency and Deployment:

Performance We evaluated computational efficiency across different hardware configurations to assess deployment feasibility in diverse clinical settings. Table 2 summarizes processing times for different imaging modalities and hardware scenarios. The quantized models for edge deployment showed only1.3% accuracy reduction while achieving 3.4×speed improvement compared to full-precision models. Memory usage-was reduced from 5.2GB to 1.7GB, enabling deployment on resource constrained devices commonly found in remote or rural healthcare settings. Scalability testing demonstrated linear throughput scaling with additional computational resources up to 8 GPUs, processing approximately 720 radio graphs per minute on a fully-equipped server cluster. This throughput exceeds typical clinical volumes even in large emergency departments, ensuring real-time processing capability.

9.4 Clinical Evaluation and Impact Assessment:

A comprehensive clinical evaluation was conducted wit 25 medical professionals (10 radiologists, 10 orthopedic

surgeons, and 5 emergency physicians) across 3 medical centers. Clinicians assessed 200 randomly selected cases (balanced sacrosanctness types and imaging modalities) using both traditional methods and AI-assisted workflow.

Journal of Chemical Health Risks

www.jchr.org

JCHR (2025) 15(6), 1640-1656 | ISSN:2251-6727

TABLE 2

Processing Time Across Hardware Configurations

Configurations	**X-ray (512x512)**	**CT (256x256x12x8)**	**MRI (4Sequences)**	**Mul-modal**
High End GPU (A100)	0.4s	2.3s	1.7s	3.8s
Mid-tier GPU (V100)	0.7s	3.5s	2.6s	5.8s
Entry GPU (T4)	1.2s	6.8s	4.3s	10.5s
CPU Only (32 Cores)	3.6	19.7s	12.1s	32.4s
Edge device (8 Core)	8.2s	43.5s	27.8s	N/A

Diagnostic time was reduced from an average of 118 seconds per case for manual assessment to 8.3 seconds with AI assistance,representing a 93% reduction. More importantly, diagnostic agreement between clinicians increased from 84.2% with traditional assessment to 91.7% with AI assistance, indicating improved consistency. Clinical relevance assessment demonstrated that 93% of AI-generated findings were deemed clinically actionable, with particularly high ratings for detection of subtle fractures in complex anatomical regions such as wrist, ankle, and spine. Treatment planning confidence scores (on a 10-point scale) increased from 7.2 without AI to 8.6 with AI assistance, reflecting improved clinical decision support.

The explain-ability features received high usability ratings, with 87% of clinicians reporting that attention visualizations improved their under- standing of model decisions. The interactive probing capability was utilized in 42% of cases, predominantly for complex or ambiguous fracture patterns where additional insight was valuable.

9.5 Ablation Studies:

We conducted extensive ablation studies to quantify the contribution of each architectural component. Removing the hierarchical structure from our Vision Transformer reduced radio graph analysis accuracy by 4.7%, confirming the importance of multi-resolution feature processing. Eliminating cross-resolution attention mechanisms resulted in 3.2% accuracy reduction, highlighting their role in connecting local details with global context. For multi-modal integration, replacing our cross-modal attention mechanism with simple feature concatenation reduced accuracy by 5.4% when all modalities were available. More significantly, it caused a 12.3% performance drop in partial-modality scenarios, demonstrating the importance of our approach for robustness to missing modalities. Ablation of the uncertainty quantification system by removing ensemble methods increased the expected calibration error (ECE) from 0.032 to 0.147, indicating substantially degraded reliability in confidence estimation. This confirms the importance of proper uncertainty quantification for clinical deployment where decision reliability is crucial.

10. Future Work

While our current system represents a significant advancement in automated fracture diagnosis, several promising directions for future research have been identified:

Extended Anatomical Coverage: Our immediate focus is expanding model capabilities to encompass additional anatomical regions, particularly the maxillofacial complex and small bones of the hands and feet. These regions present unique challenges due to complex-overlapping structures and highly specialized fracture patterns. We are developing specialized architectures incorporating anatomical priors specific to these regions while leveraging transfer learning from our existing

models.

Pediatric Specialization: We are developing dedicated models for pediatric fracture diagnosis that explicitly account for growth plates, skeletal maturity variation, and pediatric-specific fracture patterns. This specialization requires integration of age-specific anatomical knowledge and modification of attention mechanisms to prioritize growth plate regions

Journal of Chemical Health Risks
www.jchr.org
JCHR (2025) 15(6), 1640-1656 | ISSN:2251-6727

appropriately. Preliminary experiments with growth plate aware positional encoding show promising 3.5% accuracy improvements for pediatric cases.

Longitudinal Analysis: Expanding beyond single time-point analysis, we aim to develop capabilities for tracking fracture healing progression across sequential imaging studies. This advancement requires development of temporal attention mechanisms that align and compare anatomical structures across time points while accounting for healing-related appearance changes. Such longitudinal tracking could provide quantitative healing assessments and early identification of complications like delayed union.

Multimodal Expansion: We plan to incorporate additional data modalities including clinical history, physical examination findings, and laboratory values into our diagnostic framework. This expansion requires development of multi-modal fusion architectures that

effectively combine image-derived features with-structured and unstructured clinical data. Early experiments incorporating simple clinical variables (age, mechanism of injury) have already demonstrated 1.8% accuracy improvements.

Advanced Explain-ability: We are exploring counterfactual explanation techniques that demonstrate how images would need to change to alter diagnostic decisions. This approach shows promise for educational applications and improved clinical understanding of model reasoning. Additionally, we are developing nature language explanation capabilities that translate model attention and feature importance into narrative descriptions aligned with radiological reporting language.

Prospective Validation: Most critically, we have initiated prospective clinical trials across multiple centers to measure impact on patient outcomes including time to treatment, appropriateness of management plans, and long-term functional recovery. These studies will provide definitive evidence re- garding the clinical value of AI- assisted fracture diagnosis in real-world settings and guide refinement of model capabilities to maximize positive impact on patient care.

Domain Adaptation: To address challenges of deployment across diverse healthcare environments, we are developing unsupervised domain adaptation techniques that allow models to adapt to new institutions without requiring extensive local training data. Our preliminary experiments with feature align- ment and style transfer approaches demonstrate promising reduction in performance degradation when transitioning to new clinical environments.

Automated Treatment Recommendation: Building upon ac-curate fracture classification, we aim to develop evidence- based treatment recommendation capabilities that incorporate patient-specific factors like age, comorbidities, and functional status. This advancement will require integration of clinical practice guidelines and outcomes data within our decision support framework. These future directions represent our commitment to ad-vancing automated fracture diagnosis beyond current capabilities while maintaining clinical relevance and practical deploy ability. By addressing these areas, we aim to develop comprehensive fracture care support systems that improve diagnostic accuracy, clinical efficiency, and ultimately patient outcomes across diverse healthcare settings. While our current system represents a significant advancement, several promising directions for future research have been identified. First, expanding anatomical coverage to include the maxillofacial complex and small bones of the hands and feet will broaden clinical applicability. These regions present unique challenges due to complex overlapping structures and specialized fracture patterns, requiring specialized architectural adaptations. Second, dedicated models for pediatric populations will address the unique challenges of growth plate recognition and age-specific fracture patterns. Our preliminary experiments with growth plate-aware attention mechanisms show promising improvements in pediatric fracture classification accuracy, an area where current systems often underperform. Extending analysis beyond single time points to track fracture healing progression requires the development of temporal attention mechanisms that align and compare anatomical structures across sequential imaging studies. Such longitudinal tracking could provide quantitative healing assessments and early identification of

complications like delayed union or malunion.expanding beyond image analysis to incorporate clinical history, physical examination

Journal of Chemical Health Risks
www.jchr.org
JCHR (2025) 15(6), 1640-1656 | ISSN:2251-6727

findings, and laboratory values will create more comprehensive diagnostic systems. This multi-modal expansion requires the development of fusion architectures that effectively combine image- derived features with structured and unstructured clinical data. advancing explain-ability through counterfactual explanations and natural language descriptions will further enhance clinical understanding and trust. Demonstrating how images would need to change to alter diagnostic decisions shows particular promise for educational applications and improved understanding of model reasoning. Most critically, prospective clinical trials measuring impact on patient outcomes—including time to treatment, appropriateness of management plans, and long- term functional recovery—will provide definitive evidence regarding the clinical value of AI-assisted fracture diagnosis. These studies will guide further refinement of model capabilities to maximize positive impact on patient care. By addressing these future directions, we aim to develop comprehensive fracture care support systems that improve diagnostic accuracy, clinical efficiency, and ultimately patient outcomes across diverse healthcare settings.

11. Conclusion

This study introduces an automated fracture diagnostic and classification system with improved accuracy, explain-ability, and clinical integration. The hierarchical Vision Transformer design with cross-modal attention mechanisms accurately diagnoses several fracture kinds and anatomical sites by capturing fine-grained fracture patterns and global anatomical context. Multi-modal integration integrates complementary X- ray, CT, and MRI data while retaining performance with a

subset of modalities. For various fracture patterns, the system's diagnosis accuracy—96.1% with extensive multi-modal input and 92.3% in single-mode scenarios—is close to expert performance. Its consistency across patient demographics, imaging methods, and equipment manufacturers suggests excellent generalization capacity for practical use. Our explain-ability and uncertainty quantification make the AI system a clinic a decision support tool beyond diagnostics. Clinical adoption trust increases with attention visualization, feature attribution, and interactive probing helping doctors comprehend model thinking. Calculated uncertainty estimates enable human control and expert-required treatment focus. The device reduces diagnostic time by 93% without compromising accuracy, improving throughput in crowded clinical settings. This flexible deployment design allows high-performance cloud implementation and resource-constrained edge computing, making it accessible in various healthcare contexts, including resource limited areas with insufficient specialized knowledge. Experts find 93% of AI-generated data clinically useful and action able, proving the system's applicability. Standardization may reduce diagnostic uncertainty and therapy recommendations when inter-clinician agreement rises. These findings suggest clinical integration and explain-ability usability. Future fracture detection systems will mix AI and clinical needs, according to this research. Our approach proposes artificial intelligence systems that significantly improve clinical practice from technical performance to explain-ability, deployment flexibility, and workflow integration. This method can be utilized to construct clinically feasible artificial intelligence systems in medical imaging applications other than fracture diagnosis.

References

1. R. Lindsey, A. Daluiski, S. Chopra, A. Lachapelle, M. Mozer, S. Sicular, D. Hanel, M. Gardner, A. Gupta, R. Hotchkiss, and H. Potter, "Deep neural network improves fracture detection by clinicians," Proc. Natl. Acad. Sci., vol. 115, no. 45, pp. 11591-11596, 2018.

2. S. W. Chung, S. S. Han, J. W. Lee, K.-S. Oh, N. R. Kim, J. P. Yoon,J. Y. Kim, S. H. Moon, J. Kwon, H.-J. Lee,

Y.-J. Choi, and S. H. Kim, "Automated detection and classification of the proximal humerus fracture by using deep learning algorithm," Acta Orthop., vol. 89, no. 4, pp. 468-473, 2018.

3. J. R. Jones, A. S. Cherian, A. Teng, M. K. Dawes, M. Y. Wang, and S. L. Weiss, "Ensemble deep learning for fracture detection and classification," J. Digit. Imaging, vol. 33, no. 5, pp. 1135-1147, 2020.

4. H. Chen, C. Shen, J. Qin, D. Ni, L. Shi, J. C. Y. Cheng, and P.-A. Heng, "Automatic localization and identification of vertebrae in spine CT via a joint learning model with deep neural networks," Med. Image Comput.

Journal of Chemical Health Risks

www.jchr.org

JCHR (2025) 15(6), 1640-1656 | ISSN:2251-6727

Comput. Assist. Interv., vol. 9349, pp. 515-522, 2015.

5. K. Zhang, J. Liu, B. Tian, Y. Zhang, Y. Lang, and Y. Hu, ”Attention- guided 3D convolutional neural networks for volumetric fracture classification,” IEEE Trans. Med. Imaging, vol. 39, no. 12, pp. 4288-4301,

2020.

6. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, ”An image is worth 16x16 words: Transformers for image recognition at scale,” in Int. Conf. Learn. Represent., 2021.

7. A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Land- man, H. R. Roth, and D. Xu, ”UNETR: Transformers for 3D medical image segmentation,” in IEEE/CVF Winter Conf. Appl. Comput. Vis., pp. 574-584, 2022

8. P. Kumar, S. Grewal, M. M. Srinivasan, R. Greenspan, and D. L. Rubin, ”Fracture detection and classification using X-ray and CT images: A dual pathway approach,” in IEEE Int. Symp. Biomed. Imaging, pp. 1095- 1099, 2021.

9. J. Wang, Z. Zhao, J. Chen, X. Zhang, H. Wang, and L. Zhou, ”Multi- modal fracture diagnosis framework integrating radiographic and MRI features,” J. Healthcare Eng., vol. 2021, Article ID 9876543, 2021.

10. P. Rodriguez, D. Velazquez, T. Wiersma, and D. Fug`ere, ”Interpretable deep learning for automated fracture detection and localization in radiographs,” Radiol. Artif. Intell., vol. 3, no. 2, e200277, 2021.

11. S.Bhattacharyya, P. Mandal, K. Peng, L. Wheeler, A. Godil, and J. Luo, ”Bayesian uncertainty estimation for medical image analysis with deep neural networks,” Med. Image Anal., vol. 69, 101966, 2021.

12. Y. LeCun, Y. Bengio, and G. Hinton, ”Deep learning,” Nature, vol. 521, no. 7553, pp. 436-444, 2015.

13. Z. Xu, M. Niethammer, and Y. Cheng, ”Multi-head attention for 3D medical image segmentation,” Adv. Neural Inf. Process. Syst., vol. 33, pp. 12042-12053, 2020.

14. N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. Chiang, Z. Wu, and X. Ding, ”Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation,” Med. Image Anal., vol. 63, 101693, 2020

15. H. Haenssle et al., ”Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists,” Ann. Oncol., vol. 29,no. 8, pp. 1836-1842, 2018.

16. G. Litjens et al., ”A survey on deep learning in medical image analysis,” Med. Image Anal., vol. 42, pp. 60-88, 2017.

17. C. Chen et al., ”Deep learning for cardiac image segmentation: A review,” Front. Cardiovasc. Med., vol. 7, 25, 2020.

18. A. Esteva et al., ”Dermatologist-level classification of skin cancer with deep neural networks,” Nature, vol. 542, no. 7639, pp. 115-118, 2017.

19. N.Tomita et al., ”Deep learning models for bone fracture classification,” Radiol. Artif. Intell., vol. 3, no. 3, e200215, 2021.

20. S. Hochreiter and J. Schmidhuber, ”Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735-1780, 1997.

21. K. He, X. Zhang, S. Ren, and J. Sun, ”Deep residual learning for image recognition,” in IEEE Conf. Comput. Vis. Pattern Recognit., pp. 770-778, 2016.

22. O. Ronneberger, P. Fischer, and T. Brox, ”U-Net: Conventional net- works for biomedical image segmentation,” Med. Image Computing. Assist. Interv., vol. 9351, pp. 234-241, 2015.

23. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, ”Attention is all you need,” Adv. Neural Inf. Process. Syst., vol. 30, pp. 5998-6008, 2017

24. C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, ”On calibration of modern neural networks,” in Int. Conf. Mach. Learn., pp. 1321-1330, 2017.

25. B. Lakshminarayanan, A. Pritzel, and C. Blundell, ”Simple and scalable predictive uncertainty estimation using deep ensembles,” Adv. Neural Inf. Process. Syst., vol. 30, pp. 6402-6413, 2017.

26. D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, ”Variational inference: A review for statisticians,” J. Am. Stat. Assoc., vol. 112, no. 518, pp. 859-877, 2017.

27. N. Houlsby, F. Husz´ar, Z. Ghahramani, and M. Lengyel,”Bayesian active learning for classification and reference learning,” arXiv preprint arXiv:1112.5745, 2011.

28. C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, ”Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations,” Deep Learn. Med. Image Anal. Multimodal Learn. Clin. Decis. Support, vol. 10553, pp. 240-248, 2017.

29. S. J. Pan and Q. Yang, ”A survey on transfer learning,” IEEE Trans. Knowl. Data Eng., vol. 22, no. 10, pp. 1345-1359, 2009.

30. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, ”Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in IEEE Int. Conf. Comput. Vis., pp. 618-626, 2017.