International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Comparative Analysis of Neural Network Architectures for Automated Fracture Detection in Hand X-Ray Images

Jie Zhnag[1], Hongzhen Chen[2]

[1]Assumption University

Abstract

The application of several neural network architectures—including Fully Connected Networks (FCN), Convolutional Neural Networks (CNN), pretrained ResNet, Vision Transformer (ViT-B-16)—for the classification of hand X-ray images into "Fractured" and "Not Fractured"—categories is investigated in this work. The main goals are to evaluate these models' fracture detection ability and determine which architectural design fits this work. Because transfer learning let the model use past information from big-scale picture datasets, the pretrained ResNet model emerged as the most effective with high accuracy, stability, and resilience. The bespoke CNN also performed well, displaying excellent feature extraction powers especially for medical imaging. But the non-pretrained ResNet model overfitted, meaning deeper networks find it difficult to generalize without pretraining. Though innovative, the Vision Transformer performed poorly since it depends on a lot of training data and finds difficult learning of intricate spatial properties from little datasets. Although acting as a baseline, the FCN's simple architecture and incapacity to detect spatial hierarchies in images meant it could not match the efficacy of CNN models. Emphasizing the important function of transfer learning in clinical applications, the results show that pretrained CNN architectures, especially ResNet, offer the most consistent and accurate method for automatic fracture diagnosis in medical pictures.

1. Introduction

Among the most often occurring medical disorders compromising the human body are fractures. Among men and women globally, the frequency of fractures is really significant; moreover, age is also associated to this tendency. Common forms of fractures are metacarpal and phalangeal ones, usually resulting from direct external stresses like heavy physical exercise. These fractures range in kind: distal phalanx fractures, often comminuted; intermediate phalanx fractures, which also commonly displace towards the palm; and proximal phalanx fractures, which tend to displace towards the palmar side. Because of possible internal bleeding, metacarpal and phalangeal fractures usually produce pain and fast swelling at the damage site; some individuals may also develop abnormalities at the fracture point. Clinically, swift recovery depends on fast localisation of the fracture site and appropriate treatment to minimise permanent damage and minimise patient suffering.

X-ray, CT, and MRI are clinical techniques for identifying fractures; X-ray imaging is the most often utilised since it is so reasonably priced. Using its great penetrating qualities for medical diagnosis, X-ray technology generates radiographic images. Although X-ray is important for fracture detection, accurate evaluation still depends on the knowledge of seasoned doctors. The enormous volume of imaging data and

the time-consuming nature of the diagnostic process present various difficulties for current diagnostic techniques including the possibility of misdiagnosis by less experienced radiologists. Furthermore affecting diagnosis accuracy are X-ray pictures of fractures, which are sometimes partial or unclear. Unlike fractures in other areas of the body, metacarpal and phalangeal fractures include complicated joint systems and are typically numerous, which makes little, hidden fractures more challenging to find.

Deep learning has developed recently to enable the application of deep learning-based computer-aided systems for medical imaging diagnostics. Deep learning technology integration in the medical industry has produced many discoveries that greatly advance computer-aided diagnosis. Early medical artificial intelligence applications have mostly concentrated on picture recognition tasks and treatment outcome evaluation, showing significant benefits in the detection of metacarpal and phalangeal fractures. Still, the discovery of minor fractures in these regions gives chances for more research and remains difficult.

Common skeletal ailment in daily life is hand fractures. The hand consists of several linked, sophisticated, complicated bones unlike other skeletal systems like ribs and leg bones. Modern clinical diagnostics depend heavily on X-ray imaging, which provides radiologists with convenience for evaluating hand bone injuries, fast image collection, and simplicity in operation. Still, hand fractures are somewhat common and usually show up as minor, challenging-to-identify injuries. Radiologists impose a great reading load on a small number of doctors since they must carefully search every image to find all fracture sites. Furthermore, the radiologist's subjective assessment forms the basis of the diagnostic process; so, the physician's expertise and knowledge can limit this judgement and result in regular incidents of missed or inaccurate diagnosis. On the other hand, computers' continuing labour capability and objective judgement show enormous promise for X-ray picture diagnosis.

Particularly helping less experienced doctors, deep learning-based systems have greatly increased diagnostic accuracy in studies using computer-aided diagnostic systems for picture reading. Studies on diagnostic efficiency find that using deep learning-based computer-aided diagnostic systems significantly lowers the time needed for image interpretation by both trainee and attending radiologists. By means of objective diagnostic recommendations, advanced hand X-ray fracture detection algorithms improve radiologists' diagnosis confidence and speed, so minimising misdiagnosis and missed diagnosis rates and so minimising conflicts between doctors and patients and greatly lowering radiologists's workload. By preventing health losses resulting from misdiagnosis, missed diagnosis, or ineffective diagnostics, these systems help patients to improve their whole medical experience and thereby increase public well-being by means of healthcare automation. Moreover, accurate hand fracture detection systems can provide patients with self-diagnostic recommendations, which might eventually prove to be a handy regular health check.

In order to solve present diagnostic difficulties, this work creates a CNN-based model for hand fracture detection in X-ray pictures. It summarises the body of knowledge, points up important problems, describes the approach, shows findings, and ends with the efficacy of the model and future study recommendations.

2. Literature review

2.1 Computer-Aided Diagnosis of Fractures

Medical artificial intelligence (AI) technology is now extensively used in many different medical disciplines and generates a lot of curiosity among scholars all around. Medical diagnostics has seen a tsunami of AI-driven developments out of this. Deep learning-based computer-aided diagnosis systems have shown great promise in diagnosing many diseases, including breast cancer detection and recognition,

lung nodule detection and classification, colorectal polyp detection and classification, retinal disease detection, vascular segmentation, skin disease diagnosis and classification, osteoporosis prediction, prostate cancer localisation, and more.

Scholars both here and abroad have made important strides in the diagnosis and categorisation of fractures. Since fractures are systemic disorders, study on their diagnosis is done on several body areas. Key areas of research include the radius (Kim et al., 2021), scaphoid (Yoon et al., 2021), humerus (Chung et Two main applications of computer-aided diagnosis in fractures are first, determining the presence of a fracture, which falls under the domain of image classification in deep learning; second, identifying not only the presence of a fracture but also predicting its specific location, a more difficult task categorised under object detection.

Researchers like Kim et al. have investigated the use of convolutional neural networks (CNNs) for fracture diagnosis in wrist X-rays in the categorisation of fractures. Using DenseNet-161 and ResNet-152, they trained models and tested their performance on a collection of 990 wrist X-rays, obtaining test accuracies of 90.3% and 88.5%, respectively (Kim et al., 2020). Using augmented data from wrist X-rays, another paper used transfer learning with the Inception-V3 network to distinguish between "fracture" and "non-fracture".

Common in the elderly, osteoporotic fractures (OVFs) are caused by undetected osteoporosis, usually asymptomatic until a fracture results. Using ResNet and aggregates these characteristics using Long Short-Term Memory (LSTM) networks, along with three other rule-based feature aggregation techniques, Tomita et al. created an OVF detection system first extracting features from chest, abdomen, and pelvic CT images.

Dual-energy X-ray absorptiometry (DXA) is the clinical gold standard for osteoporosis diagnosis. Using a mix of InceptionResNetV2 and DenseNet, Derkatch et al. looked at whether DXA scans would show spinal fractures and compared the model's accuracy to that of orthopaedic experts (Derkatch et al., 2019). Another often occurring clinical fracture type, hip fractures, seriously affect daily life. Using an InceptionV3-based CAD system, Tanzi et al. applied a deep learning strategy to categorise femoral proximal fractures and discovered that the diagnostic accuracy increased by 14% when using CAD assistance rather by unaided specialists (Tanzi et al., 2020).

Task related to fracture detection and localisation usually call for more complicated CNN models. A Dilated Convolutional Feature Pyramid Network (DCFPN) for femur fracture detection and localising was proposed by Guan et al. with an average precision of 82.1%, therefore proving the potential of DCFPN in clinical settings (Guan et al., 2019). Developing ParalleNet, a two-stage detection network with a parallel backbone and feature fusion structure, Wang et al. obtained better performance than existing detection models (Wang et al., 2021).

Some research have also concentrated on the viability of CNN-based models in identifying concealed fractures in X-rays, including scaphoid fractures detectable on MRI but invisible on X-rays. A two-stage classification algorithm developed by Yoon et al. first trained on X-rays with apparent fractures and subsequently fine-tuned on X-rays with concealed fractures, therefore obtaining high sensitivity and specificity in identifying these difficult cases (Yoon et al., 2021).

Additionally used in segmentation and detection of rib and spinal fractures are deep learning techniques. A sensitivity of 92.9% with an average of 5.27 false positives per scan was obtained by Jin et al. using a 3D U-Net model for rib segmentation and detection (Jin et al., 2020). For the automated segmentation of

X-ray pictures of compression fractures, Cheol et al. coupled deep learning with level set approaches to show exact recognition and segmentation of vertebrae (Cheol et al., 2020).

Apart from these investigations, attempts are under way to automate fracture diagnosis in other particular domains such dental root fractures in panoramic radiographs, mandibular fractures, and calcaneal fractures in CT scans. Underlining the viability of using deep learning for fracture detection across many imaging modalities and anatomical areas, these studies frequently employ CNNs combined with feature extraction techniques like Speeded-Up Robust Features (SURF) to improve classification and localisation accuracy.

2.2 Research on Small Object Detection Algorithms

The evolution of computers and the extensive usage of the internet have resulted in ever more varied approaches of information collection. Methods including speech, video, and graphics have become indispensable tools for information communication because of their simple, graphic, and extensive traits. Computer vision has emerged to enable more simply, effectively, and intelligibly processing of vast volumes of data. Emerging as a field, computer vision lets computers interpret and evaluate visually available material, hence improving human perception and acquisition of information. Found in biological areas, video surveillance, facial identification, aerospace, among others, object detection—a fundamental topic of computer vision—finds great use (Lin et al., 2017). Object identification mostly aims to designate the areas of interest for particular objects—such as faces, cars, cats, dogs, or flowers—in given photos or videos. Object detection is more difficult than picture categorisation since it entails identifying and exactly locating several items inside an image.

Traditional object detection algorithms and deep learning-based object detection algorithms are the two primary two categories into which object detection techniques mostly fit. Usually depending on low-level and mid-level visual attributes (e.g., colour, texture), traditional object detection algorithms find objects. Setting up sliding windows of various scales to identify candidate areas across the entire image to locate objects, extracting features from these candidate regions, and subsequently using classifiers, such Support Vector Machines (SVM) (Saunders et al., 2002) and AdaBoost (Fu, 2011), to classify the regions and ascertain whether they contain objects and their category attributes forms the primary workflow. With generally used and effective features including Haar features, Scale Invariant Feature Transform (SIFT), Histogram of Orientated Gradient (HOG), and Speeded Up Robust Feature (SURF), the core of traditional object detection is the feature extraction stage. Originally employed in the 1990s for handwritten digit recognition, convolutional neural networks (CNNs) did not see much progress due to technology constraints. Along with major developments in processing capacity, Hinton's team suggested deep learning in 2006 and addressed the gradient vanishing issue during training, hence generating great interest and application for deep learning (Hinton & Salakhutdinov, 2006). AlexNet greatly raised classification accuracy in the 2012 ImageNet ILSVRC object identification competition, hence revolutionising CNNs (Krizhevsky et al., 2012). Unlike conventional methods, AlexNet After this, fresh models including GoogLeNet, ResNet (He et al., 2016), and VGGNet (Simonyan & Zisserman, 2014) were suggested progressively, each helping to shape deep learning-based detection systems.

Deep learning has brought object detection techniques into a fresh phase. Deep learning-based object identification systems greatly enhance detection performance by using convolutions to extract features from training data, therefore allowing the collection of intricate and detailed pixel distribution patterns. Two-stage and single-stage object identification algorithms are the two categories used to classify present mainstream deep learning-based object detection systems. Two-stage object detection algorithms

grounded on region proposals: Following selective search to identify potential areas from input photos, R-CNN (Region with CNN feature) was proposed in 2014 followed by CNN feature extraction, classification, and localisation (Girshick et al., 2014). The production of many candidate boxes during training kept the computational speed slow even with the increases in detection accuracy. Later models as SPPNet and Fast R-CNN included methods to combine retrieved features and simplify computations, so considerably improving training speed without compromising detection accuracy (Girshick, 2015). Faster R-CNN, which replaced selective search with Region Proposal Networks (RPN), achieved end-to- end model structure and significantly raised speed and accuracy, thereby marking a major turning point (Ren et al., 2017). Mask R-CNN later expanded this framework by including pixel-level segmentation for contour identification tasks, hence extending its uses beyond object detection to include pose estimation (He et al., 2020).

Regression-Based Single-Stage Object Detection Algorithms: Although two-stage detection systems have great accuracy, their limited detecting speed results from the phase of the region proposal. This led to the creation of single-stage detection techniques based on CNN predictions of item classes and positions straight forwardly. Notable single-stage algorithms that introduced multi-layer feature extraction and combined classification and regression in one step, so attaining accuracy and speed comparable to two-stage detectors, are the YOLO (You Only Look Once) series (Redmon et al., 2016) and SSD (Single Shot MultiBox Detector). RetinaNet addressed the precision problem by including focus loss to offset class imbalance, therefore attaining detection accuracy sometimes exceeding two-stage detectors (Lin et al., 2020).

One of the toughest problems still in object detection is small object detection. Either their absolute size—with dimensions lower than 32×32 pixels—where the length and width of the target region are less than 1/10 of the original image—or their relative size, where the dimensions describe little objects. CNN models have great difficulty with little objects because of their low resolution, fuzzy features, and complicated backdrops. Solutions include: Super-resolution techniques help to increase the resolution of small objects, therefore enhancing the detection accuracy. Low-level detailed information combined with high-level semantic information enhances small item detection by feature fusion. Widely employed for this aim are techniques include Feature Pyramid Networks (FPN), which combine multi-level feature maps across top-down paths (Lin et al., 2017). Contextual information helps to augment object knowledge and thereby improve detection accuracy. To improve detection performance, AC-CNN for example uses contextual windows around candidate areas.

These methods show the field's development towards more advanced and effective algorithms by stressing the continuous improvements and difficulties in identifying little things inside complicated images.

3. Problem statement

Timeliness and precise diagnosis depend on the identification and classification of fractures using medical images, such X-rays. Radiologists' manual review of these pictures, however, can be time-consuming and prone to human error—particularly in cases of subtle or minute fractures that are difficult to detect. This work intends to use deep learning models—especially convolutional neural networks (CNNs)—to automate the classification of hand X-ray pictures into "Fractured" and "Not Fractured" categories, therefore addressing these issues.

Because CNNs can learn and extract intricate features from visual data, they have shown rather good performance in image recognition challenges. In this work, we will investigate the features of a customized

CNN model intended especially for hand X-ray picture fracture identification. The aim is to assess CNN's performance in spotting fractures and contrast it with other models, comprising Transformer-based models, pretrained CNN models, and a simple Fully Connected Network (FCN).

As a basic model, the FCN will offer a straightforward architecture that clarifies the enhancements CNNs offer to fracture detection tasks. We will also incorporate a pretrained CNN model employing weights trained on extensive picture datasets, hence leveraging transfer learning. With few training data, this method lets the model generalize better. At last, a Vision Transformer (ViT) model will be used to evaluate how transformer-based architectures—known for their self-attention mechanisms—perform in respect to conventional CNNs.

We want to find the advantages and disadvantages of every method by methodically contrasting these models. The investigation will provide light on how various network designs might influence the classification accuracy of fracture detection in medical imaging, so providing possible routes for enhancing automated diagnostic instruments in clinical environments. This comparison will finally serve to clarify the most efficient model architecture for the purpose of hand fracture diagnosis and might help to improve the whole diagnostic procedure in medical practice.

4. Methodology

4.1 Dataset Description

Hand X-ray images—specifically meant to distinguish between "Fractured" and "Not Fractured" conditions—make up the dataset used in this work. Three subsets—training, validation, and testing—as well as two classes—"Fractured" and "Not Fractured"—are arranged in this dataset. The main objective is to assess the capacity of several deep learning models—including CNNs and Transformer-based models—to automatically identify fractures in hand X-rays, which might be quite important in clinical environments for fast and accurate diagnosis.

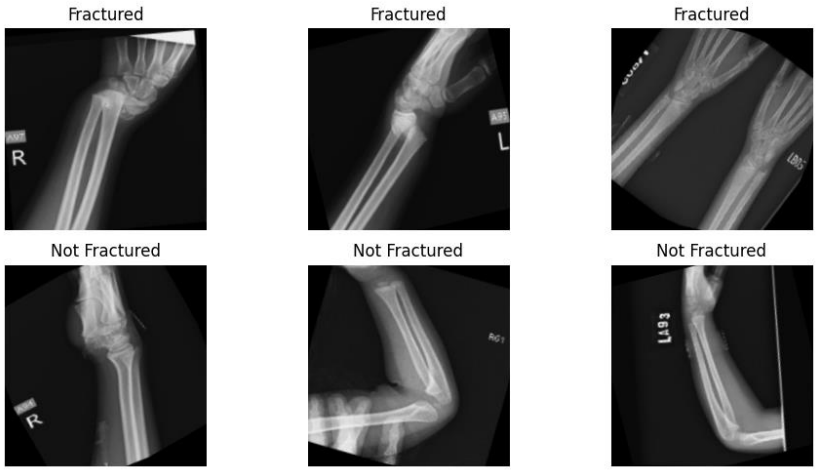

Figure 1 Sample Images of Fractured and Not Fractured Hand X-rays

The X-ray images depict various conditions of the hand, including multiple types of fractures and normal bone structures. The images are of varying quality and complexity, reflecting real-world scenarios where fractures can be subtle and difficult to detect, even for trained radiologists. This complexity makes the dataset an ideal testbed for evaluating the capabilities of advanced neural networks, such as CNNs and Transformer-based models, in automating fracture detection.

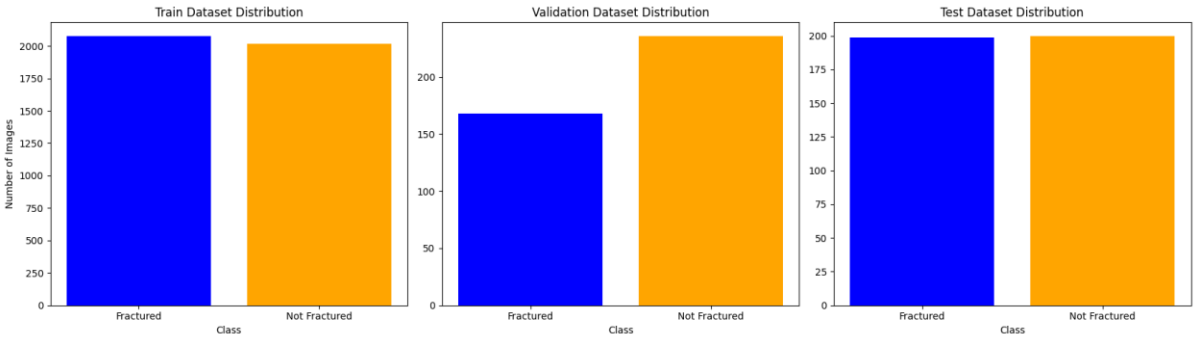

Figure 2 Sample Distribution of the Dataset

4.2 Data Preprocess

Aimed at preparing the hand X-ray pictures for efficient training of the neural network models, data preprocessing is a vital phase in this work. X-ray pictures classified as "Fractured" and "Not Fractured" make up the dataset; each image is preprocessed in order to guarantee consistency and improve model performance.

All pictures are first rescized to a set size of 224x224 pixels. Deep learning models—especially CNNs and Transformer-based models—need uniform dimensions of input images to effectively learn patterns, so this standardizing is essential. Resizing also helps to lower computing expenses without appreciably sacrificing the requirements for fracture detection.

The training images then undergo random rotation and horizontal flipping among other data augmentation methods. By mimicking various angles and orientations that might be faced in real-world circumstances, these augmentations are meant to provide diversity in the training set, therefore enabling the model to generalize better to unknown data. For medical photos, where little differences could affect diagnosis results, this stage is very crucial.

After that, all images are turned into tensors and normalized with mean and standard deviation values common of pre-trained models (e.g., ImageNet). By balancing the learning process, normalisation balances the pixel values to a standard range thereby improving convergence during training. Particularly, the pixel values for the RGB channels are standardized to mean values [0.485, 0.456, 0.406] and standard deviations [0.229, 0.224, 0.225], therefore satisfying the preconditions for models such as ResNet and ViT. Only scaling and normalizing are used for the validation and test datasets to guarantee that the assessment measures fairly the performance of the model on reasonable, unaffected images without augmentation. Images are shuffled during training to lower overfit and improve learning; the data loaders are then set to manage batch processing.

4.3 Model

Various deep learning architectures are investigated in this work to categorize hand X-ray pictures into "Fractured" and "Not Fractured" classes. This work uses a fully connected network (FCN), a traditional convolutional neural network (CNN), a pretrained ResNet model, and a Vision Transformer (ViT-B-16) model. Every model shows a different method of image classification, so we can evaluate their performance and find architectural strengths.

For this work, the baseline model is the fully connected network (FCN). Multiple layers of neurons make up FCNs, and each neuron in one layer is linked to every other in adjacent layers. Simple feedforward neural networks with a sequence of dense layers followed by ReLU activation functions and dropout layers to prevent overfitting make up the FCN employed in this work. Although FCNs are good for simple classification problems, they lack the capacity to collect spatial features hence they are not well suited to

handle image data. For complicated picture classification applications such fracture detection, which mostly depends on spatial patterns, this restriction reduces the efficiency of FCNs.

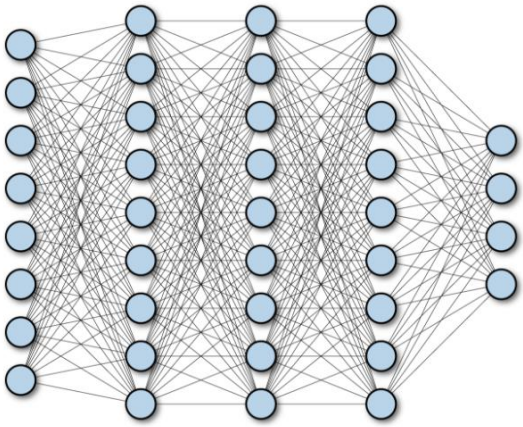

Figure 3 Fully connected neural network

By using convolutional layers especially for picture input, the custom Convolutional Neural Network (CNN) is meant to surpass the constraints of FCNs. CNNs comprise fully connected layers for final classification, pooling layers that lower data dimensionality while preserving significant features, and convolutional layers that automatically learn spatial hierarchies of features via kernels. Multiple convolutional and pooling layers as well as dropout layers to help with overfitting define the CNN utilized in this project Because CNNs can capture edges, textures, and more complicated patterns that are essential in differentiating between fractured and non-fractured X-rays, they are well-suited for jobs requiring images.

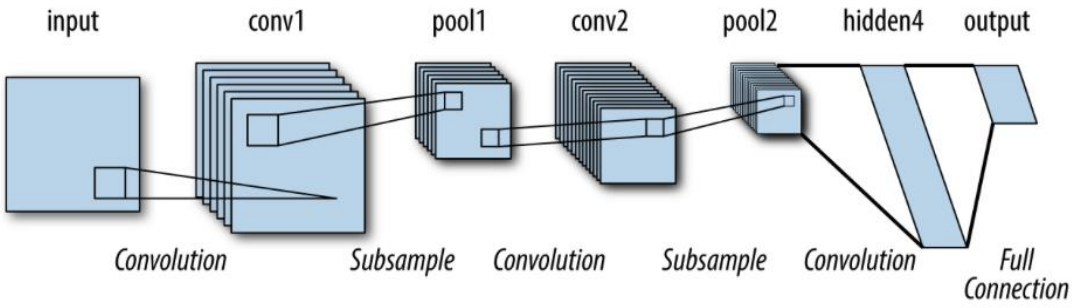

Figure 4 Convolutional neural network

Popular CNN architecture including residual connections to support deeper networks is the ResNet model—more especially, ResNet18. Residual connections allow the network to learn efficiently even with several layers by helping to solve the vanishing gradient issue. Designed as a lightweight variant of the ResNet family, ResNet18 computationally efficient while still gains from pretrained model deep learning capabilities. In this work, we employ transfer learning using a pretrained ResNet18 model trained on vast-scale image datasets. Comparatively to training from start, the pretrained weights enable the model to rapidly adapt to the challenge of fracture identification, so improving accuracy and lowering training time.

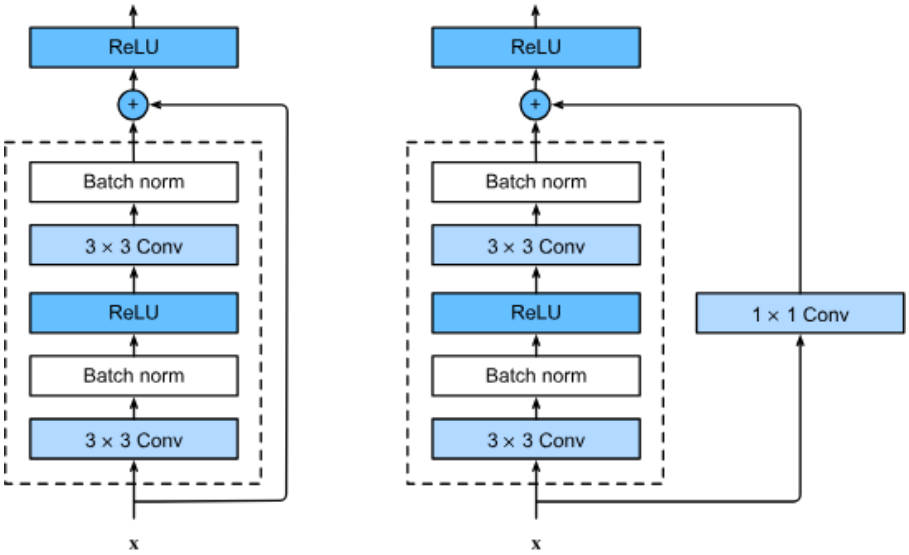

Figure 5 Resnet

In image categorization, the Vision Transformer (ViT-B-16) offers a novel method. Unlike conventional CNNs, ViTs capture global dependencies throughout the whole image by means of a transformer-based design based on self-attention processes. Divining images into patches, the ViT-B-16 model takes each patch as a token and uses self-attention to learn associations between these tokens. This method lets the model identify intricate patterns and global context that local receptive fields in CNNs could overlook. Offering state-of- the-art performance, particularly when working with complicated and delicate visual distinctions in medical images, ViT-B-16 is pretrained on vast datasets and fine-tuned for the particular purpose of fracture identification.

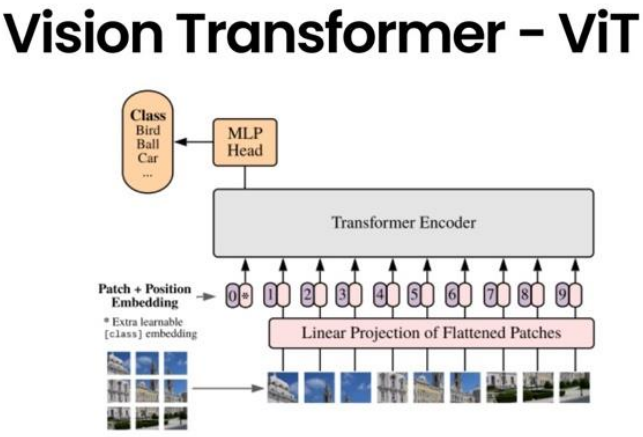

Figure 6 Vit

By means of their accuracy, loss, and other performance criteria, each of these models is assessed and offers insights on the efficiency of several neural network designs for the purpose of fracture detection in hand X-rays. The comparison study emphasizes the possibilities of modern deep learning methods in automating medical diagnoses and helps to choose the best appropriate model for clinical uses.

5. Results and analysis

The performance variations among several neural network models applied for the classification of hand X-ray pictures into "Fractured" and "Not Fractured" categories show by the figures. Together with the

confusion matrices, the loss and accuracy graphs offer a whole picture of every model's strengths and shortcomings. With consistent training and low loss and great accuracy over both training and validation datasets, the pretrained ResNet model attained the greatest performance. With just one misclassification, its confusion matrix shows almost flawless classification, hence stressing the model's resilience and the potency of transfer learning. Based on the few misclassifications it generated, the bespoke CNN also did rather well, with high accuracy but somewhat higher variability than the pretrained ResNet. From the difference between training and validation accuracy, the non-pretrained ResNet model clearly showed overfitting and generated a few more mistakes, hence emphasizing the need of pretraining. With considerable variability in loss and accuracy and a large number of misclassifications in the confusion matrix, suggesting its difficulty in learning from minimal data, the ViT, however, suffered most.

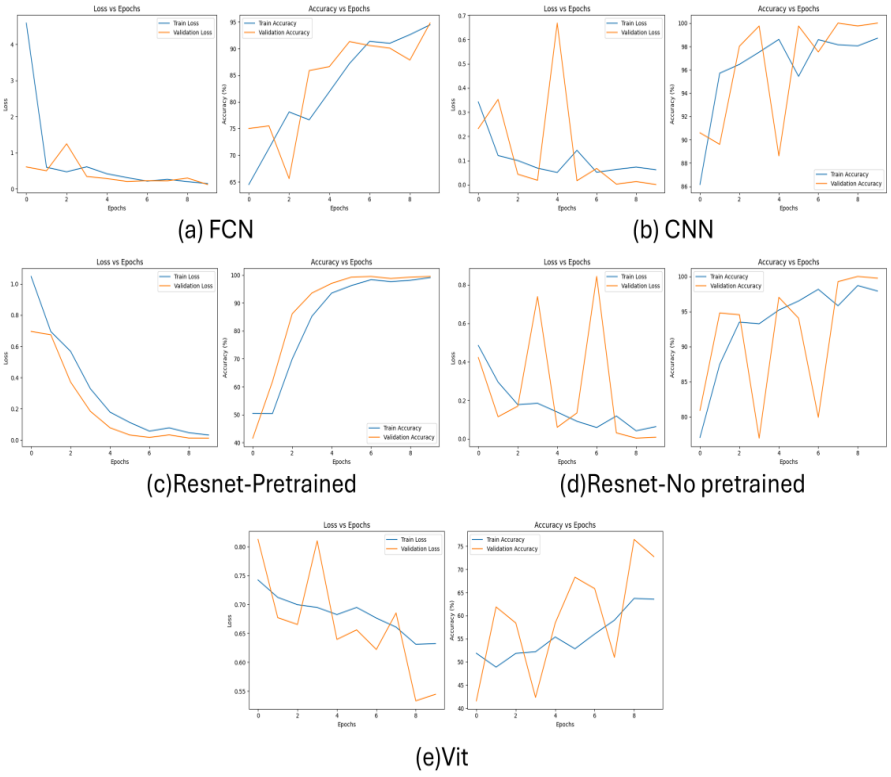

Figure 7 Training of different models

At last, the FCN showed a higher number of mistakes, which represents its incapacity to collect intricate spatial characteristics required for fracture diagnosis, thereby performing somewhat well but not as effectively as CNN-based models. These findings show generally that although pretrained CNN models—especially ResNet—are quite effective for this work, Transformer-based models like ViT may need more data or more tuning to get equivalent outcomes in medical imaging environments.

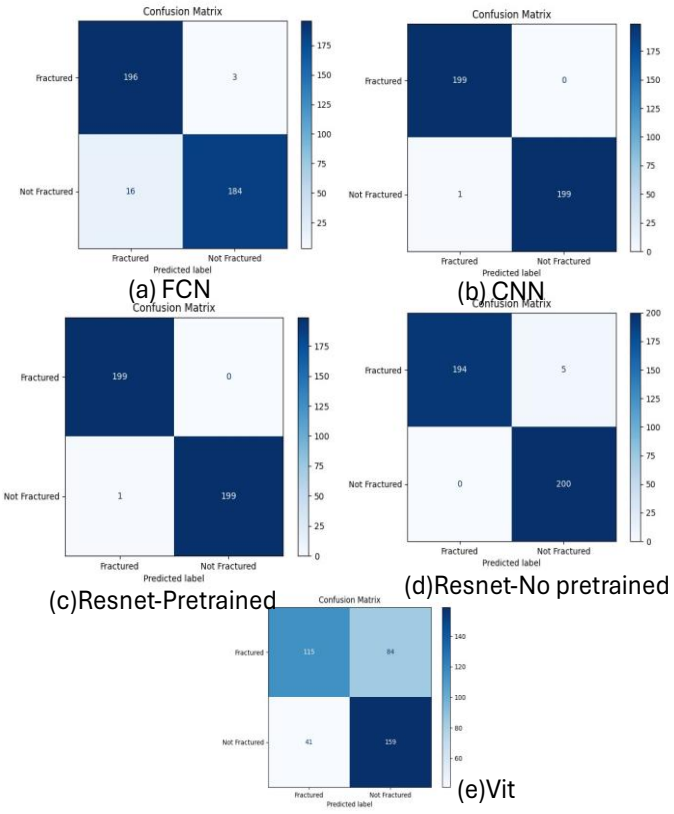

Figure 7 Testing of different models

The performance of multiple neural network models in the job of fracture identification from hand X-ray pictures is highlighted by the results analysis in several important directions. Although the pretrained ResNet obviously outperformed other models, this achievement emphasizes the reliance on transfer learning—which makes use of enormous external data and pretrained feature representations. Although helpful, this dependence on pretrained models poses questions regarding the inherent biases from the original training datasets, which might not always coincide exactly with medical imaging data. Strong performance of the bespoke CNN model also highlights that CNNs can reach high accuracy without necessarily depending on external pretrained weights, so providing a more flexible approach with enough architectural customization. Conversely, the low performance of the ViT highlights the limits of modern transformer models in specialized domains such as medical imaging when training data is limited or not adequately customized. This implies that whereas transformer models shine in huge, varied datasets, their use in more specialized or data-constrained settings like medical diagnostics calls for careful review of data volume and quality. Furthermore indicating a possible trade-off between model complexity and practical applicability in real-world clinical settings, the overfitting observed in the non-pretrained ResNet model reminds us sharply of the hazards connected with complex deep learning architectures without robust regularization or sufficient training data.

6. Conclusion

This work shows that the performance of fracture identification in hand X-ray pictures is much influenced by choosing the suitable neural network design. Thanks to the benefits of transfer learning, the pretrained ResNet performed better among the models evaluated, obtaining strong generalization and great accuracy. The bespoke CNN also showed good performance, proving that because of their capacity to collect spatial

characteristics, CNN designs are especially fit for medical picture categorization. Nevertheless, overfitting in the conventional ResNet without pretraining emphasizes the need of pretrained weights in improving model stability and accuracy. Though a state-of- the-art architecture, the Vision Transformer (ViT) suffered because of its data-intensive character and poor generalizing capabilities with the given training data. Although the Fully Connected Network (FCN) was a baseline, its simplicity lacked the depth of CNN models, therefore reducing accuracy. Emphasizing the importance of transfer learning in medical image analysis, pretrained CNN models—especially ResNet—offer generally the most dependable and accurate method for fracture detection.

Reference

1. Bay, H., Tuytelaars, T., & Gool, L. V. (2006). SURF: Speeded up robust features. Proceedings of the 9th European Conference on Computer Vision, 404–147. https://doi.org/10.1007/11744023_32
2. Cheol, K., Cheol, H., Jun, T., et al. (2020). Automatic detection and segmentation of lumbar vertebrae from X-ray images for compression fracture evaluation. Computer Methods and Programs in Biomedicine, 200(11), 105833. https://doi.org/10.1016/j.cmpb.2020.105833
3. Chung, S. W., Han, S. S., Lee, J. W., et al. (2018). Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. Acta Orthopaedica, 89(4), 468–473. https://doi.org/10.1080/17453674.2018.1453714
4. Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 1, 886–893. https://doi.org/10.1109/CVPR.2005.177
5. Derkatch, S., Kirby, C., Kimelman, D., et al. (2019). Identification of vertebral fractures by convolutional neural networks to predict nonvertebral and hip fractures: A Registry-based Cohort Study of Dual X-ray Absorptiometry. Radiology, 293(2), 404–411. https://doi.org/10.1148/radiol.2019190201
6. Fu, Z. (2011). Real AdaBoost algorithm for multi-class and imbalanced classification problems. Computer Research and Development, 48(12), 2326–2333. https://doi.org/10.1080/01932691003662381
7. Girshick, R. (2015). Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision, 1440–1448. https://doi.org/10.1109/ICCV.2015.169
8. Girshick, R., Donahue, J., Darrell, T., et al. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 580–587. https://doi.org/10.18127/j00338486-202109-11
9. Guan, B., Yao, J., Zhang, G., et al. (2019). Thigh fracture detection using deep learning method based on new dilated convolutional feature pyramid network. Pattern Recognition Letters, 125, 521–526. https://doi.org/10.1016/j.patrec.2019.06.015
10. He, K., Gkioxari, G., Dollár, P., et al. (2020). Mask R-CNN. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(2), 386–397. https://doi.org/10.1109/TPAMI.2018.2844175
11. He, K., Zhang, X., Ren, S., et al. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 770–778. https://doi.org/10.1109/CVPR.2016.90
12. Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the Dimensionality of Data with Neural Networks. Science, 313(5786), 504–507. https://doi.org/10.1126/science.1127647

13. Jin, L., Yang, J., Kuang, K., et al. (2020). Deep-learning-assisted detection and segmentation of rib fractures from CT scans: Development and validation of FracNet. EBioMedicine, 62, 103106. https://doi.org/10.1016/j.ebiom.2020.103106
14. Kim, D. H., & Mackinnon, T. (2018). Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. Clinical Radiology, 73(5), 439–445. https://doi.org/10.1016/j.crad.2017.11.015
15. Kim, M. W., Jung, J., Park, S. J., et al. (2021). Application of convolutional neural networks for distal radio-ulnar fracture detection on plain radiographs in the emergency room. Clinical and Experimental Emergency Medicine, 8(2), 120–127. https://doi.org/10.15441/ceem.20.091
16. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 2, 1097–1105.
17. Lin, T. Y., Dollár, P., Girshick, R., et al. (2017). Feature pyramid networks for object detection. Proceedings of the IEEE conference on computer vision and pattern recognition, 2117-2125. https://doi.org/10.1109/CVPR.2017.106
18. Lin, T. Y., Goyal, P., Girshick, R., et al. (2020). Focal Loss for Dense Object Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(2), 318–327. https://doi.org/10.1109/TPAMI.2018.2858826
19. Liu, W., Anguelov, D., Erhan, D., et al. (2016). SSD: Single shot multibox detector. Springer, Cham, 21–37. https://doi.org/10.1007/978-3-319-46448-0_2
20. Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2), 91–110. https://doi.org/10.1023/B.0000029664.99615.94
21. Pranata, Y. D., Wang, K. C., Wang, J. C., et al. (2019). Deep learning and SURF for automated classification and detection of calcaneus fractures in CT images. Computer Methods and Programs in Biomedicine, 171, 27–37. https://doi.org/10.1016/j.cmpb.2019.02.006
22. Raghavendra, U., Bhat, N. S., Gudigar, A., et al. (2018). Automated system for the detection of thoracolumbar fractures using a CNN architecture. Future Generation Computer Systems, 85, 184–189. https://doi.org/10.1016/j.future.2018.03.023
23. Redmon, J., Divvala, S., Girshick, R., et al. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition, 779–788. https://doi.org/10.1109/CVPR.2016.91
24. Ren, S., He, K., Girshick, R., et al. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031
25. Saunders, C., Stitson, M. O., & Weston, J. (2002). Support Vector Machine. Computer Science, 1(4), 1–28. https://doi.org/10.1007/978-3-642-27733-7_299-3
26. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. Computer Science. https://doi.org/10.48550/arXiv.1409.1556
27. Tomita, N., Cheung, Y. Y., & Hassanpour, S. (2018). Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans. Computers in Biology and Medicine, 98, 8–15. https://doi.org/10.1016/j.compbiomed.2018.05.011
28. Urakawa, T., Tanaka, Y., Goto, S., et al. (2019). Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network. Skeletal Radiology, 48(2), 239–244. https://doi.org/10.1007/s00256-018-3016-3

29. Wang, M., Yao, J., Zhang, G., et al. (2021). ParallelNet: multiple backbone network for detection tasks on thigh bone fracture. Multimedia Systems, 27(6), 1091–1100. https://doi.org/10.1007/s00530-021-00783-9
30. Yoon, A. P., Lee, Y. L., Kane, R. L., et al. (2021). Development and validation of a deep learning model using convolutional neural networks to identify scaphoid fractures in radiographs. JAMA Network Open, 4(5), 1–11. https://doi.org/10.1001/jamanetworkopen.2021.6096