Computer Assisted Methods in Engineering and Science
30(3): 369–385, 2023, doi: 10.24423/cames.682

This article belongs to the *Special Issue on Digital Twin Empowered Internet of Intelligent Things for Engineering Cyber-Physical Human Systems* edited by C. Venkatesan, Y.-D. Zhang, Q. Xin

A Supervised Approach to Musculoskeletal Imaging Fracture Detection and Classification Using Deep Learning Algorithms

Santoshachandra Rao KARANAM[1)*], Y. SRINIVAS[2)], S. CHAKRAVARTY[1)]

[1)] *Computer Science and Engineering, Centurion University of Technology and Management, Odisha, India*

[2)] *Information Technology, GITAM University, Visakhapatnam, India*

[*] *Corresponding Author e-mail: kschandra.rao@gmail.com*

Bone fractures break bone continuity. Impact or stress causes numerous bone fractures. Fracture misdiagnosis is the most frequent mistake in emergency rooms, resulting in treatment delays and permanent impairment. According to the Indian population studies, fractures are becoming more common. In the last three decades, there has been a growth of 480 000, and by 2022, it will surpass 600 000. Classifying X-rays may be challenging, particularly in an emergency room when one must act quickly. Deep learning techniques have recently become more popular for image categorization. Deep neural networks (DNNs) can classify images and solve challenging problems. This research aims to build and evaluate a deep learning system for fracture identification and bone fracture classification (BFC). This work proposes an image-processing system that can identify bone fractures using X-rays. Images from the dataset are pre-processed, enhanced, and extracted. Then, DNN classifiers ResNeXt101, InceptionResNetV2, Xception, and NASNetLarge separate the images into the ones with unfractured and fractured bones (normal, oblique, spiral, comminuted, impacted, transverse, and greenstick). The most accurate model is InceptionResNetV2, with an accuracy of 94.58%.

Keywords: musculoskeletal images, image processing, image enhancement, fracture diagnosis, fracture classification, deep neural networks.

1. Introduction

Fractures are a common emergency room (ER) condition [1]. Injuries or diseases such as osteoporosis can cause bone fractures. Fractures can cause lasting

damage or even death. An X-ray of the questionable organ is the most frequent method of diagnosing bone fractures. Reading X-rays in emergency departments (EDs), where people are in agony and fractures are not usually obvious, is difficult [2–4]. Magnetic resonance imaging (MRI), computed tomography (CT), and X-rays are the few imaging modalities that may be used to examine the musculoskeletal system. A musculoskeletal X-ray is the most common method of diagnosing fractures. This procedure involves radiologists who classify musculoskeletal imaging and doctors in the ER, where patients with acute injuries are admitted. Emergency doctors are unskilled in interpreting X-rays and make mistakes and misclassifications [26–28]. Image-classification software can assist emergency practitioners in detecting a fracture [5], which is important in ERs when a second opinion is typically unavailable.

Deep learning (DL) is a new advancement in artificial intelligence that has shown its ability to learn and prioritize essential aspects of a data set without being specifically taught. Deep learning is ideally adapted for tackling computer vision challenges because of its unique characteristics. Object detection and classification are just a few computer vision challenges solved using deep learning [29].

The studies [30–35] did not expand their investigation to numerous fracture diagnostic classes. In addition, previous research has been hindered in effectiveness either by using constrained pre-trained networks or by identifying certain layers of a network.

An automated computer-aided diagnosis (CAD) system for the BFC is described in this research. To find the optimal strategy for BFC, we compared the results of four pre-trained DNNs (ResNeXt101, InceptionResNetV2, Xception, and NASNetLarge) models. The paper breaks down into the following. Section 2 discusses the work that is linked with this subject. Section 3 discusses the procedures/methods used for this investigation. Section 4 discusses the results and any problems that may have arisen. Sections 5 and 6 provide a conclusion as well as some suggestions for the future.

2. Related work

Berlin [1] specified that in complicated clinical situations, where experienced radiologists or other practitioners may be unavailable, clinicians lack the subspecialized skill and experience required to detect fractures on radiographs effectively. There is a requirement for accurate fracture detection. Clinicians may also be subjected to heavy workloads, leading to weariness and making them more vulnerable to interpreting mistakes. According to Guly [2], good clinical skills are essential. The bulk of the abnormalities missed in the radiographs were relatively straightforward. Junior doctors should get specialized training and be

evaluated on their ability to accurately interpret musculoskeletal scans before they are permitted to practice independently. Hallas and Ellingsen [3] identified that a significant percentage of fractures were misdiagnosed during the first visit to an ED visit. The incidence of fracture misdiagnosis varied throughout the day, with a substantial surge between 8 p.m. and 2 a.m. When fracture diagnosis was incorrect, the patients did not appear to have a characteristic profile in regard to capability to communicate with the ED personnel. Improving radiological consultation services may lower the number of diagnostic mistakes, especially on nights between 8 p.m. and 2 a.m.

Wei *et al.* [4] studied an ED and extensively explored fractures in the extremities that were undetected in the initial radiological examination, especially of plain radiographs. A consensus was reached on the various causes of misdiagnosis. According to an analysis of the various causes, the most prevalent explanation for missed fractures was the fractures' subtlety. As a proportion of all fractures, the foot was the most common place for missed fractures. By partitioning the problem into smaller subproblems in support vector machine's (SVM) kernel space, He *et al.* [5] proposed a novel divide-and-conquer technique for fracture detection. As the subproblems are easier to solve than the whole problem, lower-level SVMs are trained to supplement higher-level SVMs. SVM accuracy and reliability may be improved using this method.

Liang *et al.* [6] used a partial threshold algorithm to locate fractures in bone images. After segmentation, mathematical morphology removes the needed margin and wraps fracture edges. Superposing the target border image and covering the extracted skeleton allows for detecting the precise location of fractures and therefore recognizing fractures in the X-ray photos. Bandyopadhyay *et al.* [7] proposed a method for segmenting bone parts in X-ray images based on entropy by setting a threshold for entropy values in the window next to the given image. In many cases, the bone boundary became fragmented after thresholding. To fix this problem, a multilevel-averaging filter was used to improve the bone structure of the segmented image. The automated technique outperformed two earlier approaches to the segmentation entropy quantitative evaluation metric.

Chai *et al.* [8] devised a way to use automated gray-level co-occurrence matrix strategies to diagnose femur bone fractures automatically. Using this method, we can classify whether or not there is a bone fracture based on the parameter value that comes from the matrix value. A value of 0.95 is the threshold between not having a bone fracture and having one. The suggested method has an accuracy of at least 86.67% in diagnosing bone fractures automatically. The findings imply that the system can produce consistent and reproducible outputs. Mahendran and Shereef [9] focused on automatically detecting leg bone fractures. Pre-processing, segmentation, and feature extraction were employed. The col-

lected characteristics fed into a fusion-based classification system to detect the presence or absence of fractures in an image. The proposed amalgamation of methodologies enhanced the accuracy and speed of fracture identification in experiments. Future research will examine how shape affects the detection rate.

Krizhevsky *et al.* [10] found that large, deep convolutional networks (DCN) achieved good accuracy with the help of supervised learning. Even removing a single convolutional layer degrades the performance of their network. The authors would like to use DCNs on video sequences because the temporal structure of these sequences could contain important information that is either missing or less clear in still photographs. A fracture diagnostic technique was suggested by Al-Ayyoub *et al.* [11] and was based on wavelet transformations applied to the radiography image. This technology makes use of filter techniques to locate edges as well as fractures in the hand bones. Roth *et al.* [12] developed DCN to automate the identification of fractures in spine CT. In their research, DCN, often called ConvNets, are utilized to automatically locate fractures in the spine's posterior elements. In the first step, a cross-label fusion is used to separate the spine's vertebra bodies and their back parts in a spine CT. After that, the edge maps of the elements in the back are computed. These edge maps serve as candidate regions to estimate a set of probabilities for fractures along the image edges using ConvNets.

Dimililer [13] used an intelligent bone fracture detection system (IBFDS). The author used neural networks after a pre-processing step to determine the state of a bone fracture. In his study, the author employed a training set of 30 images and examined a testing set of 70. In preparation, the images were processed using scale invariant feature transform (SIFT) and the Haar wavelet transform. The IBFDS used 1024 input neurons using a three-layered back propagation neural network technique. During various experiments, the learning rate was adjusted to achieve the required amount of inaccuracy and minimize mistakes. The author obtained a 94.3% accuracy in his result. Olczak *et al.* [14] looked at whether orthopedic imaging can detect fractures using standard deep learning. The authors used five DNN models, including the VGG-CNN network, the BVLC-CaffeNet Reference network, a network-in-network, and the VGG-CNN-S network, to accomplish their aim. It was discovered that the VGG16 network had the highest raw score, with an accuracy of 83%.

Rajpurkar *et al.* [15] trained a simple model to determine whether an X-ray is acceptable or flawed to limit the number of patients who require additional diagnosis or intervention. Class activation maps (CAMs) wereincluded to emphasize the fracture diagnosis. Chung *et al.* [16] discussed humerus fracture diagnosis in their thesis. ResNet-152 performed as good or better than most general doctors and orthopedic surgeons. To avoid overfitting, the authors used all labeled images in their processing. In this study, the authors employed data enhancement

to extend the dataset; rotations of 90, 180, and 270 degrees added data information. Due to diagnostic imaging conducted in a specified direction, only a fraction of the population could match the model. Maintaining the bone's fundamental angle would have been easy. Lindsey *et al.* [17] indicated that deep neural networks may improve fracture detection by clinicians. The authors studied wrist fractures. Data dropout, early stopping, and data augmentation reduced overfitting. This helps expert evaluations. Every study should use this method to show how CAD can aid decision-making. When pre-training a model, you change its parameters for the task, such as the skeleton, rather than using weights from a dataset trained on different parameters, such as, images. Kim and MacKinnon [18] defined fractured and unfractured wrist images. It was observed that CNNs, pre-trained on non-medical images, can be used for automatic fracture detection in plain radiographs. The accuracy was 0.954. This study showed that CNNs pre-trained on numerous data sets may be used medical imaging. All computations utilized enhanced data. Every iteration should provide a fresh image, or real-time data mixing should develop new concepts.

According to Kitamura *et al.* [19], a voting method was used to consolidate the output generated by three radiographic views and model ensemble. For single radiographic views, the ensemble of all five models, used in their study, produced the best accuracy at 76%. Despite the small dataset size, an accuracy of 81% was obtained by using an ensemble of models and three views for each instance. This level of accuracy was comparable to that attained by other models using a substantially higher number of cases with pre-trained models, and models that implemented the manual extraction of features.

Cheng *et al.* [20] suggested using a DCN to diagnose and locate fractures in the hip. To determine whether or not the model is accurate, a visualization technique known as the gradient-weighted class activation mapping (Grad-CAM) algorithm was used. The accuracy of the algorithm was 91%. Jiménez-Sánchez *et al.* [21] proposed a CAD tool based on DL algorithms that recognizes, identifies, and categorizes proximal femur fractures on musculoskeletal X-ray images. The recommended CAD tool's performance in organizing X-ray images into "A", "B", and "not-fractured" (87% accuracy) is similar to that of an average expert. In comparison, its performance in recognizing fractures versus "not-fractured" occurrences is even better (precision of 94%). The recommended CAD tool's implementation into clinical practice was also studied to increase the interaction between humans and AI-powered technology in supporting medical decisions.

Zdolsek *et al.* [22] developed a deep neural network approach to assist clinicians in diagnosing atypical femur fracture (AFF). Three common DNN architectures were used to classify the images as AFF or normal femur factures (NFF): VGG19, InceptionV3, and ResNet50. Transfer learning was used to pre-

train these networks using ImageNet images. This was accomplished via the use of five-fold cross-validation. With a 94% accuracy, ResNet50 is the most accurate. Chada [23] investigated novel model architectures and deep transfer learning to improve the identification of upper limb abnormalities while simultaneously training with little data. Deep learning models such as DenseNet-169, DenseNet-201, and InceptionRevV2 were implemented. These models were implemented because of their high recognition accuracy by employing a massive dataset of radiographic images of the human humerus and finger readily accessible to the public from MURA. Deep transfer learning models DenseNet-201 and InceptionResNetV2 identified anomalies in humerus radiographs with 83–92% accuracy and sensitivity greater than 0.9.

Jones *et al.* [27] in their research studied a deep learning system that could help doctors learn more about imaging of the musculoskeletal system. The authors developed a deep learning system to diagnose fractures in the musculoskeletal system. This system was trained using data that prominent orthopedic surgeons and radiologists manually annotated. The overall AUC of the system that used deep learning was 0.974.

Kandel *et al.* [28] studied how transfer learning (TL) affected the classification of musculoskeletal images. We discovered that in the 168 results obtained using six different CNN architectures and seven distinct bone types, TL achieved better results than training a CNN from scratch. Only 3 out of the 168 results showed that training-from-scratch produced better outcomes than transfer learning. The weaker performance of the training-from-scratch approach could be related to the choice of the hyperparameters. In the CNNs considered, there is a large number of weights (trainable parameters). Additionally, the number of images used to train these networks is too small to develop a workable model. Jin *et al.* [30] suggested using a DL system known as FracNet to locate and differentiate rib fractures by examining CT image data. In addition, the developed deep learning system performed very well in the earlier and published rib fracture segmentation test. Deep learning enhanced rib fracture detection sensitivity by 94%, with acceptable false positives and less clinical time. Chen *et al.* [31] showed that plain abdominal anterior images used for clinical reasons may be utilized to diagnose vertebral fractures with excellent accuracy. The proposed technology may help doctors repair fractures more efficiently and affordably. The visualization algorithm Grad-CAM was use for model interpretation. Ma and Luo [32] delivered a novel method for determining the location of fractures in X-rays. Crack-sensitive convolutional neural network (CrackNet), a newly developed categorization network, can identify fracture lines and characterize them accurately. The proposed technique can locate fractured bone and pinpoint its location in an X-ray image. It might assist medical professionals in finding a bone fracture. Extensive testing of the Radiopaedia dataset provided conclusive evidence that

the proposed technique is successful, outperforming other methods in terms of accuracy (88.39%), recall (87.50%), and precision (89.09%).

Karanam *et al.* [33] concluded that fracture detection and classification methods have traditionally consisted of steps such as data preparation, image preprocessing, feature extraction, and classification. Radiologists with extensive training are required to examine and classify radiographic images; this is a labor-intensive and expensive process that might be expedited with the use of fracture classification algorithms. The lack of labeled training data sets is cited as the most critical obstacle on the path to constructing a high-performance classification model by several academics addressed in this research. The authors have made an effort to provide several distinct ideas and points of view, all of which have the potential to play a role in the creation of an ideal model that can identify various types of fractures in a range of anatomical sites. Karanam *et al.* [34] researched classified bone fractures using machine learning and deep learning. K-nearest neighbor (KNN), SVM, random forest, Inception V3, and ResNeXt101 were used to evaluate, categorize and recognize fractures. ResNeXt101 improved test scores by 93.75%. The suggested approach helps doctors and radiologists determine which bone is shattered, classify the fracture, and propose therapy. According to the studies that were conducted, the majority of researchers are interested in the diagnosis of bone fractures, but in more recent years, there have been some academics who have been interested in the classification of bone fractures.

According to the statistics presented above, out of the 30 researchers, 27 focused on the binary classification of bone fractures, while only 3 considered the multi-classification of bone fractures. The detection of bone fractures was the primary focus of most early investigations rather than their classification.

With the assistance of ResNeXt101, InceptionResNetV2, Xception, and NASNetLarge deep learning algorithms, in this study the diagnosis and classification of bone fractures is conducted. In this article, we expand the work by looking at how a CNN is used to categorize X-ray images to identify bone fractures using TL.

3. Materials and methods

A general framework for the multi-class classification of bone fractures is shown in Fig. 1. This was one of our suggestions. An X-ray picture of a bone fracture goes through a pre-processing phase to combine it with the input dimension of the architectures that are being utilized in this inquiry. This step is conducted on X-ray images. Images that have been processed are used by the architecture so that features may be isolated and characteristics can be fine-tuned. In the end, all of the components are merged to produce the output image, which

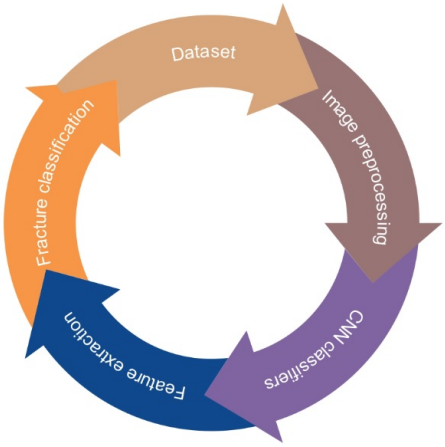

FIG. 1. Proposed architecture for multi-class BFC.

is then sorted into one of six categories based on the type of bone fracture it depicts. To conduct this investigation, the Anaconda platform for Python 3.7 was used.

3.1. Dataset

There are several ongoing studies where researchers and radiologists gather and annotate radiographic images. One of the major drawbacks in making direct comparisons between different systems is the lack of a publicly accessible standard data set, which has an impact on the model's performance in private data sets used by academics. Some popular dataset sources are the MURA data set (40 000 images) [24, 26], Medpix – an online database of medical images, and Radiopaedia where over 4800 fracture cases with diagnoses are publicly available. The sample X-ray images dataset was gathered from the National Orthopedic Hospitals and Open Access Medical Image Repositories. Sample X-ray images of bone fracture types are shown in Fig. 2.

FIG. 2. Sample images of different types of bone fractures.

3.2. Preprocessing

To ensure the proposed method applies to a wider range of contexts, the preprocessing stages have been preserved. The X-ray image is converted to grayscale so that it can be processed more rapidly and with less strain on the memory than would otherwise be required. The Crimmins technique is applied during the process of eliminating noise from the X-ray image, while the contrast limited adaptive histogram equalization (CLAHE) is utilized during the process of smoothing and edge sharpening. Both of these processes are described in more detail below [33]. The integrated tool offered by the Keras Image Data Generator allowed us to successfully execute the basic pre-processing step. The original images had a resolution that ranged from 450 to 600 pixels. We decreased the resolution of the bone fracture X-ray images in the dataset to 299×299 pixels so that they could be used as input to the deep learning classifier models. Figure 3 displays sample images that have been pre-processed and are of various bone fractures.

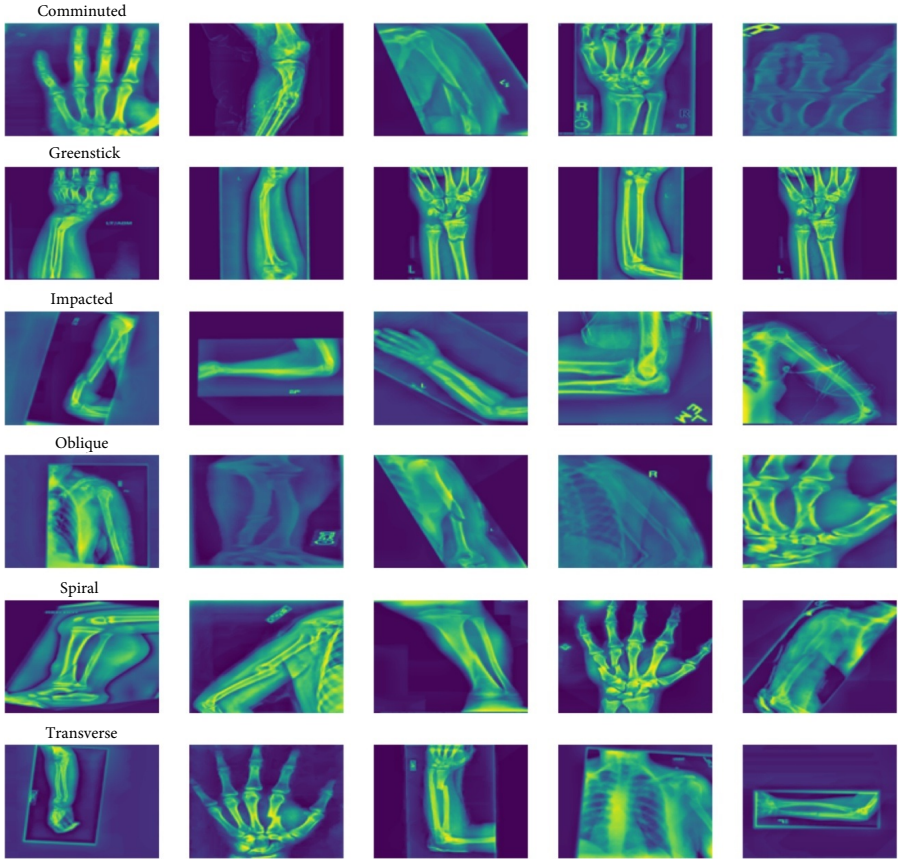

FIG. 3. Sample pre-processed images of bone fracture types from the dataset.

3.3. Classification models

The subsequent stage is the categorization of the fracture type, which is carried out using CNN classifier approaches such as ResNeXt101, InceptionV3, Xception, and InceptionResNetV2. These classifiers were primarily chosen because they are quick and easy to use, which speeds up training and improves classification accuracy.

The proposed architecture for multi-class BFC is shown in Fig. 4. Flow chart of BFC is presented in Fig. 5.

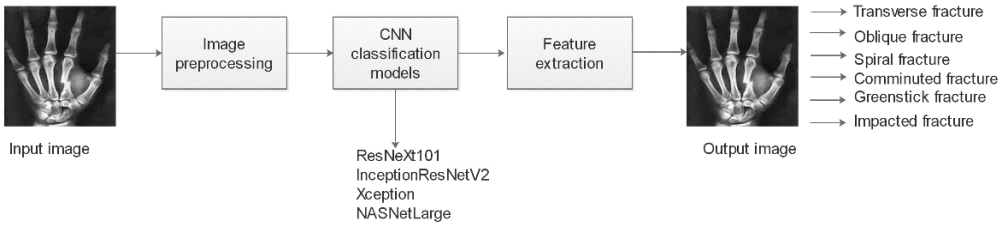

Fig. 4. Architecture for multi-class BFC.

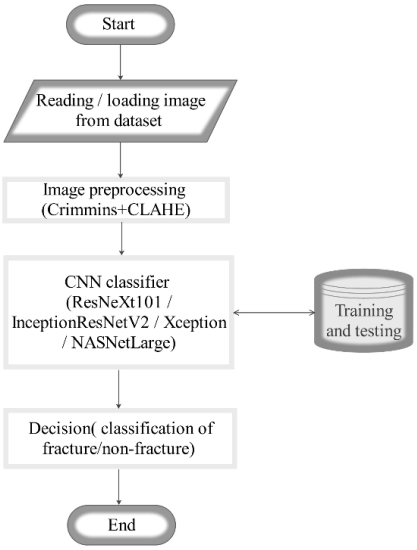

Fig. 5. Flow chart of BFC.

3.3.1. ResNeXt101. To deal with accuracy saturation and degradation, the ResNet system was proposed, which applied residual connections as a way to improve network depth. There are three versions of ResNet: 50, 101, and 152. ResNeXt101's residual learning architecture makes it simpler to train deeper networks and reformulate the layers of those networks. The increased depth of the ResNeXt101 model makes it easier to converge on a solution, which ulti-

mately leads to an improvement in the model's accuracy. The inclusion of a thick layer to ResNeXt101, which included a rectified linear unit (ReLU) activation, softmax layers, and dropouts which have seven outputs, led to a considerable improvement in the performance of the software. This architecture was developed by examining 1200 images throughout 30 epochs, with a learning rate of 0.0001 and 0.9 stochastic gradient descent (SGD) momentum, respectively [23, 34].

3.3.2. Inception-ResNet-v2. Over a million images from the ImageNet collection were used during the training process for the CNN known as Inception-ResNet-v2. This layered network is capable of grouping images into a thousand separate object categories. Because of this, the network learns a wide range of different rich feature representations for several image categorizations. As input, the network is given an image that is 299 by 299 pixels, and as output, it generates a list of anticipated class probabilities [23, 34].

The inception structure and the residual connection have been merged into one another to form the basis of this formulation. The Inception-ResNet block is created by coupling convolutional filters of different sizes with residual connections. When residual connections are used, not only is it possible to circumvent the degradation problem that is brought by deep structures, but it also makes it possible to reduce the amount of time that is spent being trained. The graphic offers a summary of the core network architecture that underlies Inception-ResNet-v2 in its current iteration. This architecture was optimized using 1200 images throughout 30 epochs with a learning rate of 0.0001 per image and 0.9 SGD momentum.

3.3.3. Xception. The Xception architecture is composed of a linear mixture of layers combined with residual connections. This helps to reduce the complexity of the design. Compared to earlier iterations of DCNs, the Xception model focuses more on efficiently using all of the model parameters at its disposal. Furthermore, it does away with the inception modules and replaces them with depth-wise independent convolution layers [23, 34]. To enhance Xception's performance of the dataset, we updated its design by including a dense layer with a softmax layer and the ReLU activation with seven outputs. This was accomplished by adding a softmax layer. Each of these two levels has seven outputs available to it. To speed up model optimization, we use the adaptive moment estimation (Adam) optimizer to fine-tune the modified version of Xception on 1200 images at a learning rate of 0.001. During the process of adjusting the system to its optimal state, 30 epochs were used. Xception is a deep convolutional neural network (CNN) that consists of 71 layers. You can load a version of the network that has already been trained from the ImageNet dataset. This version of the network has learned from looking at more than a million images. The net-

work has already been trained to put images into a thousand different categories based on what they are about. This design was developed by using a learning rate of 0.0001 per image and 0.9 SGD momentum throughout 30 epochs, involving a total of 1200 images.

3.3.4. NASNetLarge. NASNet designs provide common and lessening cell ideas, both of which are amenable to modification via the use of the reinforcement learning search strategy. Training over very large datasets was the motivation for the development of NASNetLarge architecture, which was released in 2017 [31, 32]. Because the cost of training on a big dataset is so significant, the search for an intelligent building block is first carried out on a dataset that has a relatively small amount of data. The block is then transferred to a larger dataset using the NASNet search space. A fundamental component of NASNetLarge is the regularization approach known as ScheduledDropPath, which significantly improves the level of generalization achieved by NASNet models. The pre-trained architecture of NASNetLarge was modified such that it now includes a dense layer with 'ReLu' activation, softmax layers, and dropouts which have seven outputs. The design was developed by using a learning rate of 0.0001 per image and 0.9 SGD momentum throughout 30 epochs, involving a total of 1200 images.

3.4. Feature extractions

To combine two pre-trained models, the integrated feature extractor was used. When it comes to the construction and training of deep convolutional neural networks, each strategy is effective and helps save time. To extract features, it makes use of transfer learning, which is a method in which a model is trained and refined in one problem before being applied to another. Pre-trained networks are used because they learn low-level properties such as lines, borders, and other aspects that may be used for future projects. Finally, we integrate the output of the pre-trained model with additional layers. This training started with pre-trained models and was then customized to our problem. The pre-trained model's weights were fixed throughout training to avoid weight changes during the new model training. Hidden layers in convolutional neural networks are responsible for the transmission of input data to higher-level internal representations. As inputs move up the network, they become more important to the system. Each layer's characteristics are recorded in an image for classification.

3.5. Performance matrix

To test the model's performance, we looked at 1200 X-ray images. The method used divides the expected image into four subsets: TP represents the figure of

properly recognized positives. TN is the figure of properly recognized negative occurrences. FP is the figure of misclassified positives. FN is the figure of incorrectly classified negatives [24, 29, 33]. The accuracy of a model may be determined by looking at its accuracy. Equation (1) demonstrates this. Precision, sensitivity, specificity, and F1-score are four additional performance metrics for multi-class classification, represented by Eqs. (2)–(5):

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})}, \tag{1}$$

$$\text{Precision} = \frac{(\text{TP})}{(\text{TP} + \text{FP})}, \tag{2}$$

$$\text{Sensitivity} = \frac{(\text{TP})}{(\text{TP} + \text{FN})}, \tag{3}$$

$$\text{Specificity} = \frac{(\text{TN})}{(\text{TN} + \text{FP})}, \tag{4}$$

$$\text{F1-score} = 2 * \frac{(\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}. \tag{5}$$

4. Results and analysis

The conclusion was reached after analyzing validation data that included 1200 images (960 training set + 240 testing set) dataset of six types of bone fracture. Keras library was utilized to implement the deep models employed in this study for multimedia tools and applications. Keras has been able to run other profound learning frameworks such as TensorFlow or Theano. Possible outcomes of the confusion matrix are shown in Table 1 and the specificity, sensitivity, precision, accuracy, and F1-score of CNN classifier models are shown in Table 2.

Table 1. Fracture analysis results.

CNN classifier model	Truth data	
	N	F
ResNeXt101	TN-44	FN-4
	FP-12	TP-180
InceptionResNetV2	TN-45	FN-3
	FP-10	TP-182
Xception	TN-42	FN-6
	FP-14	TP-178
NASNetLarge	TN-43	FN-5
	FP-12	TP-180

Table 2. Specificity, sensitivity, precision, accuracy, and F1-score of CNN classifier models.

Classifier model	Specificity [%]	Sensitivity [%]	Precision [%]	Accuracy [%]	F1-score
ResNeXt101	78.57	97.82	93.75	93.33	93.53
InceptionResNetV2	81.81	98.37	94.79	94.58	94.68
Xception	75.00	96.73	92.70	91.66	92.17
NASNetLarge	78.18	97.29	93.75	92.91	93.32

For categorization of bone fractures into six classes: transverse, oblique, spiral, comminuted, greenstick, and impact fracture, we tested the performance of four different models: ResNeXt101, InceptionResNetV2, Xception, and NASNetLarge. Categorical accuracy was determined to be 93.33%, 94.58%, 91.66%, and 92.91% for ResNeXt101, InceptionResNetV2, Xception, and NASNetLarge, respectively, as shown in Table 2. ResNeXt101 and InceptionResNetV2 have the highest accuracy.

5. Conclusion

This research aimed to report on the many deep-learning algorithms that may be used to determine bone fractures from radiographs. The primary focus of this research was on the challenges associated with distinguishing fractures from images of healthy bone and identifying many types of fractures. Pre-processing, feature extraction, training, and testing were the system's four primary phases. The primary consideration in selecting the classifiers was how easily they could be applied and how accurate their results were. Classifiers were used to differentiate between the many types of fracture patterns that were present in the photos. The results of the system have been empirically validated through the use of two different study methodologies. All available types – TP (prediction of true as true), TN (prediction of true as false), FP (prediction of false as false), and FN (prediction of false as true) were used to evaluate the effectiveness and precision of the model performance. The findings of the tests indicate that the utilization of Inception-ResNet-v2 resulted in a success rate of 94.58%. The new technology facilitates a quick and straightforward diagnosis for medical practitioners and radiologists.

6. Future research

Identification and classification of bone fractures using automation improve diagnostic precision. The medical staff may then evaluate the severity of the crack. It elucidates the shape, location, and difficulty of the fracture. Inputs

relating to the categorization of the fracture may be used to get information about the fracture that exists. In further research, the emphasis may be placed on the creation of deep-learning computer-assisted fracture detection systems that are more effective. Future research may develop a novel approach for developing computer-assisted fracture detection systems.

References

1. L. Berlin, Defending the "missed" radiographic diagnosis, *American Journal of Roentgenology*, **176**(2): 317–322, 2001, doi: 10.2214/ajr.176.2.1760317.
2. H.R. Guly, Diagnostic errors in an accident and emergency department, *Emergency Medicine Journal*, **18**(4): 263–269, 2001, doi: 10.1136/emj.18.4.263.
3. P. Hallas, T. Ellingsen, Errors in fracture diagnoses in the emergency department-characteristics of patients and diurnal variation, *BMC Emergency Medicine*, **6**, article number 4, 2006, doi: 10.1186/1471-227X-6-4.
4. C.-J. Wei, W.-C. Tsai, C.-M. Tiu, H.-T. Wu, H.-J. Chiou, C.-Y. Chang, Systematic analysis of missed extremity fractures in emergency radiology, *Acta Radiologica*, **47**(7): 710–717, 2006, doi: 10.1080/02841850600806340.
5. J.C. He, W.K. Leou, T.S. Howe, Hierarchical classifiers for detection of fractures in X-ray images, [in:] W.G. Kropatsch, M. Kampel, A. Hanbury [Eds.], *Computer analysis of images and patterns. CAIP 2007. Lecture Notes in Computer Science*, vol. 4673, Springer, Berlin, 2007, doi: 10.1007/978-3-540-74272-2_119.
6. J. Liang, B.-C. Pan, Y.-H. Huang, X.-Y. Fan, Fracture identification of X-ray image, [in:] *2010 International Conference on Wavelet Analysis and Pattern Recognition*, pp. 67–73, 2010, doi: 10.1109/ICWAPR.2010.5576438.
7. O. Bandyopadhyay, B. Chanda, B.B. Bhattacharya, Entropy-based automatic segmentation of bones in digital X-ray images, [in:], S.O. Kuznetsov, D.P. Mandal, M.K. Kundu, S.K. Pal [Eds.], *Pattern Recognition and Machine Intelligence*, *PReMI 2011, Lecture Notes in Computer Science*, Vol. **6744**, Springer, Berlin, Heidelberg, 2011, doi: 10.1007/978-3-642-21786-9_22.
8. H.Y. Chai, L.K. Wee, T.T. Swee, S. Salleh, A.K. Ariff, Kamarulafizam, Gray-level co-occurrence matrix bone fracture detection, *American Journal of Applied Sciences*, **8**(1): 26–32, 2011, doi: 10.3844/ajassp.2011.26.32.
9. S.K. Mahendran, I.K. Shereef, An enhanced tibia fracture detection tool using image processing and classification fusion techniques in X-ray images, *Global Journal of Computer Science and Technology*, **11**(14): 23–28, 2011.
10. A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, *Communications of the ACM*, **60**(6): 84–90, 2017, doi: 10.1145/3065386.
11. M. Al-Ayyoub, I. Hmeidi, H. Rababah, Detecting handand bone fractures in X-ray images, *Journal of Multimedia Processing and Technologies* (JMPT), **4**(3): 155–168, 2013, doi: 10.13140/RG.2.1.2645.8327.
12. H.R. Roth, Y. Wang, J. Yao, L. Lu, J.E. Burns, R.M. Summers, Deep convolutional networks for automated detection of posterior-element fractures on spine CT, [in:] *Pro-*

ceedings SPIE 9785, Medical Imaging 2016: Computer-Aided Diagnosis, 97850P, 2016, doi: 10.1117/12.2217146.

13. K. Dimililer, IBFDS: Intelligent bone fracture detection system, *Procedia Computer Science*, **120**: 260–267, 2017, doi: 10.1016/j.procs.2017.11.237.

14. J. Olczak *et al.*, Artificial intelligence for analyzing orthopedic trauma radiographs, *Acta Orthopaedica*, **88**(6): 581–586, 2017, doi: 10.1080/17453674.2017.1344459.

15. P. Rajpurkar *et al.*, MURA: Large dataset for abnormality detection in musculoskeletal radiographs, *arXiv:1712.06957v4*, 2017, doi: 10.48550/arXiv.1712.06957.

16. S.W. Chung *et al.*, Automated detection and classification of the proximal humerus fracture by using deep learning algorithm, *Acta Orthopaedica*, **89**(4): 468–473, 2018, doi: 10.1080/17453674.2018.1453714.

17. R. Lindsey *et al.*, Deep neural network improves fracture detection by clinicians, *Proceedings of the National Academy of Sciences of the United States of America*, **115**(45): 11591–11596, 2018, doi: 10.1073/pnas.1806905115.

18. D.H. Kim, T. MacKinnon, Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks, *Clinical Radiology*, **73**(5): 439–445, 2018, doi: 10.1016/j.crad.2017.11.015.

19. G. Kitamura, C. Chung, B.E. Moore, Ankle fracture detection utilizing a convolutional neural network ensemble implemented with a small sample, de novo training, and multiview incorporation, *Journal of Digital Imaging*, **32**(4): 672–677, 2019, doi: 10.1007/s10278-018-0167-7.

20. C.-T. Cheng *et al.*, Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs, *European Radiology*, **29**(10): 5469–5477, 2019, doi: 10.1007/s00330-019-06167-y.

21. A. Jiménez-Sánchez *et al.*, Precise proximal femur fracture classification for interactive training and surgical planning, *International Journal for Computer Assisted Radiology and Surgery*, **15**(5): 847–857, 2020, doi: 10.1007/s11548-020-02150-x.

22. G. Zdolsek, Y. Chen, H.P. Bögl, C. Wang, M. Woisetschläger, J. Schilcher, Deep neural networks with promising diagnostic accuracy for the classification of atypical femoral fractures, *Acta Orthopaedica*, **92**(4): 394–400, 2021, doi: 10.1080/17453674.2021.1891512.

23. G. Chada, Machine learning models for abnormality detection in musculoskeletal radiographs, *Reports*, **2**(4): 26, 2019, doi: 10.3390/reports2040026.

24. L. Tanzi, E. Vezzetti, R. Moreno, S. Moos, X-ray bone fracture classification using deep learning: a baseline for designing a reliable approach, *Applied Sciences*, **10**(4): 1507, 2020, doi: 10.3390/app10041507.

25. A. Oyeranmi, B. Ronke, R. Mohammed, A. Edwin, Detection of fracture bones in X-ray images categorization, *Journal of Advances in Mathematics and Computer Science*, **35**(4): 1–11, 2020, doi: 10.9734/JAMCS/2020/v35i430265.

26. D. Joshi, T.P. Singh, A survey of fracture detection techniques in bone X-ray images, *Artificial Intelligence Review*, **53**: 4475–4517, 2020, doi: 10.1007/s10462-019-09799-0.

27. R.M. Jones *et al.*, Assessment of a deep-learning system for fracture detection in musculoskeletal radiographs, *NPJ Digitital Medicine*, **3**: 144, 2020, doi: 10.1038/s41746-020-00352-w.

28. I. Kandel, M. Castelli, A. Popovic, Musculoskeletal images classification for detection of fractures using transfer learning, *Journal of Imaging*, **6**(11): 127, 2020, doi: 10.3390/jimaging6110127.

29. T. Bergs, C. Holst, P. Gupta, T. Augspurger, Digital image processing with deep learning for automated cutting tool wear detection, *Procedia Manufacturing*, **48**: 947–958, 2020, doi: 10.1016/j.promfg.2020.05.134.

30. L. Jin *et al.*, Deep-learning-assisted detection and segmentation of rib fractures from CT scans: Development and validation of FracNet, *eBioMedicine*, **62**: 103106, 2020, doi: 10.1016/j.ebiom.2020.103106.

31. H.-Y. Chen *et al.*, Application of deep learning algorithm to detect and visualize vertebral fractures on plain frontal radiographs, *PLoS ONE*, **16**(1): e0245992, 2021, doi: 10.1371/journal.pone.0245992.

32. Y. Ma, Y. Luo, Bone fracture detection through the two-stage system of crack-sensitive convolutional neural network, *Informatics in Medicine Unlocked*, **22**: 100452, 2021, doi: 10.1016/j.imu.2020.100452.

33. S.R. Karanam, Y. Srinivas, S. Chakravarty, A systematic review on approach and analysis of bone fracture classification, *Materials Today: Proceedings*, 2021, doi: 10.1016/j.matpr.2021.06.408.

34. S.R. Karanam, Y. Srinivas, S. Chakravarty, A systematic approach to diagnosis and categorization of bone fractures in X-ray imagery, *International Journal of Healthcare Management*, pp. 1–12, 2022, doi: 10.1080/20479700.2022.2097765.

35. S.R. Karanam, Y. Srinivas, S. Chakravarty, A statistical model approach based on the Gaussian Mixture Model for the diagnosis and classification of bone fractures, *International Journal of Healthcare Management*, pp. 1–12, 2023, doi: 10.1080/20479700.2022.2161146.

Received June 30, 2022; revised version August 13, 2022;
accepted August 26, 2022.