Advanced Engineering Technology and Application
An International Journal

http://dx.doi.org/10.18576/aeta/130111

Comparative Analysis of Deep Learning Models for Fracture Detection and Classification in X-ray Images

A.K. Hamzat[1], M.S. Murad[1], M. Kanan[2,] and R. Asmatulu[1,*]*

[1] Department of Mechanical Engineering Wichita State University, 1845 Fairmount, Wichita, KS, 67260, USA.
[2] Department of Industrial Engineering, University of Business and Technology, Jeddah, 21448, Saudi Arabia.

Received: 27 Jul. 2024, Revised: 2 Aug. 2024, Accepted: 5 Aug. 2024.
Published online: 1 Sep. 2024.

Abstract: For efficient medical diagnosis and treatment planning, which have historically relied on expert interpretation of radiographic images, accurate classification of bone fractures is essential. In this work, we introduce an automated method that uses machine learning techniques to improve fracture classification accuracy and efficiency. We use a large dataset with various bone fracture images to compare the effectiveness of two different models: Multilayer Perceptron (MLP) and Convolutional Neural Network (CNN). By utilizing sophisticated preprocessing methods to maximize feature extraction and reduce noise, we thoroughly assess the models' classification performance. Our results show a substantial difference in performance between the two models: the CNN model achieves an amazing classification accuracy of 94%, while the MLP model only manages 73%. This significant advancement highlights how well the CNN model can represent complex fracture patterns, highlighting its potential to transform orthopedic medicine's diagnostic procedures and improve patient care. These machine learning algorithms present a promising option for improving treatment results for individuals with bone fractures by automating fracture categorization and lowering reliance on subjective human interpretation.

Keywords: Deep learning; Classification; Convolutional Neural Network; Healthcare.

1 Introduction

An accident or continuous pressure on the bones can cause a bone fracture, which is the partial or total breaking of a bone. When it comes to medical diagnosis and treatment planning, the categorization of bone fractures such as Transverse (Type1), Oblique (Type2), Spiral (Type3), Comminuted (Type4), Greenstick (Type5), and Impacted fractures (Type6) is crucial.[1] Fracture identification using traditional manual inspection and X-ray methods has shown to be ineffective, potentially resulting in missed fractures and delayed diagnosis.[2] The demand for prompt and reliable identification of bone fractures has increased due to an increase in their prevalence worldwide. It is impossible to overestimate how profoundly deep learning and artificial intelligence (AI) have changed the field of fracture detection and classification in this context.[3]

The 206 bones that make up the human body vary greatly in size, form, and complexity, making it difficult for medical professionals to properly diagnose and categorize fractures.[4] Bone fractures in the lower limbs are especially prevalent. The promise of machine learning as a pattern identification technique to analyze medical imaging and help doctors diagnose patients more accurately has drawn attention recently. [5] Determining the severity and the best course of therapy for a fracture requires early identification and accurate categorization. The field of fracture diagnosis could undergo a revolution with the incorporation of cutting-edge technology such as artificial intelligence and deep learning, given the increasing global prevalence of bone fractures.

A key component of the medical image distribution process is the Digital Imaging and Communications in Medicine (DICOM) standard. X-rays are among the most often utilized diagnostic tools for bone fractures because of their speed, affordability, and user-friendliness.[6] Medical imaging has advanced quickly since Wilhelm Roentgen discovered X-rays in 1895, and it is now an essential component of modern diagnostics. With complex algorithms needed to identify and analyze anomalies in medical imaging of the human skeleton, machine learning has become an essential tool for medical data analytics.[7]

The constraints of manual inspection and standard X-ray technologies have resulted in inefficiencies in fracture identification, even with the availability of traditional diagnostic methods. [8] By combining deep learning approaches with computer vision systems, there is a chance to improve fracture diagnosis accuracy and efficiency by screening X-ray pictures for anomalies.

*Corresponding author e-mail: ramazan.asmatulu@wichita.edu

Since human inspection and traditional X-ray procedures present some problems, the creation of an automated diagnostic tool has long been an appealing idea. The intricacies of fracture detection have been investigated through the integration of machine learning, encompassing preprocessing, feature extraction, and fracture identification. By providing a thorough review of the use of deep learning techniques more particularly, Multilayer Perceptron (MLP) and Convolutional Neural Network (CNN) models in the automated categorization of bone fractures, this study seeks to advance this rapidly developing subject.[9]

We present a thorough examination of the state-of-the-art in fracture identification and classification in this paper, emphasizing the shift from traditional machine learning techniques to the most recent advancements made possible by deep learning methods. We examined a few selected studies that deal with fracture detection and classification, highlighting the benefits of deep learning techniques in creating a universal tool that can identify all kinds of fractures in the different bones in the human body. The paper goes on to describe our methodology for using CNN and MLP models, displaying impressive outcomes that demonstrate CNNs' potential to greatly increase fracture classification accuracy. Our goal is to provide insightful information through this research that may facilitate automated fracture detection systems and improve patient outcomes in orthopedic care.

2 Related works

The growing frequency of bone fractures which are more common as a result of injuries and accidents is the subject of this study. The authors [10] investigate different methods of fracture detection, with special attention to the Indian setting, where hospital records show a considerable increase in fracture incidence during the previous thirty years. The introduction and data preparation, review of related work, feature extraction methods, traditional and deep learning-based fracture detection methods, performance evaluation approaches, and challenges faced by fracture detection researchers are the six sections that make up the structure of the paper. This research highlights the significance of appropriately identifying fractures, despite the fact that many studies only look for fractures. Its objective is to further fracture detection technologies by offering guidance and assistance to researchers who are working to create models that can automatically identify and categorize bone fractures.

In order to overcome the difficulties caused by the intrinsic fuzziness of ordinary X-ray scans, this work [11] investigates the application of machine learning methods in enhancing bone fracture identification from X-ray images. The study highlights how computers and other cutting-edge technologies are affecting many facets of human existence, including healthcare. To improve diagnostic precision and streamline surgeons' workflow, the suggested approach integrates critical phases such pre-processing, edge detection, feature extraction, and machine learning classifications. Using a dataset of 270 X-ray pictures, the study assesses a number of machines learning methods, including Naïve Bayes, Decision Tree, Nearest Neighbors, Random Forest, and SVM. The accuracy measures of these algorithms vary from 0.64 to 0.92. Among them, SVM has the best accuracy, outperforming other models and proving to be useful in the identification of bone fractures. By demonstrating machine learning's potential to improve diagnostic capabilities and, eventually, patient care outcomes, this research adds to the expanding field of machine learning applications in healthcare.

The challenge of misdiagnosed or undetected bone fractures is the main topic of this study.[12] This is a serious problem in orthopedics that can result in longer treatment times and discomfort for patients. Using a dataset that includes both normal and broken bones, the research uses a variety of machine learning approaches to identify and categorize fractures. X-ray pictures undergo initial preprocessing, and then features are extracted using techniques including Harris corner detection, Hough line detection, and Canny and Sobel edge detection. The collected features are then analyzed by twelve distinct machine learning classifiers, whose hyperparameters are chosen using grid search and testing is done by 10-fold cross-validation. Comparisons between the findings are shown, emphasizing testing, training, and accuracy rates. Interestingly, with an AUC of 0.89, linear discriminant analysis (LDA) attains the maximum accuracy rate of 88.67%. By correctly recognizing fractures, the suggested computer-aided diagnosis system (CAD) has the potential to reduce medical staff workloads and enhance patient outcomes.

The goal of this study [13] is to increase radiologists' productivity by streamlining their workflow and addressing the urgent demand for computer-aided diagnosis and detection of bone fractures in contemporary medicine. While several image processing techniques have been used in the past to detect fractures, this study focuses on using deep learning models, more specifically, convolutional neural networks (CNNs) for this goal. The study attempts to automate the process of determining fracture occurrence, locations, and severity by extending the use of deep learning models to the identification of bone fractures in X-ray pictures. To improve the reliability of fracture detection, X-rays of patients' elbows, hands, and feet are analyzed using a variety of deep learning models, with an emphasis on precise feature extraction.

The problem of early categorization and identification of bone cancer, a rare but highly metastatic cancer that carries severe dangers to patients, is the main focus of this work.[14] The study presents Adaptive Fuzzy Clustering by Local Approximation of Membership (AFLAME), a novel approach that may be used as a bone cancer detection technique. For many applications, it is crucial to accurately classify and segment bone cancers. Nevertheless, current approaches such as medical imaging techniques often falter because the pictures' non-homogeneous and contrast intensities are insufficient. Support Vector Machine (SVM) classifiers are used in the study to speed up the classification procedure. Through the introduction of a new technique for bone cancer segmentation, the study opens up new research directions in this important field of diagnosis and treatment.

The difficulties in using deep learning models to the diagnosis of bone fractures from X-ray pictures are covered in this thorough review.[15] The study finds that the development and comparison of techniques is hampered by the absence of precise criteria for tasks related to location, detection, classification, and recognition. In order to address this, the review examines and assesses forty current articles, providing accurate definitions for various deep learning tasks. Key findings from each study are compiled to provide a generalized processing framework. The paper also highlights important directions for further investigation, such as improving interpretability, incorporating multimodal clinical data, offering treatment suggestions, and creating sophisticated visualization techniques. All things considered, this work closes the gap in exact task definitions and lays the groundwork for next developments that will enhance deep learning models for bone fracture diagnosis in terms of interpretability, clinical decision support, and visualization methods.

The significance of bones in the human body and the frequent occurrence of bone fractures which are frequently detected by X-ray imaging are the main topics of this study.[16] Efficient diagnostic tools are necessary since traditional fracture detection methods are laborious and prone to mistakes. Broken bone identification is frequently aided by deep learning methods like Deep Neural Networks (DNN). A DNN framework for differentiating between healthy and damaged bones is presented in the study. Nevertheless, the DNN model first suffers from overfitting because of the scarcity of available data. Data augmentation strategies are used to enhance the amount of the dataset in order to address this. The paper evaluates the suggested system's performance through three experiments that make use of Softmax and Adam optimization approaches. Using 5-fold validation, the results demonstrate a high accuracy rate of 99.54% in categorizing both healthy and damaged bones. The effectiveness of the proposed method in diagnosing bone fractures is demonstrated by its accuracy, which beats equivalent existing algorithms.

The purpose of this study [17] was to evaluate the performance of models based on convolutional neural networks (CNNs) for the detection and categorization of maxillofacial fractures in computed tomography (CT) images. A regional trauma hospital provided a dataset of 3407 CT scans, 2407 of which had maxillofacial fractures. The dataset was gathered retrospectively. Multiclass object identification models using Faster R-CNN and YOLOv5, and multiclass image classification models using DenseNet-169 and ResNet-152, were built. While the detection models automatically positioned bounding boxes to identify fracture lines, the classification models classified fractures into frontal, midface, mandibular, and no fracture classes. The best multiclass detection model (Faster R-CNN) had a mean average precision of 0.78 and the best multiclass classification model (DenseNet-169) had an overall accuracy of 0.70, according to performance evaluation on an independent test dataset. To sum up, DenseNet-169 and Faster R-CNN show potential in identifying and categorizing maxillofacial fractures from CT scans.

This retrospective epidemiological study [18] looked at the epidemiological features, risk factors, classification, mechanisms of injury, and early therapy of femur fractures in Somalia. 402 individuals received treatment for femur fractures over a four-year period; 36% of cases were female and 64% were male. 47.7 years was the mean age of the patients. Anatomically, proximal femur fractures were more prevalent, with femur neck fractures predominating. Femur shaft fractures were most commonly caused by gunshots, especially in young males between the ages of 19 and 40. Elderly patients were more likely to suffer falls from standing height. Age groups and genders differed significantly in the mechanisms of damage; older females were more likely to have falls injuries and younger males were more often harmed by firearms. The study emphasizes the high rate of gunshot-related fractures in a nation plagued by protracted conflict, as well as the substantial morbidity and death linked to femur fractures in Somalia. It also highlights the significance of low-energy injuries in the older population.

3 Methodology

For fracture identification and classification, two models are used: a Convolutional Neural Network (CNN) and a Multilayer Perceptron (MLP).

Data Collection and Preprocessing:

This study's dataset comes from the "Bone Break Classification Image Dataset." First, preprocessing is applied to the photos to make sure they are consistent and appropriate for training the model. The TensorFlow 'image_dataset_from_directory' function is used to load the photos and divide them into training and validation sets. The directory structure is used to infer the labels, giving rise to a supervised learning framework. The photos are downsized to a common 256x256 pixel size in order to aid in model convergence and performance. To ensure numerical stability during training, normalization is then used to scale pixel values to the range [0, 1].

Multilayer Perciptron:

Multiple tightly connected layers make up the MLP design, which makes it easier to identify complex patterns and relationships in the input data. A Leaky ReLU activation function, which adds non-linearity to the model and facilitates feature extraction, comes after each dense layer. Each dense layer is followed by batch normalization layers, which standardize the inputs to the layers that come after it. This helps to minimize the effects of covariate shift and speeds up model convergence. Ten units make up the final output layer, which corresponds to the ten fracture classes the dataset contains. Applying a softmax activation function to the output layer creates probability distributions across the classes, allowing for multi-class classification. The stochastic gradient descent version Adam optimizer is used to train the model, and categorical cross-entropy is used as the loss function.

Convolutional Neural Network:

The CNN architecture is especially well-suited for image classification tasks since it is specifically intended to catch local patterns and spatial hierarchies within the input data. The CNN model consists of several convolutional layers, each of which is followed by dropout layers, batch normalization, and leaky ReLU activation. The convolutional layers convolve over the input pictures using learnable filters in order to extract pertinent characteristics at various spatial scales. To do classification, the feature maps are then flattened and run through layers with a high degree of connectivity. The CNN's output layer has ten units with softmax activation for multi-class classification, much like the MLP model. The Adam optimizer and categorical cross-entropy loss function are used to train the model.

Evaluation metrics:

The model was assessed for performance using a variety of measures after training was completed. The two main measures taken into account were accuracy and loss, which provide information about the generalization performance and predictive power of the model. To further visualize the model's classification results and pinpoint any misclassifications, a confusion matrix was created. For each class, precision, recall, and F1-score were computed to assess the model's performance on distinct categories. Ultimately, classification reports were produced in order to give a thorough summary of the model's advantages and disadvantages as well as to summarize its performance across various evaluation metrics.

Classification metrics:

Accuracy: The percentage of correctly identified samples relative to the total number of samples is the measure of accuracy. It offers a general evaluation of the model's ability to predict well.

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

Where:

TP: True Positives

TN: True Negatives

FP: False Positives

FN: False Negatives

Confusion Matrix: Displaying the quantity of true positives, true negatives, false positives, and false negatives, the confusion matrix offers a thorough analysis of the model's predictions. Additional metrics, including recall, precision, and F1-score, can be obtained from the confusion matrix.

Precision: The percentage of true positive predictions among all of the model's positive predictions is measured by precision, which is sometimes referred to as positive predictive value. It shows how well the model can prevent false positives.

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

Where:

TP: True Positives

FP: False Positives

Recall (also called sensitivity): The percentage of true positive predictions among all actual positive samples in the dataset is measured by recall, which is also referred to as sensitivity. It measures how well the model can identify good examples.

F1-Score: This balanced indicator of the model's performance is calculated as the harmonic mean of precision and recall. It is appropriate for imbalanced datasets since it takes into account both false positives and false negatives.

$$F1\ score = 2 \times \frac{Precison \times Recall}{Precision + Recall} \quad (3)$$

4 Results

Several measures, such as loss and accuracy, were used to assess the fracture detection and classification performance of the Multilayer Perceptron (MLP) and Convolutional Neural Network (CNN) models.

Multilayer Perceptron:

During the course of the 16-epoch training procedure, the Multilayer Perceptron (MLP) model showed appreciable gains in accuracy and loss. After the first epoch, the model's accuracy was comparatively poor at initially, reaching about 13.96%, but it progressively improved over the next epochs. By the last epoch, 72.96% of the data was accurate. Similarly, after training, the categorical cross-entropy loss dropped from 2.4601 to 0.8692, demonstrating the model's successful convergence. After testing the model on the validation dataset, the MLP model obtained a good accuracy of 73.21%. This performance indicates the model's potential usefulness in practical applications by showing how effectively it generalizes to new data.

Convolutional Neural Network:

The Convolutional Neural Network (CNN) model demonstrated remarkable efficacy in fracture identification and classification. The CNN model is specifically tailored to capture spatial hierarchies in image data. Over the duration of training, CNN consistently improved its accuracy and loss metrics after 10 epochs of training. The model's accuracy grew over time, peaking at 99.12% in the final epoch after initially increasing to about 16.22% in the first epoch. Simultaneously, the loss dropped from 2.6400 to 0.1165, suggesting that learning was successful. When the CNN model was tested using the validation dataset, it produced an impressive accuracy of 94.64%. This outstanding result highlights how well CNN can identify intricate patterns and spatial correlations in X-ray images, which helps with fracture classification and detection.

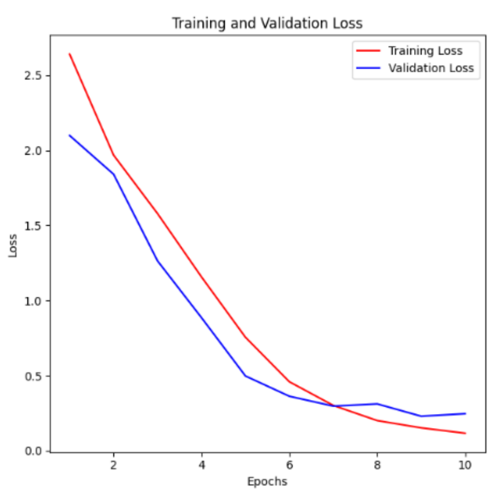

Fig. 1: training and validation loss results.

Fig. 2: training and validation accuracy results.

Comparative analysis:

It is clear from comparing the MLP and CNN models' performances that the CNN performed noticeably better in terms of accuracy and predictive power than the MLP. CNN performed better in fracture identification tasks because of its capacity to extract spatial information and hierarchical characteristics from images.

Acknowledgment

We extend our heartfelt appreciation to the dedicated scientists and researchers whose contributions have been instrumental in the development of this track. Their pioneering work and invaluable insights have laid the foundation upon which our research builds. We are grateful for their innovative ideas, groundbreaking discoveries, and tireless efforts in advancing the field. Their expertise and commitment to excellence have inspired and guided our own endeavors, shaping the trajectory of this project. We deeply admire and acknowledge their significant contributions, which have enriched our understanding and propelled the progress of this track.

5 Conclusion

In order to identify and categorize fractures from X-ray pictures, this study examined the effectiveness of Multilayer Perceptron (MLP) and Convolutional Neural Network (CNN) models. The outcomes demonstrated that, although both models showed encouraging potential, the CNN model outperformed the MLP in terms of accuracy and efficacy when it came to fracture detection tasks. The CNN model performed better by utilizing its capacity to extract intricate patterns and capture spatial hierarchies from images. This suggests that the model has the potential to automate fracture detection procedures and enhance diagnostic precision in clinical settings. These results highlight the value of deep learning approaches in the analysis of medical images and indicate that more study concentrating on improving CNN architectures and model interpretability could result in the creation of reliable computer-aided diagnosis systems with important ramifications for patient care and clinical practice in orthopedics and beyond.

References

[1] T. Sanada, R. Murakami, H. Iwaso, E. Honda, H. Yoshitomi, and M. Inagawa, "Predictive factors for the bone union disorder of intramedullary screw fixation in proximal fifth metatarsal bone fractures," Arch. Orthop. Trauma Surg., vol. 143, no. 4, pp. 1949–1955, 2023, doi: 10.1007/s00402-022-04398-1.

[2] M. Modrák, M. Trebuňová, A. F. Balogová, R. Hudák, and J. Živčák, "Biodegradable Materials for Tissue Engineering: Development, Classification and Current Applications," J. Funct. Biomater., vol. 14, no. 3, 2023, doi: 10.3390/jfb14030159.

[3] R. F. Dash, G. J. Minshall, and I. M. Wright, "Comparison of radiography and CT for the evaluation of third carpal bone slab fractures in Thoroughbred racehorses," Vet. Radiol. \& Ultrasound, vol. 64, no. 4, pp. 661–668, 2023, doi: https://doi.org/10.1111/vru.13255.

[4] O. Bouzeboudja, B. Haddad, A. Taleb-Ahmed, S. Ameur, M. El Hassouni, and R. Jennane, "Multifractal analysis for improved osteoporosis classification," Biomed. Signal Process. Control, vol. 80, p. 104225, 2023, doi: https://doi.org/10.1016/j.bspc.2022.104225.

[5] L.-G. Koepke et al., "Analysis of Three-Dimensional Bone Microarchitecture of the Axis Exposes Pronounced Regional Heterogeneity Associated with Clinical Fracture Patterns," Calcif. Tissue Int., vol. 112, no. 5, pp. 563–572, 2023, doi: 10.1007/s00223-023-01070-7.

[6] Y. Koike et al., "Artificial intelligence-aided lytic spinal bone metastasis classification on CT scans," Int. J. Comput. Assist. Radiol. Surg., vol. 18, no. 10, pp. 1867–1874, 2023, doi: 10.1007/s11548-023-02880-8.

[7] N. Vishwanath et al., "The Role of Antibiotics in Nasal Fractures after Closed Reduction.," Plast. Reconstr. surgery. Glob. open, vol. 11, no. 4, p. e4886, Apr. 2023, doi: 10.1097/GOX.0000000000004886.

[8] C. Li, W. Lin, G. Wu, L. Zhang, and P. Zhang, "Clinical effect of minimally invasive technique of internal fixation device removal in patients with healed long tubular bone fractures.," Pakistan J. Med. Sci., vol. 39, no. 1, pp. 91–95, 2023, doi: 10.12669/pjms.39.1.6967.

[9] S. Razaeian et al., "Predicting the deltoid tuberosity index in proximal humerus fractures using fracture characteristics and patient age: development of the LBQ-PHF score," BMC Musculoskelet. Disord., vol. 24, no. 1, p. 754, 2023, doi: 10.1186/s12891-023-06883-z.

[10] S. Rao Karanam, Y. Srinivas, and S. Chakravarty, "A systematic review on approach and analysis of bone fracture classification," Mater. Today Proc., vol. 80, pp. 2557–2562, 2023, doi: https://doi.org/10.1016/j.matpr.2021.06.408.

[11] K. Dlshad Ahmed and R. Hawezi, "Detection of bone fracture based on machine learning techniques," Meas. Sensors, vol. 27, p. 100723, 2023, doi: https://doi.org/10.1016/j.measen.2023.100723.

[12] M. E. Sahin, "Image processing and machine learning-based bone fracture detection and classification using X-ray images," Int. J. Imaging Syst. Technol., vol. 33, no. 3, pp. 853–865, 2023, doi: https://doi.org/10.1002/ima.22849.

[13] L. Lydia, K. V. Kumar, and G. N. Rao, "Analysis of Advanced Deep Learning Approaches for the Multiple Bone Fracture detection," 2023, doi: 10.21203/rs.3.rs-2995613/v1.

[14] A. George and B. Ayshwarya, "Adaptive Flame based segmentation and classification for bone cancer detection," in 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), 2023, pp. 1–5. doi: 10.1109/ICECONF57129.2023.10083670.

[15] Z. Su, A. Adam, M. F. Nasrudin, M. Ayob, and G. Punganan, "Skeletal Fracture Detection with Deep Learning: A Comprehensive Review," Diagnostics, vol. 13, no. 20, 2023, doi: 10.3390/diagnostics13203245.

[16] L. Visalatchi, B. Dwarakanath, M. Navaneethakrishnan, V. Pandimurugan, L. Sathish Kumar, and A. Gadgil, "Biomaterials Analysis and Deep Learning Outperforms Clinical Experts in Diagnosis and Classification of Human Bone Fractures," in Computational Vision and Bio-Inspired Computing, S. Smys, J. M. R. S. Tavares, and F. Shi, Eds., Singapore: Springer Nature Singapore, 2023, pp. 115–123.

[17] K. Warin, W. Limprasert, S. Suebnukarn, T. Paipongna, P. Jantana, and S. Vicharueang, "Maxillofacial fracture detection and classification in computed tomography images using convolutional neural network-based models," Sci. Rep., vol. 13, no. 1, p. 3434, 2023, doi: 10.1038/s41598-023-30640-w.

[18] Y. B. Ibrahim et al., "Risk factors, classification, and operative choices of femur fractures at a Tertiary Hospital: first report from Somalia," Sci. Rep., vol. 13, no. 1, p. 12847, 2023, doi: 10.1038/s41598-023-39671-9.