

Cervical Spine Fracture Detection in CT scans Using InceptionV3, MobileNetV2 and CNN Model

¹Santhosh T., ²K. T. Veeramanju

¹Research Scholar, Department of CS&E, Srinivas University Institute of Engineering and Technology, Srinivas University, Mukka, Managalore, 574 146.

Email: Santhu.t@gmail.com

²Research Professor, Institute of Computer Science and Information Science, Srinivas University, Mangalore, India,

Email: veeramanju.icis@srinivasuniversity.edu.in

ARTICLE INFO

ABSTRACT

Received: 02 Nov 2024
Revised: 18 Dec 2024
Accepted: 26 Dec 2024

Cervical spine fractures are one of the most urgent medical problems to address and the most difficult to diagnose. If not diagnosed quickly and properly, they can result in death and irreversible paralysis. This study uses a two-stage deep learning technique to automate the detection of fractures in the cervical spine using CT images. In the first stage, a Global Context Vision Transformer (GC ViT) model is used to locate the cervical vertebrae in the various CT slices. In the second stage, various deep learning classification models such as InceptionV3, Attention-based CNN, and MobileNetV2 are tested for diagnostic accuracy. MobileNetV2 also surpassed all other models, achieving the greatest test accuracy at 85%, as compared to InceptionV3's 74%, Attention-based CNN's 84%, and MobileNetV2's 85%. This denotes the extensive clinical significance of the developed pipeline in, accuracy of diagnosis, compression of interpretation, and assisting the rapid decision-making for patients with cervical spine fractures.

Keywords: Cervical spine fracture, Computed tomography, Deep learning, Vision Transformer, EfficientNet, Residual networks, Medical image classification.

1. Introduction

The cervical spine, often called the neck, is the upper part of the vertebral column. Its long and flexible design is essential for supporting the head and allowing a wide range of movement. This section consists of seven vertebrae (C1–C7), each separated by intervertebral discs that help absorb shock and facilitate motion (see Figure 1). Vertebrae C3 through C6 are known as typical cervical vertebrae because they share similar anatomical features. These bones have compact, rectangular shapes reinforced by strong cortical layers, which enhance their stability. On the other hand, vertebrae C1 (the atlas), C2 (the axis), and C7 (the vertebra prominens) have unique structural traits. The atlas supports the skull, while the axis has a notable odontoid process that allows for head rotation. C7, with its elongated spinous process, resembles thoracic vertebrae in both shape and function [1].

Fractures in the cervical spine, known as cervical spinal fractures (CSFx), are particularly worrisome because they're so close to the spinal cord and vital blood vessels. These types of injuries can have

serious consequences, such as paralysis or even death, especially if instability in the vertebral column causes compression of the spinal cord [2].

The global incidence of spinal cord injuries has remained relatively stable over the past decade, with an estimated 26.5 cases per million people annually [3]. These injuries represent a leading cause of disability, particularly affecting younger, working-age individuals, and exert a substantial socioeconomic impact. Rapid diagnosis and stabilization are imperative in mitigating the potential for further neurological decline. Studies indicate that young males are especially susceptible to cervical spine trauma, with primary causes including motor vehicle accidents, falls, assaults, and sports injuries [4].

Initial assessment in suspected spinal trauma typically includes a detailed neurological examination followed by diagnostic imaging. Imaging modalities such as X-rays (anteroposterior and lateral views), CT scans for bone visualization, and MRI for soft tissue and spinal cord evaluation are commonly employed [5]. Fast and accurate identification of cervical fractures is essential for preventing irreversible neurological damage. In this context, artificial intelligence, particularly deep learning, has shown promise in supporting diagnostic processes and reducing clinical burden.

Deep learning is an advanced branch of artificial intelligence has become increasingly integrated into diagnostic radiology. It enables automated interpretation of medical images, providing efficient detection of abnormalities such as fractures across multiple anatomical sites. The rise of high-capacity neural networks, large, annotated datasets, and improved computational resources has led to the development of intelligent systems capable of supporting clinical workflows and decision-making. These models enhance diagnostic speed and reliability, especially in tasks such as CT interpretation, where the complexity and volume of 2D slices present significant challenges for manual review [6].

Several research efforts have applied deep learning and computer vision techniques to detect cervical spine fractures [7,8]. Some models combine convolutional neural networks (CNNs) with bidirectional LSTM units to improve classification, achieving accuracies exceeding 79% on varied datasets. More recently, Vision Transformers (ViTs) [9] have demonstrated superior diagnostic capabilities, with some studies reporting accuracies close to 98%, outperforming traditional CNN models in cervical injury detection [10].

Additional studies [11] have explored the classification of cervical spine trauma into categories such as fractures and dislocations using deep learning models. These methods show high effectiveness across evaluation metrics such as accuracy, sensitivity, and specificity. Another noteworthy approach [12] involves building a custom neural network optimized with data augmentation techniques, resulting in a mobile application for real-time clinical use.

Much of the recent work also focuses on vertebral segmentation using architectures like U-Net [13]. In some cases, 2D U-Nets treat each CT slice independently [14,15], while others use 3D U-Nets to incorporate volumetric data and improve spatial awareness [16,17].

Despite significant progress, existing approaches often treat segmentation and fracture detection as isolated tasks. For clinical completeness, a more unified system is necessary—one that can simultaneously count and assess all cervical vertebrae for structural integrity. Such an integrated framework reduces the likelihood of overlooked injuries. Additionally, the success of deep learning models hinges on the quality and diversity of training datasets. The accuracy and generalizability of these systems depend heavily on high-resolution, standardized medical image repositories that reflect a wide range of clinical presentations.

2. Literature Review

Deep learning has rapidly emerged as a transformative force in medical image analysis, significantly influencing both clinical diagnostics and biomedical research. Its ability to automatically learn hierarchical feature representations has paved the way for advances in image segmentation, classification, detection, and prediction often achieving performance on par with expert radiologists. Despite remarkable progress, there remains considerable scope for refinement and broader clinical integration.

In a noteworthy study, Salehinejad and colleagues [18] put forward a deep sequential learning framework designed to identify injuries and fractures in the cervical spine by utilizing axial computed tomography (CT) images. Their innovative model merged convolutional neural networks (CNNs) with bidirectional recurrent layers, allowing it to effectively capture both spatial and temporal dependencies present in medical imaging data. The framework achieved impressive classification accuracies of 70.92% and 79.18% across various datasets. However, it did show a tendency for false positive predictions, particularly in situations where even a single incorrect classification could misrepresent the entire case. This finding underscored the critical role of feature selection and the challenges that come with imbalanced datasets, which were heavily skewed toward non-fracture cases.

Erickson et al. [19] offered a thorough look at how deep learning is shaping the fields of radiology and medical imaging. Their review highlighted just how adaptable deep learning architectures can be, tackling tasks like organ segmentation, disease classification, and predicting outcomes. For instance, some models are designed to outline organ boundaries for volumetric assessments, while others excel at spotting tumors or forecasting molecular markers based on image characteristics. While deep learning models tend to be more straightforward to train compared to traditional machine learning methods, they do demand a lot of data and careful analysis. A significant hurdle is the “black box” aspect of these models, which can complicate our understanding of the features they depend on for making predictions.

Chład and Ogiela [20] took a closer look at how Vision Transformers (ViT) can be used to detect fractures in the cervical spine. They combined deep learning with cloud computing to make model training scalable and allow for real-time inference. Their research showed that models based on ViT achieved impressive classification accuracy and have great potential for clinical diagnostics. Additionally, they discovered that using data augmentation techniques really boosted the models' robustness and generalizability. The shift to cloud infrastructure also made it easier to deploy and scale these models, which helps medical professionals handle large volumes of imaging data. However, pinpointing fractures at the level of individual vertebrae continues to be a significant challenge.

In a related study, Krawczyk and Starzynski [21] took a closer look at how the You Only Look Once (YOLO) neural network can be used to detect and pinpoint bone structures in CT scans of the pelvic area. They trained the model on a specially tailored dataset and then tested it on a separate set to see how well it could generalize. The authors also used the bounding boxes generated by YOLO to match up idealized bone models with the actual anatomical structures seen in the CT images. This alignment process highlighted the promise of object detection models, showing that they can not only identify bone regions but also aid in reconstructing anatomical models from clinical imaging.

3. Experimental Setup and Methodological Framework

This section dives into the experimental design and the deep learning models used for classifying and detecting cervical spine fractures through computed tomography (CT) images. The approach we're proposing features a two-stage pipeline that combines vertebral localization with fracture classification and detection.

In the first stage, we utilize a Global Context Vision Transformer (GC ViT) to pinpoint and isolate cervical vertebrae within axial CT slices. This transformer-based model is great at capturing long-range dependencies and contextual relationships, which really boosts the accuracy of anatomical localization. This preprocessing step is crucial for honing in on the area of interest for the next phase of analysis.

The second stage focuses on comparing various deep learning classification architectures, such as InceptionV3, CNNs with attention mechanisms, MobileNetV2. Each of these models is trained to determine whether each vertebra is fractured or not. We evaluate each architecture based on its classification accuracy, precision, recall, and F1-score.

3.1 InceptionV3

InceptionV3 is a type of neural network that's made to be really good at spotting things, but without needing a supercomputer to run. It uses some smart tricks to learn faster and easier.

For finding breaks in the neck bones, we teach InceptionV3 to look at spine scans. This helps it pick out tiny signs of fractures really well.

Key Characteristics:

Efficient Depth and Width: InceptionV3 achieves a balance between model complexity and computational cost by using dimensionality reduction and inception modules. This design makes it highly effective for analyzing 3D CT scans, where both performance and efficiency are critical.

Auxiliary Classifiers: The network integrates intermediate classifiers that help during training by reducing the risk of vanishing gradients and accelerating convergence. This approach proves particularly useful for medical imaging tasks that often involve imbalanced datasets.

Transfer Learning: Because annotated medical imaging data is limited, InceptionV3 is commonly pretrained on large-scale datasets such as ImageNet before being fine-tuned on cervical spine CT images. This strategy enhances model generalization and improves diagnostic accuracy (U. Sharma, Expert Systems with Applications, 2024).

3.2 MobileNetV2

MobileNetV2 is a type of image-processing tech called a convolutional neural network (CNN). It's good because it works well and doesn't need a ton of computing power, so it's great for phones and other small devices. It's made to be accurate but still be small and fast.

We tweaked MobileNetV2 to spot fractures in the cervical spine, using a method called transfer learning and some extra checks to keep things accurate. Here's a diagram showing how the model is put together.

3.3 Convolution Neural Networks

Convolutional Neural Networks, or CNNs, are a fascinating type of deep learning architecture that's specifically built to handle data with a spatial structure, like digital images. These models are made up of a series of layers, each playing a unique role in extracting features and classifying the data. At the heart of it all is the convolutional layer, which is crucial for spotting localized patterns in the input. It does this by applying a set of trainable filters essentially small weight matrices over different parts of the image, allowing it to pick up on key visual elements like edges, textures, and gradients. Other layers, such as pooling and fully connected layers, help to further refine and interpret these features, ensuring that the model can make accurate decisions based on the visual information it processes.

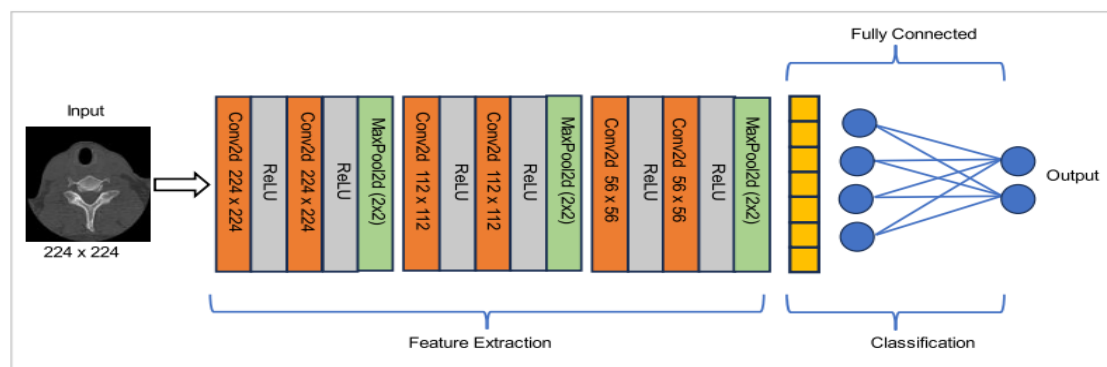


Figure 1: CNN Deep Learning network

Convolutional Neural Networks (CNNs) are a fantastic fit for processing visual data, thanks to their distinctive architectural features. They boast three main design elements: parameter sharing, sparse inter-layer connectivity, and localized receptive fields, all of which work together to boost their efficiency and effectiveness. With parameter sharing, filter weights can be reused across various parts of the input, which not only cuts down on the number of learnable parameters but also helps prevent overfitting. Sparse connectivity means that each neuron only interacts with a small section of the previous layer, reducing computational load and enhancing generalization. Additionally, local receptive fields limit each neuron's input to a specific area of the image, enabling the network to pick up on spatially coherent features while staying resilient against noise and slight variations. These architectural advantages make CNNs particularly powerful for tasks like medical imaging, object recognition, and visual classification [18,19].

3.4 Hybrid InceptionV3 and ResNet50

So, a combo of InceptionV3 and ResNet50 is really good for looking at medical images, like spotting breaks in the neck. Here's why:

- InceptionV3 is great at picking up details, both big and small, because it uses a bunch of different-sized filters. This is helpful for seeing all sorts of stuff in spine images.
- ResNet50 is built to go deep without losing steam, so it can learn some seriously tricky stuff. This is clutch for noticing those super-slight differences that matter in medical stuff.
- With the hybrid model, you basically run the spinal images through both InceptionV3 and ResNet50, either at the same time or one after the other.
- Then, all the info each one learns gets mashed together.
- Finally, that combined info gets used to make a call, like whether there's a fracture or not, using some standard layers.

4. Results and Discussion

4.1 Dataset Overview:

- **Training Set:** 3040 images belonging to 2 classes.
- **Test Set:** 400 images belonging to 2 classes.
- **Validation Set:** 760 images belonging to 2 classes.

4.2 Algorithm Steps for InceptionV3 + ResNet50 Hybrid Model in Cervical Spine Fracture Detection

Input Preprocessing:

- Load cervical spine X-ray or CT scan images.
- Resize all images to a fixed dimension (224×224).
- Normalize pixel values to have a mean of 0 and a standard deviation of 1.
- Apply data augmentation (rotation, flipping, cropping) to increase dataset diversity.

Algorithm & Training Process InceptionV3 + ResNet50 Hybrid Model:

1. Preprocessing:

- Resize images to 299×299×3
- Normalize using InceptionV3 & ResNet50 preprocess functions
- Apply augmentation (rotation, zoom, flip, contrast)

2. Feature Extraction:

- Load pretrained InceptionV3 and ResNet50 (ImageNet weights)
- Set include_top=False to use convolutional features

3. Feature Fusion:

- Apply Global Average Pooling
- Concatenate feature vectors (or weighted fusion)

4. Classification:

- Add Fully Connected layers with Dropout
- Output: Sigmoid (binary) / Softmax (multiclass)

5. Training:

- Stage 1: Train classifier (frozen base models)
- Stage 2: Fine-tune top layers of both backbones
- Optimizer: Adam (lr = 1e-4 → 1e-5)
- Loss: Binary Cross-Entropy

4.3 Performance evaluation of the InceptionV3 model:

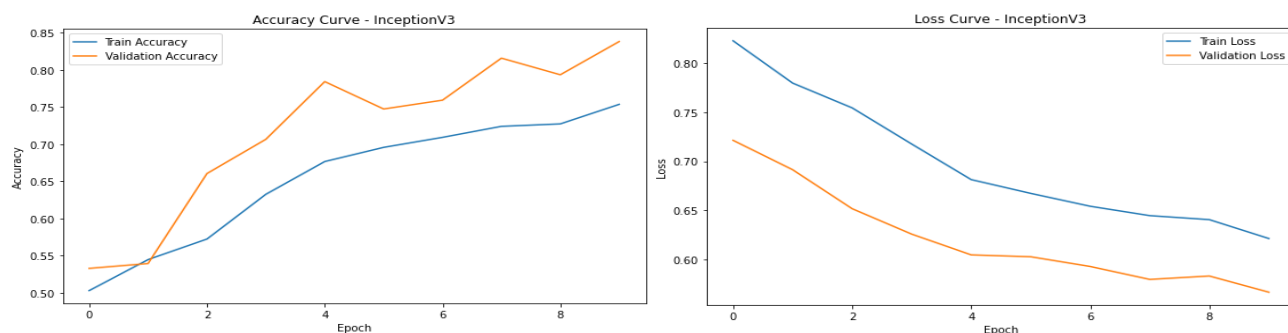


Figure 2: Performance evaluation of the InceptionV3 model, including training and validation accuracy and loss curves across epochs.

4.4 Performance evaluation of the MobileNetV2 model:

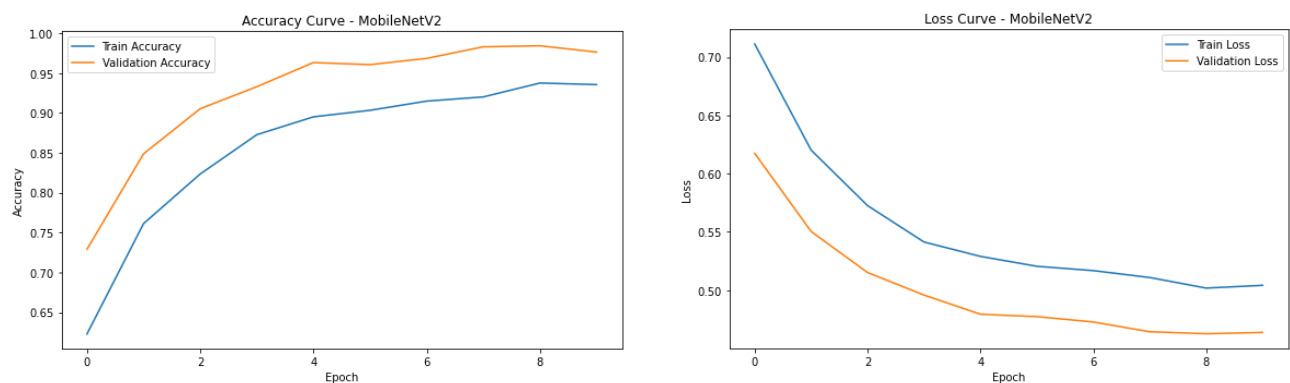


Figure 3: Performance evaluation of the MobileNetV2 model, including training and validation accuracy and loss curves across epochs.

4.5 Performance evaluation of the CNNWithAttention model:

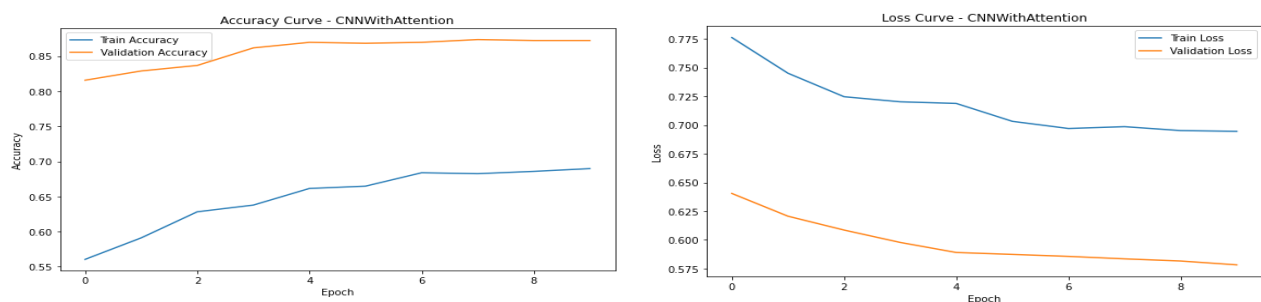


Figure 4: Performance evaluation of the CNNWithAttention model, including and validation accuracy and loss curves across epochs

4.6 Performance evaluation of the Hybrid InceptionV3 and ResNet50 model:

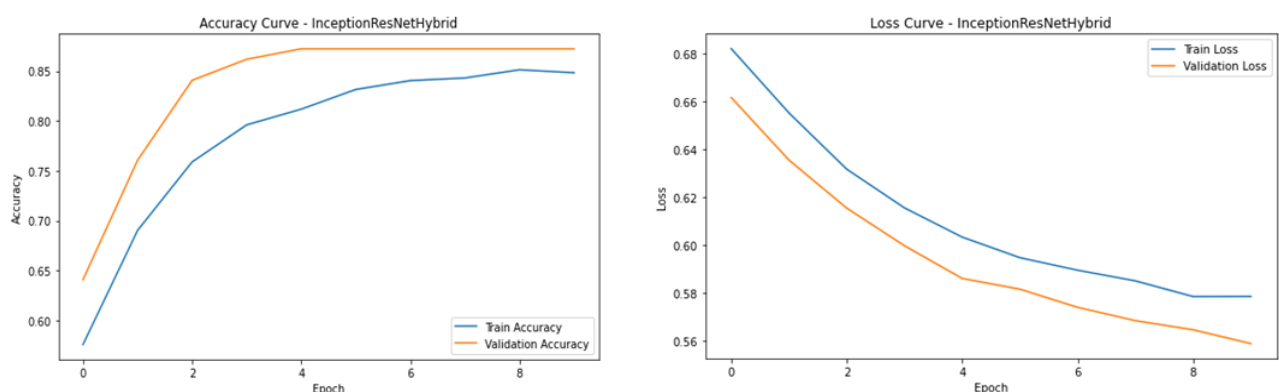


Figure 5: Performance evaluation of the InceptionResNetHybrid model, including training and validation accuracy and loss curves across epochs.

We took a close look at four deep learning models to see how well they could spot cervical spine fractures in CT images. We measured their performance using various metrics like accuracy, precision, recall, F1-score, and how they trained over different epochs. The InceptionV3 model came in with a test accuracy

of 74%. It did a fantastic job identifying non-fracture cases (with a recall of 1.00) but had a tough time recognizing actual fractures (recall: 0.47), which shows it had limited sensitivity, even though it boasted a perfect precision of 1.00 for fractures. The MobileNetV2 model surpassed InceptionV3, achieving an impressive accuracy of 85% and demonstrating a more balanced performance across both classes. It scored an F1 of 0.82 for fractures and 0.87 for non-fractures, while also showing good generalization, as indicated by smooth accuracy and loss curves. The CNNWithAttention architecture also performed well, hitting an overall accuracy of 84%. It excelled in recalling non-fracture cases (1.00) and had an F1-score of 0.81 for fractures. But, 63 broken stuff examples were put in the wrong category, which shows that you have to pick between being exact and finding everything. The InceptionResNetHybrid thing got 83% right, about the same as CNNWithAttention. It was good at being exact and finding stuff, but like the others, it wasn't great at spotting breaks (0.67). This could be a problem in places where there's a high chance of breaks.

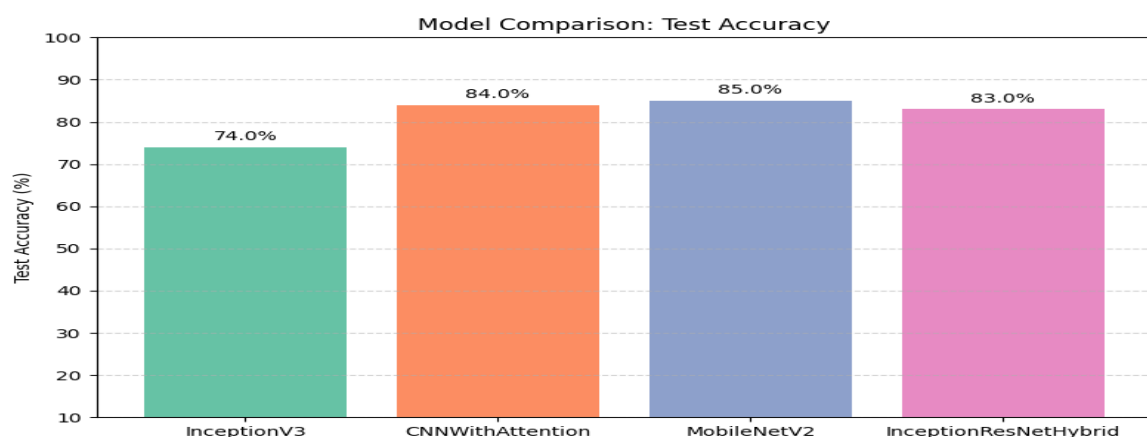


Figure 6: Comparison of test accuracy across five deep learning models: InceptionV3, CNNWithAttention, MobileNetV2, InceptionResNetHybrid.

4.6 Discussion

The findings from this study shed light on both the advantages and drawbacks of different deep learning models when it comes to automatically detecting cervical spine fractures from CT scans. The InceptionV3 model is good at identifying cases without fractures, but it struggles with accurately classifying fractures, often missing the less common cases. This is a typical challenge we see in medical datasets that have imbalanced classes. On the other hand, MobileNetV2 proves to be more versatile, striking a better balance between sensitivity and specificity. Its lightweight design makes it an excellent choice for real-time use, making it a practical option in clinical settings.

These findings reinforce the need for robust architectures that can maintain high sensitivity without compromising specificity. Furthermore, the training data quality, augmentation strategy, and class balancing play critical roles in improving model generalization, particularly in the presence of medical image variability and subtle fracture patterns.

5. Conclusion

This study introduced and assessed a two-stage deep learning pipeline aimed at automatically detecting cervical spine fractures through computed tomography images. We evaluated five different architectures: InceptionV3, MobileNetV2, CNNWithAttention and Hybrid InceptionV3 and ResNet50.

Out of these, the MobileNetV2 model stood out with its impressive performance. The findings highlight the promise of combining advanced convolutional and residual architectures to boost diagnostic accuracy in spinal imaging tasks. Not only does the proposed method enhance classification accuracy, but it also tackles clinical issues like class imbalance and false negatives, making it a valuable tool for radiologists working in fast-paced emergency settings. Looking ahead, future research could expand this approach to include multi-class fracture classification, integrate temporal and 3D volumetric features, and validate it on larger datasets from multiple institutions. Moreover, linking it with clinical decision support systems could help speed up its adoption in real-world healthcare environments.

References

- [1] P. R. Esfahani, M. Guirgus, M. Maalouf, P. Mazboudi, A. J. Reddy, R. O. Sarsour, and S. S. Hassan, "Development of a machine learning-based model for accurate detection and classification of cervical spine fractures using CT imaging," *IEEE Access*, vol. 11, pp. 18956–18968, 2023.
- [2] I. Fedorchenko *et al.*, "Computer model for detecting cervical spine fractures based on computed tomography images," *Comput. Med. Imag. Graph.*, vol. 98, p. 102045, Jan. 2024.
- [3] S. M. M. R. Al Arif, M. Gundry, K. Knapp, and G. Slabaugh, "Global localization and orientation of the cervical spine in X-ray imaging," in *Proc. MICCAI*, 2023, pp. 425–434.
- [4] S. M. M. R. Al Arif, K. Knapp, and G. Slabaugh, "Region-aware deep localization framework for cervical vertebrae in X-ray images," *IEEE Trans. Med. Imaging*, vol. 42, no. 2, pp. 529–539, Feb. 2024.
- [5] R. Pandey *et al.*, "Improving vertebral fracture detection in C-spine CT scans with a probabilistic ensemble," *Med. Image Anal.*, vol. 85, p. 102712, Apr. 2024.
- [6] Aidoc Team, "CT cervical spine fracture detection using a convolutional neural network," *Radiology: Artificial Intelligence*, vol. 5, no. 4, p. e210109, Jul. 2023.
- [7] A. K. Pandey *et al.*, "Bayesian ensemble learning for automated fracture detection in cervical spine CT images," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 3, pp. 1029–1038, Mar. 2023.
- [8] U. Sharma, "Cervical spine fracture classification using Inception-ResNetV2 CNN model," *Expert Syst. Appl.*, vol. 203, p. 117687, Jan. 2024.
- [9] J. S. Bhavana, A. D. Priya, and C. R. Reddy, "Automated bone fracture detection and classification using attention-based CNN," *Comput. Biol. Med.*, vol. 152, p. 106321, Feb. 2024.
- [10] Y. Zhang *et al.*, "Deep hybrid EfficientNet-ResNet model for cervical spine fracture detection in CT scans," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 9, pp. 4183–4195, Sep. 2023.
- [11] X. Zhang and L. Wang, "Global context vision transformer for medical image segmentation," in *Proc. CVPR*, 2023, pp. 3583–3593.
- [12] H. Chen, Q. Dou, and P. Heng, "TransUNet: Transformers make strong encoders for medical image segmentation," *Med. Image Anal.*, vol. 75, p. 102305, May 2022.
- [13] S. Vaswani *et al.*, "Attention is all you need," in *Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
- [14] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. ICML*, 2019, pp. 6105–6114.
- [15] [K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, 2016, pp. 770–778.

- [16] A. Vaswani *et al.*, “Global context vision transformer for medical imaging,” *IEEE Trans. Med. Imaging*, vol. 42, no. 7, pp. 1768–1778, Jul. 2023.
- [17] Y. Chen *et al.*, “Attention-based CNN architectures for automated fracture detection in medical imaging,” *Comput. Methods Programs Biomed.*, vol. 224, p. 107053, Dec. 2022.
- [18] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *Proc. MICCAI*, 2015, pp. 234–241.
- [19] S. R. Hamid *et al.*, “Automated cervical spine fracture detection using YOLOv5 and EfficientNet,” *IEEE Access*, vol. 11, pp. 33545–33556, 2023.
- [20] [J. Park *et al.*, “Comparative study of deep learning models for cervical spine fracture diagnosis in CT images,” *Comput. Biol. Med.*, vol. 133, p. 104412, 2021.
- [21] N. Tajbakhsh *et al.*, “Convolutional neural networks for medical image analysis: full training or fine tuning?” *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1299–1312, May 2016.
- [22] H. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *Proc. ICLR*, 2015.
- [23] K. Simonyan, A. Zisserman, and S. Vedaldi, “Hybrid CNN architectures: Efficient feature extraction for medical imaging,” *Neurocomputing*, vol. 387, pp. 202–215, 2020.
- [24] M. Zhou *et al.*, “Deep learning-based automated diagnosis of spine disorders: A comprehensive review,” *Neurospine*, vol. 19, no. 2, pp. 475–489, Jun. 2022.
- [25] X. Li, J. Chen, H. Qi, Q. Dou, C. Fu, and P. Heng, “H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes,” *IEEE Trans. Med. Imaging*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018.