Received: 14 August 2025 | Revised: 5 March 2026 | Accepted: 23 March 2026 | Published online: 7 April 2026

RESEARCH ARTICLE

Artificial Intelligence and Applications
2026, Vol. 00(00) 1–9
DOI: 10.47852/bonviewAIA62027209

BON VIEW PUBLISHING

# MSA-CSpineNet: A Multi-Scale Spatial Attention Deep Learning Framework for Cervical Spine Fracture Diagnosis

**Oladosu Oyebisi Oladimeji[1,*] and Ayodeji Olusegun Ibitoye[2]**

[1] *Faculty of Engineering and Computing, Atlantic Technological University, Ireland*

[2] *School of Computing and Mathematical Sciences, University of Greenwich, UK*

**Abstract:** Cervical spine fractures can cause instability in the cervical spine and may result in spinal cord injuries. If not promptly detected and treated, these fractures may deteriorate over time. Hence, the diagnosis of cervical spine injuries must be conducted urgently to prevent further complications. Current deep learning models face limitations in accurately diagnosing these fractures due to issues such as insufficient attention to subtle fracture features and poor generalization across varying scales. This research proposes MSA-CSpineNet (Multi-Scale Spatial Attention Cervical Spine Network), a deep learning framework for accurate cervical spine fracture diagnosis from computed tomography scans. The pre-trained MobileNet was used for feature extraction, which was passed to the multi-scale and attention module for relevant feature selection. The results show an accuracy of 99.75%, sensitivity of 99.99%, specificity of 99.50%, and precision of 99.50%. Compared with the existing state-of-the-art approaches that used transfer learning and conventional convolutional neural network techniques, experimental results demonstrated that the proposed MSA-CSpineNet outperforms existing methods in image classification. The results of this research have the potential to greatly improve cervical spine fracture early diagnosis and treatment, which would benefit patients' outcomes. Gradient-weighted class activation mapping visualization demonstrates that the model develops spatially selective attention patterns, providing interpretability that supports clinical trust in model predictions.

**Keywords:** cervical spine fracture, attention mechanism, fracture, multi-scale, computed tomography

## 1. Introduction

The human spine plays a vital role in maintaining upright posture and enabling coordinated body movement [1]. The cervical spine (C-spine) consists of seven vertebrae in the neck region and supports the head's weight while allowing for a wide range of head and neck movements. In Southeast Norway, 2153 individuals experienced one or more C-spine fractures over a 5-year period from 2015 to 2019, with an incidence rate of 14.9 per 100,000 person-years [2]. These fractures occur due to abnormal movements or combinations of movements such as hyperflexion, rotation, hyperextension, lateral bending, and axial loading of the spinal column [3]. These fractures can lead to C-spine instability and potential spinal cord injuries [4, 5]. Consequently, C-spine injuries can result in significant morbidity and mortality among trauma patients [6]. Thus, a delay in identifying an unstable fracture, which leads to inadequate immobilization, can result in catastrophic neurological deterioration with severe consequences [7]. Therefore, quick and timely diagnosis of C-spine fractures is crucial for optimal patient outcomes.

Traditional diagnostic methods, such as X-rays and computed tomography (CT) scans, are the standard, most suitable, and commonly used imaging modality for C-spine fracture diagnosis [4]. However, these methods may miss subtle fractures or underestimate the severity of the injury. Moreover, interpreting these images often relies heavily on the expertise of radiologists, which introduces a degree of subjectivity and potential for human error [8]. C-spine fracture diagnosis using medical imaging faces several challenges. The process is time-consuming, costly, and at times not readily accessible in primary care settings, coupled with the limited availability of radiologists. Additionally, the symptoms experienced by individuals with a C-spine fracture can vary depending on which vertebra is fractured [3]. These factors underscore the need for more reliable and precise diagnostic tools. Accurate diagnosis of C-spine fractures is paramount for effective treatment planning and preventing long-term complications. Misdiagnosis could result in worsening a patient's health and increase treatment time and effort [9], and in severe cases, a misdiagnosis might even result in death.

C-spine fractures are critical injuries often resulting from high-impact trauma, such as vehicular accidents, falls, or sports-related activities. These fractures involve the bones in the neck region, specifically the seven cervical vertebrae, and can range from minor cracks to severe dislocations. Due to the proximity of

***Corresponding author:** Oladosu Oyebisi Oladimeji, Faculty of Engineering and Computing, Atlantic Technological University, Ireland. Email: oladosu.oladimeji@atu.ie

these vertebrae to the spinal cord, C-spine fractures pose significant risks, including paralysis or even death if not promptly and adequately managed [10]. The prevalence of C-spine injuries, particularly in elderly populations, is on the rise due to the increasing incidence of falls and osteoporosis [11]. C-spine fractures are a major healthcare concern because of their association with high morbidity and mortality rates, as well as their potential to severely affect neurological function and the overall well-being of the patient [3]. Studies show that 10–11% of all C-spinal fractures lead to spinal cord injuries [12]. Hence, a delay in diagnosing an unstable fracture, leading to insufficient immobilization, can cause a catastrophic deterioration in neurologic function, resulting in severe and potentially irreversible consequences [13]. The incidence of trauma-related C-spine fractures has been steadily increasing over time [7].

While prompt diagnosis and intervention can help reduce morbidity and mortality in patients with C-spine injuries, assessing the spine is time-consuming, as fractures can be very subtle, increasing the risk of underdiagnosis or delayed diagnosis [14]. Therefore, understanding the full range of C-spine fractures empowers emergency physicians to make informed clinical decisions, develop targeted treatment protocols, optimize the use of imaging resources, and provide accurate prognoses [7]. Deep learning has had a transformative impact on medical imaging, enabling the development of models that can analyze and interpret complex medical images with unprecedented accuracy. These models have been applied to various tasks, including image classification, segmentation, detection of anomalies, and even predicting patient outcomes. For instance, deep learning models have been successfully used to detect lung nodules in CT scans, identify diabetic retinopathy in retinal images, and classify skin lesions in dermatology images [14]. It has also been used for breast cancer diagnosis [15]. Consequently, deep learning models have shown promise as effective tools to support healthcare professionals in accurately detecting C-spine fractures.

Research studies have focused on applying deep learning for C-spine diagnosis. These models can analyze vast amounts of imaging data, such as MRI and CT scans, to detect fractures, assess their severity, and predict potential complications [16]. Integrating deep learning in medical imaging holds significant potential for improving diagnostic accuracy and patient care. Paul et al. [3] proposed a transfer learning method for real-time diagnosis of C-spine fracture from CT scans. The result shows MobileNetV2 performed better than other baseline models including ResNet and Inception. Similarly, AlexNet and GoogleNet architectures were utilized by Naguib et al. [10] to classify C-spine injuries as either fractures or dislocations from X-ray images. Thereafter, saliency maps were used for model explainability. In the study, GoogleNet performed better than AlexNet and the radiologists. Bezabh et al. [9] concatenated features extracted from X-ray images using LeNet and AlexNet architectures for C-spine disease classification. The model was found to perform better than the other single baseline models, including ResNet, VGG16, InceptionNet, and GoogleNet. In addition, Meadi et al. [17] developed a custom convolutional neural network (CNN) model consisting of multiple convolutional layers with progressively increasing filters and a 3-unit kernel size for C-spine fracture detection. The custom model performed better than baseline models, including ResNet50V2 and VGG16. Tanwar et al. [18] developed a CNN, which entails three pairs of convolutional layers, each pair containing two convolutional layers and $3 \times 3$ kernel filters detect C-spine fractures from CT scans. The model achieved high accuracy, though it was not compared with other baseline models.

Despite the C-spine's significance in human medical science, only a few studies have focused on detecting C-spine fractures [3]. This is because of factors like limited datasets, which can restrict the generalizability of deep learning models [19]. Also, C-spine fractures can vary greatly in appearance depending on factors like the patient's age, bone density, and the nature of the trauma. Hence, deep learning models often generalize poorly when confronted with this variability, which limits their effectiveness across diverse patient populations [20]. While multi-scale methods and spatial attention techniques have been explored in various medical imaging applications, including fracture detection and classification, this study introduces Multi-Scale Spatial Attention Cervical Spine Network (MSA-CSPineNet), an innovative approach by incorporating multi-scale spatial attention into a pre-trained MobileNet for C-spine fracture classification. This mechanism allows the model to dynamically focus on important regions of the CT scan by analyzing spatial dependencies at multiple scales. Smaller fractures, which may be overlooked by conventional methods, are captured by fine-grained attention mechanisms, while larger anatomical structures and broader contextual information are modeled using coarser scales.

This multi-scale feature extraction enables the model to better handle the varying size and complexity of C-spine fractures. The novel approach, which integrated multi-scale spatial attention into deep learning models for C-spine fracture diagnosis, provides several advantages. First, the approach would allow the model to dynamically focus on different regions of the image at various scales, which is crucial when dealing with complex and diverse medical imaging data like C-spine fractures. Additionally, focusing on critical regions of the CT scan helps reduce the model's reliance on irrelevant background information, further enhancing its classification performance. This targeted focus not only improves model accuracy but also aligns more closely with radiological interpretation practices, making the model's decision-making process more transparent and clinically interpretable. The rest of this article is structured as follows: The second section reviews the relevant literature, while the third section details the methodology employed in the study. The fourth section presents the experimental results and discusses their implications, and the fifth section offers the conclusions drawn from the findings.

## 2. Literature Review

The application of deep learning in detecting spine and bone fractures, including C-spine fractures, has shown considerable promise. CNNs have been used to develop automated systems that can identify fractures in medical images such as X-rays, CT scans, and MRIs with high precision. Recent studies [21] have demonstrated that these models can achieve diagnostic accuracy comparable to, and in some cases even surpassing, that of experienced radiologists. Deep learning models have been developed to automatically detect and classify fractures from imaging data. These models can identify subtle fractures that might be missed by traditional imaging techniques, providing a more objective and standardized assessment of fracture severity. Moreover, the integration of deep learning models into clinical workflows has the potential to significantly reduce the time required for diagnosis, which is crucial in emergency settings where rapid decision-making is essential [22]. Hence, there is a growing body of research utilizing deep learning in medical imaging and

diagnostics for C-spine and the entire human spine. ResNet-based CNN has been adapted for C-spine fracture detection by incorporating residual connections to improve gradient flow during training [23].

Another noteworthy approach involves the use of 3D CNNs, which consider the volumetric nature of CT and MRI scans, allowing for more comprehensive feature extraction and improved fracture detection accuracy [24]. Also, hybrid models that combine CNNs with Recurrent Neural Networks (RNNs) or attention mechanisms have been explored to enhance performance further. These models leverage the spatial features captured by CNNs and the temporal or sequential patterns identified by Long Short-Term Memory (LSTM), enabling better context understanding and fracture classification [25]. Gaikwad et al. [26] utilized You Only Look Once version 5 (YOLOv5) to detect both major and minor fractures across the C1–C7 vertebrae, while a deep learning network was used to classify vertebrae as either normal or fractured. The model achieved an accuracy of 89%. Using the transfer learning approach, Karno et al. [1] utilized eight variants of EfficientNet (B0–B7) for C-spine fracture classification in CT scans. Based on the experiment, B6 performed best with an accuracy of 99.4%. With the recent introduction of Vision Transformers (ViT), which perform better than CNNs and are more efficient in the usage of computational resources, Chlad and Ogiela [19] used YOLOv5 to accurately locate and extract the necessary information to identify damaged vertebrae in CT scans and ViT to classify the detected region.

The model achieved an accuracy of 98%. Despite the existing works on this subject matter, interpreting C-spine scans can be particularly challenging, especially in older adults, due to the frequent presence of overlapping degenerative conditions and osteoporosis, which complicate the detection of fractures [27]. Furthermore, C-spine fractures can have variations in appearance based on factors like the patient's age, bone density, and the nature of the trauma. Hence, deep learning models often generalize poorly when confronted with this variability, which limits their effectiveness across diverse patient populations [20]. Regardless of being a critical concern in human medicine, only a few studies have focused on detecting C-spine fractures. Moreover, only a few of the existing studies have achieved high accuracy. To the best of our knowledge, no one has explored the impact of multi-scale spatial attention for C-spine fractures to assist medical practitioners.

## 3. Research Methodology

In this research, we propose MSA-CSpineNet, a deep learning model with multi-scale spatial attention mechanisms to enhance the detection and diagnosis of C-spine fractures. This approach enables the model to focus on relevant anatomical regions at varying scales, improving diagnostic accuracy in complex medical images such as X-rays, CT, and MRI scans. The methodology involves several key stages: data preprocessing, model design and architecture, training and validation, and performance evaluation. Initially, in section 3.1, the dataset and its preprocessing are described to showcase image quality and model generalizability. Next, section 3.2, a multi-scale CNN integrated with spatial attention modules is presented, allowing the model to capture both local and global context within the images. The spatial attention mechanism emphasizes critical regions associated with fractures, while the multi-scale feature extraction ensures that subtle variations in fracture patterns across different scales are captured. After training the model, we evaluate its performance on metrics such as accuracy, precision, specificity, and F1-score, while also comparing it with existing baseline models.

### 3.1. Dataset description and preprocessing

This research utilized the Spine Fracture Prediction from the CT dataset, which consists of randomly selected images obtained from the RSNA Cervical Spine Fracture Detection Challenge [28]. This dataset was also used by [3], providing a basis for comparison with existing work. The dataset was divided into two sections for training (80%) and validation set (20%). Additionally, a separate testing set of 200 images per class (400 total), not involved in training or parameter tuning, was used for final model assessment. Each section contains two classes: normal C-spine images and fractured C-spine images. The details of the dataset are given in Table 1.

**Table 1**
**The description of the dataset**

| Class | Training | Validation | Testing | Total |
|---|---|---|---|---|
| Fracture | 1520 | 380 | 200 | 2100 |
| Normal | 1520 | 380 | 200 | 2100 |

From Table 1, the original training dataset consisted of 1900 samples per class. This was further partitioned into 80% (3040 samples) for model training and 20% (760 samples) for validation. In addition, a separate hold-out validation set comprising 200 samples per class (400 total) was used to assess the model's performance on unseen data from the same source. To prevent overfitting, the hold-out validation set was completely isolated from the training pipeline and was never accessed during model development or hyperparameter tuning. Additionally, no data augmentation techniques (rotation, flipping, scaling, or brightness adjustment) were applied during training, completely eliminating any risk of information leakage between splits.

The images were resized to 224 × 224 pixels using the OpenCV resize function to ensure uniform and compatible input sizes for the model. Additionally, min–max normalization was applied to scale the pixel values to a normalized range of 0 and 1. This normalization prevents overfitting and facilitates proper computation.

### 3.2. Architecture of MSA-CSpineNet

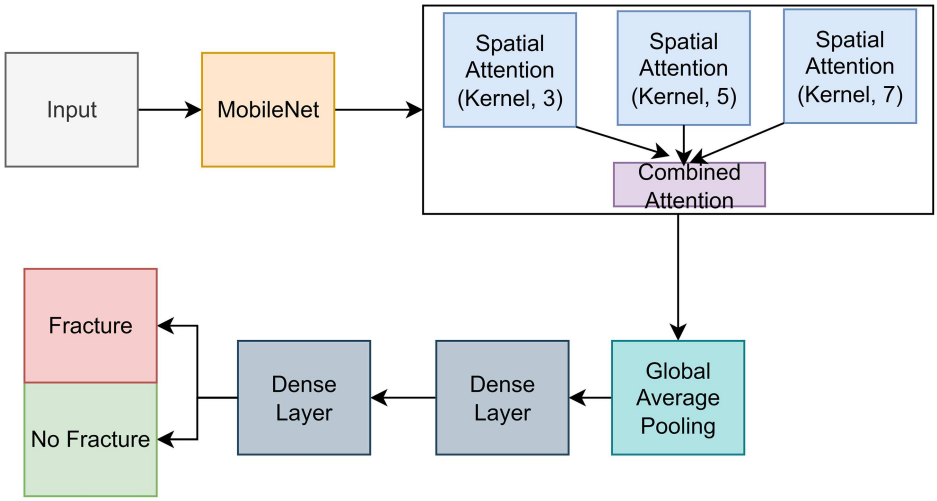
The proposed model architecture comprises a pre-trained MobileNet base integrated with a novel multi-scale spatial attention mechanism, as shown in Figure 1. The detailed components of this architecture are described as follows:

The preprocessed images are fed into a pre-trained MobileNet backbone, which serves as the feature extraction module, while the multi-scale spatial attention mechanism integrates features from sub-networks, enhancing the discovery of correlated complementary information from various scales using attention mechanisms.

#### *3.2.1. MobileNet base architecture*

The MobileNet model pre-trained on the ImageNet dataset was utilized as the base model [29, 30]. MobileNet is a lightweight and efficient CNN designed for mobile and embedded vision applications. The top layers responsible for classification were

**Figure 1**
**Architecture of MSA-CSpineNet for cervical spine fracture diagnosis**




removed to allow for customization with our multi-scale spatial attention mechanism and cervical fracture classification layers. The choice of MobileNet is justified by several factors. First, it uses depthwise separable convolutions, which factorize a standard convolution into a depthwise convolution followed by a pointwise convolution. This architecture can capture essential features in complex C-spine images, addressing challenges of variability in fracture presentations, such as different types of fractures and bone density differences. Furthermore, MobileNet significantly reduces the computational cost and model size. Additionally, MobileNet has been proven to have effective performance in medical imaging tasks [31].

### *3.2.2. Multi-scale spatial attention mechanism*

The attention mechanism was applied to the output feature maps of the MobileNet to enhance the relevant features for classification. The attention mechanism focuses on the key features of the input image, allowing the CNN to allocate computational resources to these important features during training [32]. A multi-scale spatial attention mechanism was proposed in this research to enhance the feature extraction process. This mechanism helps the model focus on relevant regions in the images at multiple scales. Three different kernel sizes (3, 5, and 7) were used to capture features at various scales, enabling the detection of fractures of different sizes and complexities. Unlike conventional attention mechanisms that operate at a single scale, the proposed MSA captures spatial dependencies across multiple receptive fields to better localize subtle fractures of varying sizes and morphologies [33].

Let $F \in \mathbb{R}^{H \times W \times C}$ denote the feature map output from the last layer of MobileNet. The MSA module applies three parallel spatial attention branches, each using a different convolutional kernel size k $\in\{3, 5, 7\}$, to capture fine, medium, and coarse-scale spatial contexts. For each kernel size in the model, the spatial attention module uses average and max pooling to generate two context descriptors, as described in Equations (1) and (2). Average pooling $A_P(F)$ and max pooling $M_P(F)$ are applied along the channel dimension to generate two spatial context descriptors:

$$\text{Average Pooling:}\quad A_P(F) = \frac{1}{C}\sum_{C=1}^{C} F(i,j,c) \tag{1}$$

$$\text{Max Pooling:}\quad M_P(F) = max_{c=1}^{C} F(i,j,c) \tag{2}$$

where $(i, j)$ denote spatial coordinates and c indexes the channel dimension.

The average pooling captures the overall feature intensity at each location, providing a smooth representation of anatomical structures. Max pooling identifies the most salient feature responses, highlighting potential fracture indicators that produce strong activations in at least one channel as described in Equation (3).

$$A_P(F),\ M_P(F) \in \mathbb{R}^{H \times W \times 1} \tag{3}$$

The input feature map $F$ undergoes both average pooling and max pooling operations along the channel dimension, producing two spatial context descriptors $A_P(F)$ and $M_P(F)$, each with dimensions $H \times W \times 1$. Average pooling captures the overall intensity distribution, while max pooling highlights the most prominent features in each spatial location. These descriptors are concatenated along the channel dimension, creating a combined spatial descriptor with dimensions $H \times W \times 2$ that captures both average and maximum response patterns.

The concatenated descriptor is passed through a convolutional layer with kernel size k, followed by a sigmoid activation function as described in Equation 4. This produces a spatial attention map $M_k \in \mathbb{R}^{H \times W \times 1}$, where each spatial location receives a weight between 0 and 1, indicating its importance for fracture detection.

$$M_k = \sigma(Conv_k\left(Concat\left(A_p(F),\ M_p(F)\right)\right)) \tag{4}$$

where $\sigma$ denotes the sigmoid activation function, ensuring attention weights are in [0, 1]. This attention map $M_k$ is then multiplied by the input feature map $F$, as shown in Equation (5), to highlight important regions and produce a scale-specific enhanced feature map. This operation amplifies relevant spatial regions while suppressing irrelevant background information.

$$F_k^{att} = M_k \bigodot F \tag{5}$$

The attention maps from each scale are refined using $1 \times 1$ convolutions. This refinement step allows the model to learn to adjust the attention maps for each channel of the input feature.

**Table 2**
**The pseudocode of the spatial attention module for the C-spine fracture diagnosis**

| Input: Feature map from MobileNet |
|---|
| Spatial Attention Module |
| Step 1: Compute the average pooling over the channel dimension of the feature map |
| Step 2: Compute the max pooling over the channel dimension of the feature map |
| Step 3: Concatenate the average pooling and max pooling |
| Step 4: Apply a Convolution with 1 filter and the specified kernel size to generate the attention map |
| Step 5: Apply sigmoid activation to obtain the attention weights |
| Step 6: Multiply the input feature map by the attention weights for the final attention-augmented feature map |
| Output: Return the attention-augmented feature map |

**Table 3**
**The pseudocode of the multi-scale module for the C-spine fracture diagnosis**

| Input: Feature Map |
|---|
| Step 1: Apply spatial_attention_module with kernel size = 3 |
| Step 2: Apply spatial_attention_module with kernel size = 5 |
| Step 3: Apply spatial_attention_module with kernel size = 7 |
| Step 4: Combine the outputs from the three attention modules using element-wise addition |
| Output: Return the combined attention-augmented feature map |

The outputs from the spatial attention modules with different kernel sizes are combined using element-wise addition as described in Equation (6). This element-wise addition aggregates the attention maps, preserving the information captured by each kernel size without increasing the dimensionality of the feature maps.

$$F^{out} = F_3^{att} + F_5^{att} + F_7^{att} \tag{6}$$

The combination of attention outputs helps the model to focus on relevant regions at different scales, potentially improving its ability to detect fractures that vary in size and appearance. Tables 2 and 3 provide a detailed stepwise pseudocode of the model.

### *3.2.3. Model interpretability*

To enhance clinical trust and enable validation of the model's decision-making process, gradient-weighted class activation mapping (Grad-CAM) was integrated as an interpretability mechanism. Grad-CAM generates visual explanations by highlighting the regions of the input image that most strongly influence the model's classification decision. Grad-CAM was applied to the final convolutional layer of the MobileNet backbone, as this layer contains high-level semantic features while maintaining spatial resolution sufficient for anatomical localization.

## 4. Experimental Results and Discussion

The algorithms used in this research were developed using Keras 2.15.0, Python 3.10.12, and the open-source computer vision library, OpenCV. The computing environment for this research was a Windows 11 (64-bit) system, utilizing an Intel(R) Core(TM) i7-12850HX processor with a 2.10 GHz clock speed. In the experiments, the Adam optimizer was used with a batch size of 16 and the binary cross-entropy loss function. The model was trained for 15 epochs.

The hyperparameters were selected based on preliminary experiments and practical considerations. The choice of 15 epochs was determined through monitoring the validation performance during initial trials, where the model demonstrated convergence and stable performance by epoch 12–15, as evidenced in Figure 2. Training beyond 15 epochs showed no significant improvement in validation accuracy while increasing the risk of overfitting. The batch size of 16 was chosen as a balance between computational efficiency and model stability. The training and validation curves were monitored closely throughout all 15 epochs (Figure 2) to detect early signs of overfitting. The close alignment between training and validation loss curves, coupled with the absence of divergence after epoch 12, indicates that the model generalized well without overfitting.

The 15 epochs ran for 3990.481 s. The performance of the model was assessed using metrics such as accuracy, precision, recall, and AUC. The results of these evaluations are summarized in Table 4. Figure 2 illustrates the accuracy and loss plot over epochs for the training and validation sets. The model was evaluated on two datasets: an internal validation set (760 samples) used for performance monitoring during development and a separate test set (400 samples) used for final evaluation.

While real-time performance is not strictly required in most C-spine trauma workflows, timely diagnosis remains critical, as it can provide immediate preliminary feedback in emergencies where radiologist availability may be delayed, particularly in rural or under-resourced settings, and it maintains workflow efficiency without creating bottlenecks in the diagnostic pipeline. Hence, the processing time of the model makes it suitable.

Based on the trends of both lines, coupled with the absence of significant divergence, the results indicate that the model performs well and maintains its generalizability, leading to a successful and

**Figure 2**
**The accuracy and loss graph of the training process of the model**

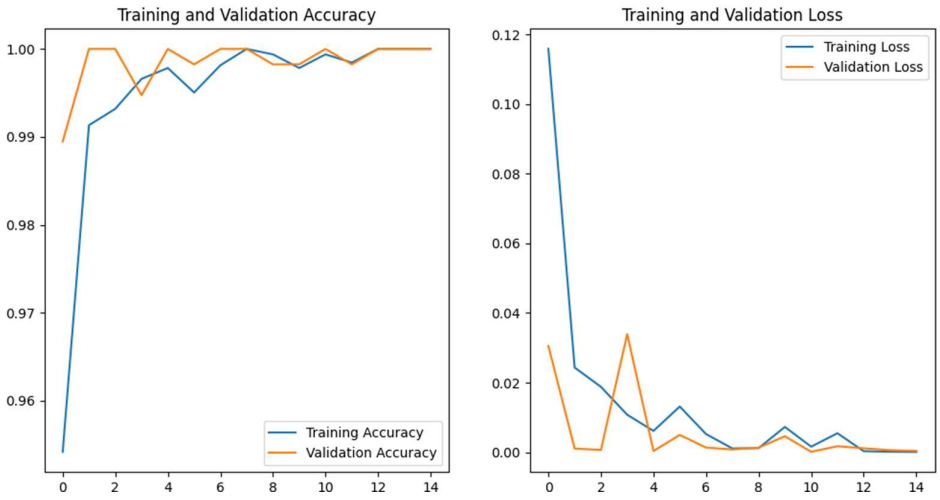



**Table 4**
**Summary of the model's performance based on evaluation metrics**

| Metric | Validation set | Test set |
|---|---|---|
| Accuracy | 0.9934 | 0.9975 |
| Precision | 0.9935 | 0.9950 |
| AUC | 0.9999 | 0.9999 |
| F1-score | 0.9934 | 0.9974 |
| Recall | 0.9934 | 0.9999 |
| Specificity | 1.00 | 0.9950 |
| Time | 7.9971 s | 4.4164 s |

desirable training outcome. The consistency in the loss shows the model's good performance without overfitting. The initial spikes in validation loss reflect the model adjustments during early epochs, but as training progresses, these smooth out, confirming that the model has learned to accurately distinguish between fractures and no fractures in the C-spine images. The Receiver Operating Characteristic Area Under the Curve (ROC-AUC), as shown in Figure 3, demonstrates a high AUC, indicating improved predictive accuracy, demonstrating the model's capability to make correct predictions across the classes.

To validate the effectiveness of the proposed MSA-CSpineNet model, a comparative evaluation against a range of modern deep learning architectures was conducted, as shown in Table 5. All models were trained under identical conditions (same dataset split, preprocessing, optimizer, batch size, and epochs) for fair comparison.

The MSA-CSpineNet outperforms all baseline models across all metrics. With the outstanding results achieved, a comparative analysis was conducted to evaluate this approach against existing state-of-the-art methods in the literature. Table 6 gives the details of the comparison. The comparative results showed that the proposed model outperformed other techniques.

To validate that the model focuses on clinically relevant anatomical features rather than spurious correlations, interpretability analysis using Grad-CAM visualization was conducted on test set images. Figure 4 presents representative examples

**Figure 3**
**The ROC-AUC curve and the confusion matrix of the model**

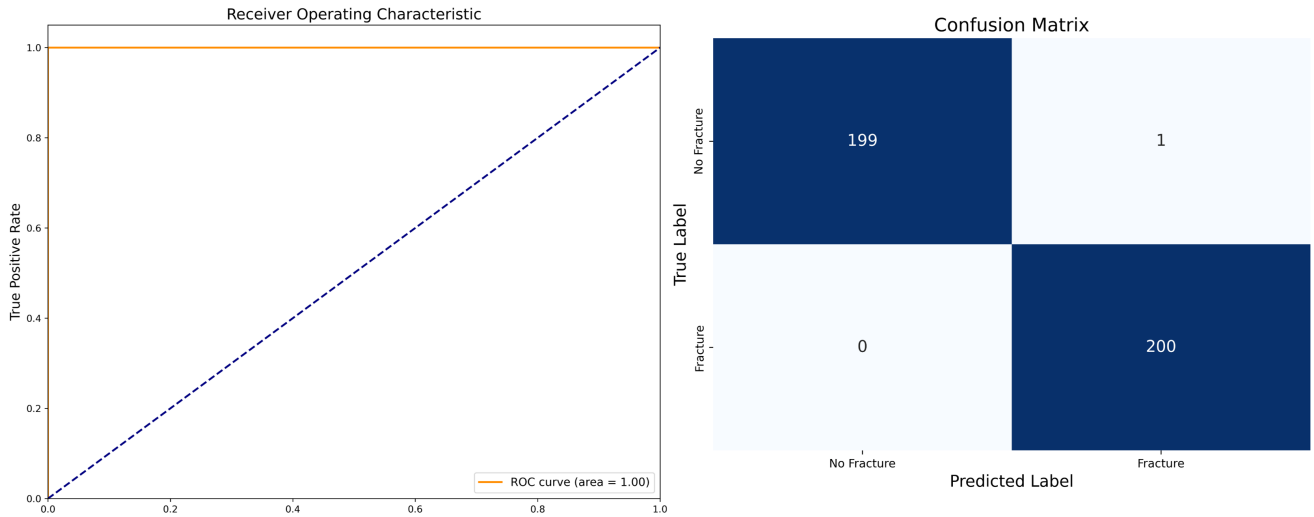


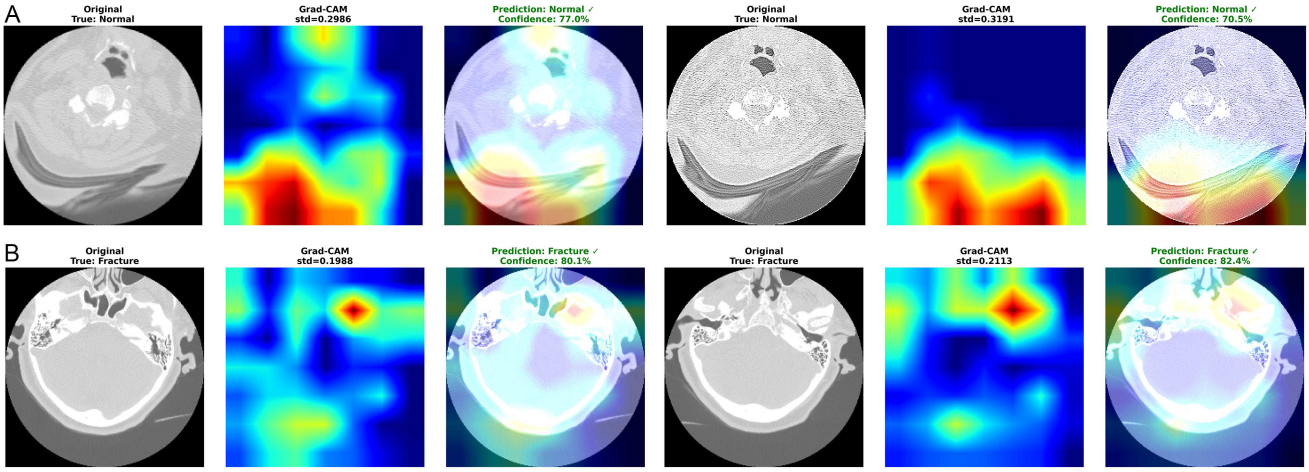
**Table 5**
**Summary of comparisons of results with related studies**

| Ref. | Architecture | Accuracy | Precision | AUC | Recall |
|---|---|---|---|---|---|
| [34] | ConvNext | 0.827 | 0.827 | 0.905 | 0.827 |
| [35] | EfficientNet | 0.793 | 0.851 | 0.9120 | 0.793 |
| [36] | ResNet50 | 0.940 | 0.946 | 0.988 | 0.940 |
| This study | MSA-CSpineNet | 0.997 | 0.995 | 0.999 | 0.999 |

**Table 6**
**Summary of comparisons of results with related studies**

| Ref. | Dataset | Architecture | Accuracy | Precision | AUC | Recall |
|---|---|---|---|---|---|---|
| [3] | RSNA C-Spine derived 2D CT slices | MobileNetV2 | 0.950 | 0.9551 | 0.9499 | 0.950 |
| [10] | Chest Pelvis CSpineScans (X-ray) | GoogleNet | 0.985 | 0.9850 | 0.9987 | 0.985 |
| [9] | Private dataset | LeNet and AlexNet | 0.950 | 0.950 | | 0.95 |
| MSA-CSpineNet | RSNA C-Spine derived 2D CT slices | MobileNet+MSA | 0.997 | 0.995 | 0.999 | 0.999 |

**Figure 4**
**The model explainability using Grad-CAM for both fractured and normal cases**




demonstrating the model's attention patterns across different prediction scenarios.

As observed in Figure 4, row A exhibits more distributed attention patterns, suggesting the model verifies the absence of fracture indicators across broader regions rather than focusing on specific abnormalities. Row B showed moderate attention concentration, with heatmap intensity focused on specific anatomical regions. In displaced fractures visible in Figure 4B, attention localizes to areas of bone discontinuity and vertebral body disruption.

C-spine fractures often result from severe trauma, causing considerable pain, limited neck movement, and possible neurological deficits. Hence, early and efficient detection of C-spine fractures is essential to prevent severe damage or adverse effects on the body. Over the years, deep learning methods have been essential in detecting critical medical conditions. Therefore, this research employed a deep learning approach to detect C-spine fractures in CT scans by integrating multi-scale spatial attention for accurate C-spine fracture diagnosis. The results of this research are promising, with an impressive accuracy of 99.75%. Compared with the existing state-of-the-art approaches that used baseline models, including GoogleNet [10] and MobileNetV2 [3] with robust custom layers, there is a significant improvement. The use of multi-scale methods enhanced the models' ability to better observe and analyze images, improving overall performance. This affirms that the information observed in an image at different scales varies significantly [37]. Unlike the commonly used pyramid approach to obtain multi-scale features by upsampling and downsampling, this research used different convolution kernels to reduce the computational complexity of the model. In addition, spatial attention helped the model focus on determining "where" important features are located within the scans.

One of the limitations of this research is the limited dataset; in future research, more datasets will be acquired to evaluate and improve the generalizability of the model. The results suggest that the multi-scale spatial attention module, combined with the MobileNet architecture, is highly effective for C-spine fracture detection. The model can reliably detect fractures in real-world applications. However, further validation on larger and more diverse datasets is necessary to ensure robustness across different clinical settings. The developed framework offers a

segmentation-free approach that eliminates the need for handcrafted features, simplifying the process and potentially improving generalization. Additionally, the model's ability to focus on a critical anatomical area without being explicitly programmed demonstrates its potential as a valuable tool in clinical practice. It could serve as an effective aid in drawing attention to regions requiring careful examination, potentially enhancing the efficiency and accuracy of radiological assessments. However, it's important to note that this analysis is based on a limited sample. Future work should involve a larger dataset to confirm the consistency of these findings across a diverse range of cases.

## 5. Conclusions

A C-spine fracture is a medical emergency that can result in permanent paralysis or even death. Hence, accurate and prompt diagnosis of suspected C-spine injuries is crucial for effective patient management. This study employs a deep learning approach that integrates multi-scale spatial attention for accurate C-spine fracture diagnosis in CT scans. The developed system achieved a performance with an accuracy of 99.75%, sensitivity of 99.99%, specificity of 99.50%, and precision of 99.50%. These results showcase the potential of attention mechanisms in improving the detection of subtle and variable fracture patterns. This demonstrates the model's strong ability for clinical application, offering a more reliable fracture detection tool. The Grad-CAM analysis provides evidence that our model develops interpretable attention patterns consistent with fracture detection. However, the study's limitations, including dataset size, warrant further research using larger, multicenter datasets to enhance generalizability. This work marks a significant step forward in AI-driven diagnostic advancements in medical imaging. In future research, other types of C-spine diseases such as cervical disc herniation, cervical spondylosis, and cervical spinal stenosis, will be included. These conditions are prevalent and share overlapping symptoms with C-spine fractures, making their differentiation crucial in clinical settings.

## Ethical Statement

The authors declare that this study did not require formal ethical approval because Atlantic Technological University does not require Institutional Review Board or ethics committee approval for retrospective research. This exemption is based on the Use of Animals for Research and Teaching Policy (4.1.1 Principle of Replacement), issued by the Quality Assurance and Enhancement Team.

## Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

## Data Availability Statement

The data that support the findings of this study are openly available in Kaggle at https://www.kaggle.com/datasets/vuppalaadithyasairam/spine-fracture-prediction-from-xrays, reference number [28].

## Author Contribution Statement

**Oladosu Oyebisi Oladimeji:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Ayodeji Olusegun Ibitoye:** Investigation, Writing – original draft, Writing – review & editing, Supervision, Project administration.

## References

[1] Karno, A. S. B., Hastomo, W., Surawan, T., Lamandasa, S. R., Usuli, S., Kapuy, H. R., & Digdoyo, A. (2023). Classification of cervical spine fractures using 8 variants EfficientNet with transfer learning. *International Journal of Electrical and Computer Engineering*, *13*(6), 7065–7077. http://doi.org/10.11591/ijece.v13i6.pp7065-7077

[2] Utheim, N. C., Helseth, E., Stroem, M., Rydning, P., Mejlænder-Evjensvold, M., Glott, T., . . . , & Linnerud, H. (2022). Epidemiology of traumatic cervical spinal fractures in a general Norwegian population. *Injury Epidemiology*, *9*(1), 10. https://doi.org/10.1186/s40621-022-00374-w

[3] Paul, S. G., Saha, A., & Assaduzzaman, M. (2023). A real-time deep learning approach for classifying cervical spine fractures. *Healthcare Analytics*, *4*, 1–14. https://doi.org/10.1016/j.health.2023.100265

[4] Izzo, R., Popolizio, T., Balzano, R. F., Pennelli, A. M., Simeone, A., & Muto, M. (2019). Imaging of cervical spine traumas. *European Journal of Radiology*, *117*, 75–88. https://doi.org/10.1016/j.ejrad.2019.05.007

[5] van den Wittenboer, G. J., van der Kolk, B. Y., Nijholt, I. M., Langius-Wiffen, E., van Dijk, R. A., van Hasselt, B. A., . . . , & Boomsma, M. F. (2024). Diagnostic accuracy of an artificial intelligence algorithm versus radiologists for fracture detection on cervical spine CT. *European Radiology*, *34*(8), 5041–5048. https://doi.org/10.1007/s00330-023-10559-6

[6] Gupta, R., Siroya, H. L., Bhat, D. I., Shukla, D. P., Pruthi, N., & Devi, B. I. (2022). Vertebral artery dissection in acute cervical spine trauma. *Journal of Craniovertebral Junction and Spine*, *13*(1), 27–37. https://doi.org/10.4103/jcvjs.jcvjs_3_22

[7] Khanpara, S., Ruiz-Pardo, D., Spence, S. C., West, O. C., & Riascos, R. (2020). Incidence of cervical spine fractures on CT: A study in a large level I trauma center. *Emergency Radiology*, *27*(1), 1–8. https://doi.org/10.1007/s10140-019-01717-9

[8] Quinn, L., Tryposkiadis, K., Deeks, J., De Vet, H. C., Mallett, S., Mokkink, L. B., . . . , & Sitch, A. (2023). Interobserver variability studies in diagnostic imaging: A methodological systematic review. *The British Journal of Radiology*, *96*(1148), 20220972. https://doi.org/10.1259/bjr.20220972

[9] Bezabh, Y. A., Salau, A. O., Abuhayi, B. M., & Ayalew, A. M. (2024). Classification of cervical spine disease using convolutional neural network. *Multimedia Tools and Applications*, *83*(41), 88963–88979. https://doi.org/10.1007/s11042-024-18970-x

[10] Naguib, S. M., Hamza, H. M., Hosny, K. M., Saleh, M. K., & Kassem, M. A. (2023). Classification of cervical spine fracture and dislocation using refined pre-trained deep model and saliency map. *Diagnostics*, *13*(7), 1273. https://doi.org/10.3390/diagnostics13071273

[11] Bruggink, C., van de Ree, C. L. P., van Ditshuizen, J., Polinder-Bos, H. A., Oner, F. C., Reijman, M., & Rutges, J. P. H. J. (2024). Increased incidence of traumatic spinal injury in patients aged 65 years and older in the Netherlands. *European Spine Journal*, *33*(10), 3677–3684. https://doi.org/10.1007/s00586-024-08310-w

[12] Lin, H. M., Colak, E., Richards, T., Kitamura, F. C., Prevedello, L. M., Talbott, J., Annotators, RSNA-ASSR-ASNR., Contributors, the Dataset Curation., . . . , Contributors, the Dataset Curation. (2023). The RSNA cervical spine fracture CT dataset. *Radiology: Artificial Intelligence*, *5*(5), e230034. https://doi.org/10.1148/ryai.230034

[13] Golla, A. K., Lorenz, C., Buerger, C., Lossau, T., Klinder, T., Mutze, S., . . . , & Goelz, L. (2023). Cervical spine fracture detection in computed tomography using convolutional neural networks. *Physics in Medicine & Biology*, *68*(11), 115010. https://doi.org/10.1088/1361-6560/acd48b

[14] Zhang, H., & Qie, Y. (2023). Applying deep learning to medical imaging: A review. *Applied Sciences*, *13*(18), 10521. https://doi.org/10.3390/app131810521

[15] Oladimeji, O., Ayaz, H., McLoughlin, I., & Unnikrishnan, S. (2024). Optimizing BI-RADS 4 lesion assessment using lightweight convolutional neural network with CBAM in contrast enhanced mammography. In R. M. Mann, T. Zhang, T. Tan, L. Han, D. Truhn, S. Li, . . . , & .. (Eds.), *Deep breast workshop on AI and imaging for diagnostic and treatment challenges in breast care* (pp. 96–106). Springer. https://doi.org/10.1007/978-3-031-77789-9_10

[16] Kalmet, P. H., Sanduleanu, S., Primakov, S., Wu, G., Jochems, A., Refaee, T., . . . , & Poeze, M. (2020). Deep learning in fracture detection: A narrative review. *Acta Orthopaedica*, *91*(2), 215–220. https://doi.org/10.1080/17453674.2019.1711323

[17] Meadi, M. N., & Benbrahim, H. (2024). Cervical spine fracture detection using deep learning algorithm. In *International Conference on Image and Signal Processing and their Applications*, 1–8. https://doi.org/10.1109/ISPA59904.2024.10536832

[18] Tanwar, V., Anand, V., Chauhan, R., & Rawat, D. (2023). Improving patient care through the use of a highly accurate CNN model for the detection of cervical spine fractures. In *International Conference on Smart Generation Computing, Communication and Networking*, 1–5. https://doi.org/10.1109/SMARTGENCON60755.2023.10442236

[19] Chłąd, P., & Ogiela, M. R. (2023). Deep learning and cloud-based computation for cervical spine fracture detection system. *Electronics*, *12*(9), 2056. https://doi.org/10.3390/electronics12092056

[20] Kutbi, M. (2024). Artificial intelligence-based applications for bone fracture detection using medical images: A systematic review. *Diagnostics*, *14*(17), 1879. https://doi.org/10.3390/diagnostics14171879

[21] Joshi, D., & Singh, T. P. (2020). A survey of fracture detection techniques in bone X-ray images. *Artificial Intelligence Review*, *53*(6), 4475–4517. https://doi.org/10.1007/s10462-019-09799-0

[22] Oladimeji, O. O., & Ibitoye, A. O. (2026). A novel attention-enhanced hybrid deep learning approach for malaria diagnosis in microscopic cell images. *Informatics and Health*, *3*(1), 41–47. https://doi.org/10.1016/j.infoh.2025.11.004

[23] Park, J., Yang, J., Park, S., & Kim, J. (2022). Deep learning-based approaches for classifying foraminal stenosis using cervical spine radiographs. *Electronics*, *12*(1), 195. https://doi.org/10.3390/electronics12010195

[24] Yamamoto, N., Rahman, R., Yagi, N., Hayashi, K., Maruo, A., Muratsu, H., & Kobashi, S. (2020). An automated fracture detection from pelvic CT images with 3-D convolutional neural networks. In *International Symposium on Community-centric Systems*, 1–6. https://doi.org/10.1109/CcS49175.2020.9231453

[25] Rashid, T., Zia, M. S., Meraj, T., Rauf, H. T., & Kadry, S. (2023). A minority class balanced approach using the DCNN-LSTM method to detect human wrist fracture. *Life*, *13*(1), 133. https://doi.org/10.3390/life13010133

[26] Gaikwad, D. P., Sejal, A., Bagade, S., Ghodekar, N., & Labade, S. (2024). Identification of cervical spine fracture using deep learning. *Australian Journal of Multi-Disciplinary Engineering*, *20*(1), 48–56. https://doi.org/10.1080/14488388.2024.2307082

[27] Small, J. E., Osler, P., Paul, A. B., & Kunst, M. (2021). CT cervical spine fracture detection using a convolutional neural network. *American Journal of Neuroradiology*, *42*(7), 1341–1347. https://doi.org/10.3174/ajnr.A7094

[28] Sairam, V. A. (2022). *Spine fracture prediction from C.T. [data set]. Kaggle*. https://www.kaggle.com/datasets/vuppalaadithyasairam/spine-fracture-prediction-from-xrays

[29] Zhu, W., Qiu, P., Chen, X., Li, H., Wang, H., Lepore, N., . . . , & Wang, Y. (2023). Beyond MobileNet: An improved MobileNet for retinal diseases. In B. Sheng, H. Chen, & T. Y. Wong (Eds.), *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 56–65). Springer. https://doi.org/10.1007/978-3-031-54857-4_5

[30] Toğaçar, M., Cömert, Z., & Ergen, B. (2021). Intelligent skin cancer detection applying autoencoder, MobileNetV2 and spiking neural networks. *Chaos, Solitons & Fractals*, *144*, 110714. https://doi.org/10.1016/j.chaos.2021.110714

[31] Chen, L., Yao, H., Fu, J., & Ng, C. T. (2023). The classification and localization of crack using lightweight convolutional neural network with CBAM. *Engineering Structures*, *275*, 115291. https://doi.org/10.1016/j.engstruct.2022.115291

[32] Wang, A., Zhao, S., Xie, K., Wen, C., Tian, H. L., He, J. B., & Zhang, W. (2024). Attention mechanism-enhanced graph convolutional neural network for unbalanced lithology identification. *Scientific Reports, 14*(1), 17319. https://doi.org/10.1038/s41598-024-64871-2

[33] Oladimeji, O. O., & Ibitoye, A. O. (2025). Multi-scale adaptive attention framework for improved lung disease classification. *Sakarya University Journal of Computer and Information Sciences*, *8*(3), 400–409. https://doi.org/10.35377/saucis...1635644

[34] Liu, Z., Mao, H., Wu, C., Feichtenhofer, C., Darrell, T., & Xie, S. A. (2022). ConvNet for the 2020s. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11966–11976. https://doi.org/10.1109/CVPR52688.2022.01167

[35] Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, 6105–6114.

[36] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. https://doi.org/10.1109/CVPR.2016.90

[37] Cao, P., Xie, F., Zhang, S., Zhang, Z., & Zhang, J. (2022). MSANet: Multi-scale attention networks for image classification. *Multimedia Tools and Applications*, *81*(24), 34325–34344. https://doi.org/10.1007/s11042-022-12792-5

**How to Cite:** Oladimeji, O. O., & Ibitoye, A. O. (2026). MSA-CSpineNet: A Multi-Scale Spatial Attention Deep Learning Framework for Cervical Spine Fracture Diagnosis. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA62027209