https://archtrauma.kaums.ac.ir/
Archives of Trauma Research. 2023; 12(2): 102-111
https://doi.org/10.48307/ATR.2023.176358

Narrative Review Open Access

Artificial intelligence in detecting mandibular fractures: A review of literature

Mohammad Soroush Sehat [1], **Amirali Hariri** [2], **Parisa Soltani** [3,4], **Abbasali Khademi** [5*]

[1] School of Dentistry, Isfahan University of Medical Sciences, Isfahan, Iran
[2] Department of Pharmaceutical Biotechnology, School of Pharmacy and Pharmaceutical Sciences, Isfahan University of Medical Sciences, Isfahan, Iran
[3] Department of Oral and Maxillofacial Radiology, Dental Implants Research Center, Dental Research Institute, School of Dentistry, Isfahan University of Medical Sciences, Isfahan, Iran
[4] Department of Neurosciences, Reproductive and Odontostomatological Sciences, University of Naples "Federico II", Naples, Italy
[5] Dental Research Center, Dental Research Institute, School of Dentistry, Isfahan University of Medical Sciences, Isfahan, Iran

[*] **Corresponding author:** Abbasali Khademi, Dental Research Center, Dental Research Institute, School of Dentistry, Isfahan University of Medical Sciences, Isfahan, Iran. **Email:** a_khademi@dnt.mui.ac.ir

Received: 4 January 2023 **Revised:** 10 March 2023 **Accepted:** 23 July 2023 **e-Published:** 29 July 2023

Abstract

Background: Mandibular fractures are a common trauma in oral and maxillofacial surgery. The accurate diagnosis of these fractures is crucial for successful treatment. However, the interpretation of radiographic scans can be time-consuming and prone to human error. The advent of artificial intelligence (AI), specifically Convolutional Neural Networks (CNNs), has opened up new possibilities for improving the accuracy and efficiency of fracture detection.
Objectives: This review aims to explore the role of AI in detecting mandibular fractures.
Methods: A comprehensive literature search was performed using PubMed, Embase, Web of Science, and Google Scholar databases. Studies were included if they used AI techniques, specifically CNNs or transformers, for the detection of mandibular fractures.
Results: The systematic search yielded 53 studies, with eight studies meeting the inclusion criteria. The AI models across these studies demonstrated a generally high degree of effectiveness in detecting mandibular fractures, with F1 scores ranging from 45% to 100%. Some studies also compared the diagnostic prowess of human clinicians and AI models, with AI models often matching or surpassing human performance.
Conclusion: The application of AI in detecting mandibular fractures represents a promising avenue of research. AI models have the potential to reduce the workload of radiologists, improve the efficiency of fracture detection, and lead to faster diagnosis and treatment. However, further research is needed to validate these findings in larger and more diverse datasets and to address challenges such as the interpretability of AI algorithms and the availability of large, well-annotated datasets.

Keywords: Artificial Intelligence, Convolutional Neural Networks, Mandibular Fractures, Fracture Detection, Radiology.

Introduction

Mandibular fractures, a common trauma in oral and maxillofacial surgery, account for a significant proportion, 59.6% of facial traumas.[1] These fractures are often the result of various incidents such as vehicle accidents, assaults, sports injuries, and falls.[2] The anatomical areas typically affected include the symphysis/parasymphysis (30-50%), body/horizontal branch (21-36%), and angle of the mandible (15-26%).[3] The complexity and severity of these fractures necessitate a range of treatment options, from nonoperative management to open reduction with internal fixation.[4]

The accurate diagnosis of mandibular fractures is crucial for the successful re-establishment of occlusion, function, and facial aesthetics.[5] Panoramic radiographs (PR) are frequently employed as the first-level imaging technique in facial trauma patients.[6] However, the limitations of PR, such as the lack of three-dimensionality and potential artifacts in regions of interest, can impede the accurate detection of fractures.[7] Computed tomography (CT) scans, which offer a more detailed and three-dimensional view, are often used to complement PR for a more accurate diagnosis.[8] Despite the advancements in imaging technology, the interpretation of these scans can be time-consuming and prone to human error, especially in busy clinical settings like emergency departments.[9]

Misdiagnosis of mandible fractures can lead to a variety of complications, including chronic pain, malocclusion, infection, and even facial deformity. It is crucial to accurately diagnose and treat these injuries to prevent such adverse outcomes.[1] In recent years, the advent of artificial intelligence (AI) has opened up new possibilities for improving the accuracy and efficiency of fracture detection.[10] In the realm of AI, Convolutional Neural Networks (CNNs), a subtype of deep learning models, have emerged as potent tools for interpreting two-dimensional and also three-dimensional medical images.[11-14] Derived from the neural networks concept, CNNs incorporate the unique feature of convolution layers that allow these networks to process input data in a grid format, ideally suited for image data.[15] They have been shown to demonstrate outstanding performance in a variety of complex perceptual tasks, including image segmentation, object detection, and classification. These algorithms have demonstrated exceptional performance in interpreting medical images, including X-rays, CT scans, MRI scans, and ultrasound images, enabling accurate diagnosis and detection of various conditions, often matching or nearing the performance of human experts.[16, 17]

However, the performance of AI systems like CNNs and transformers is highly dependent on the dataset, the hyperparameters, and the architecture itself. [18] Furthermore, certain fracture locations, such as the condyle region, present unique challenges due to their complex fracture shapes and the potential for superimposition with other structures especially in panoramic images.[3] Despite these challenges, AI systems have the potential to serve as a valuable tool in training physicians and experts to better evaluate panoramic radiographs and CT scans for fractures, thereby reducing diagnostic error and the need for advanced imaging in emergency departments.[19]

While there have been individual studies examining the role of AI in detecting mandibular fractures, to the best of our knowledge, a comprehensive review synthesizing these findings has not been conducted. This gap in the literature underscores the need for a systematic review to collate and analyze the existing evidence.

Objectives

Therefore, this review aims to explore the role of artificial intelligence in detecting mandibular fractures, with a focus on the challenges and potential solutions in this emerging field. The aim of the present study was to provide a comprehensive overview of the current state of AI applications in mandibular fracture detection, including the use of different imaging techniques like PR and CT/CBCT, and to highlight areas for future research and development.

Methods

This review was conducted following a systematic approach to ensure a comprehensive and unbiased assessment of the literature. The methodology was designed to identify, select, and extract data from these studies.

Search Strategy

A comprehensive literature search was performed using PubMed, Embase, Web of Science, and Scopus databases from their inception until June 2023. The search strategy included a combination of keywords and MeSH terms related to "Artificial Intelligence", "Mandibular Fractures", "Detection", "Diagnosis", "Convolutional Neural Networks", “Machine Learning”, "Transformers", "Panoramic Radiographs", and "Computed Tomography", “Cone-Beam Computed Tomography” and “Extraoral Radiography”. The search was limited to articles published in English. Additionally, the reference lists of included studies were manually searched to identify any additional relevant studies.

Study Selection

Studies were imported to Mendeley reference manager (Elsevier, Amsterdam, Netherlands). Two independent reviewers (A H and M S) screened the titles and abstracts of the identified studies. Full-text articles were retrieved for those that met the inclusion criteria or where there was uncertainty. Any disagreements between the reviewers were resolved through consultation with a third reviewer (P S).

Inclusion Criteria

Studies were included if they: 1) were original research articles; 2) used artificial intelligence techniques for the detection of mandibular fractures; 3) utilized imaging techniques such as panoramic radiographs or computed tomography; and 4) reported performance measures such as sensitivity, specificity, area under the curve (AUC), or F1 score. Articles were excluded if they were in vitro studies, animal studies, reports from ‘grey literature’ (conference abstracts, unpublished studies), letters to the editors, and reviews.

Data Extraction

Data were extracted from the included studies by two reviewers (A H and M S) using a standardized data extraction form. The extracted information included study characteristics (authors, year of publication, and country), details of the AI techniques used, type of imaging technique used, number of images or patients, and key findings.

Data Synthesis and Analysis

A narrative synthesis of the findings from the included studies was conducted. Due to the expected heterogeneity in AI techniques, imaging modalities, and study populations, a meta-analysis was not planned. The performance measures reported in the studies were summarized and compared. The challenges and potential solutions identified in the studies were also synthesized.

Results

The systematic search of the literature yielded 53 studies for initial screening. After removing duplicates and screening titles and abstracts, 11 studies were selected for full-text review. Following the full-text review, eight studies met the inclusion criteria and were included in the final analysis. The data extracted from the selected studies are presented in Table 1.

The included studies were published between 2021 and 2023 and originated from various countries, including Korea, Japan, China, Iran, Germany, and Thailand, reflecting a global interest in the application of artificial intelligence in detecting mandibular fractures.

The studies varied in terms of study design, AI model used, imaging technique used, and sample size. The majority of the studies used convolutional neural networks (CNNs) as the AI model and utilized computed tomography (CT) or panoramic radiography (PR) as the imaging technique. Among the studies using CT scans, one adopted axial view images, and another generated panoramic images from CT scans. The size of datasets varied considerably, ranging from 190 (a low range) to 1624 (a high range).

In terms of performance, the AI models across these studies demonstrated a generally high degree of effectiveness in detecting mandibular fractures. The performance of these models was often evaluated using the F1 score, a critical metric in AI studies that combines precision and recall. F1 score is a measure that combines precision and recall, two critical parameters in medical diagnostics and a higher F1 indicates a better performance of a model.[20] Precision is the proportion of true positive outcomes among all outcomes predicted as positive, while recall (also known as sensitivity) is the proportion of true positives correctly identified out of all actual positive cases. The F1 score is calculated using the formula 2 * (Precision * Recall) / (Precision + Recall). This score becomes especially crucial in medical contexts where both false negatives (missed fractures) and false positives (unnecessary treatment) can have severe implications. [20] Across the studies, the reported F1 scores ranged from 45% to 100%.

Two of the eight studies drew a comparison between the diagnostic prowess of human clinicians and AI models. Warin et al. [21] compared the sensitivity and specificity of five Oral and Maxillofacial Surgeons (OMFS) against those of DenseNet and ResNet AI models. The OMFS demonstrated a sensitivity of 88.32% and a specificity of 96.46%. In contrast, the DenseNet model achieved perfect scores in both metrics, and the ResNet model scored 100% in sensitivity and 99% in specificity. Shahnavazi et al. [22] set up a similar comparison involving five general dentists and an AI model, with the dentists achieving a sensitivity of 82.2% and a specificity of 92.2%, while the AI model demonstrated a perfect sensitivity score of 100% but a slightly lower specificity of 83.3%.

Table 1. Studies utilizing AI in Detecting Mandibular Fractures

	Country	X-ray Image	Dataset Size and Source	Annotation	Models	F1 score	Human Comparison
Warin, 2023 [23]	Thailand	Ct: axial view	746 Trauma Hospital	5 OMFS	yolov5	0.81	No
					Faster R	0.83	
					DenseNet169	0.56	
					ResNet152	0.45	
Shahnavazi, 2023 [22]	Iran	Panoramic	190 General Hospital, Private radiology	2 OMFR	U-net Faster R Resnet101	0.91	Yes, 5 general dentists sensitivity of 82.2 and specificity of

			office and Internet				92.2 compared to 100 and 83.3 in Ai
Wang, 2022 [24]	China	Ct: generated Panoramic	686 Dental Hospital	3 OMFS	U-net Resnet50	0.95	No
Warin, 2022 [21]	Thailand	Panoramic	855 Trauma Hospital	3 OMFS 1 OMFR	yolov5	89.07	Yes, 5 OMFS sensitivity of 88.32 and specificity of 96.46 compared to 100 and 100 in Densnet and 100 and 99 in Resnet
					Faster R	90.67	
					DenseNet169	100	
					ResNet50	100	
Son, 2022 [25]	Korea	Panoramic	360 Not mentioned	Not mentioned	YOLOv4 LAT/U-Net	0.908	No
					U-net	0.83	
					YOLOv4	0.844	
					YOLOv4 LAT	0.875	
					Mask RCNN	0.63	
Vinayahalingam, 2022 [26]	Germany	Panoramic	1624 Dental hospital	3 OMFS	Faster R-CNN with Swin-Transformer	0.947	No
Nishiyama, 2021 [27]	Japan	Cropped Panoramic	200 Dental hospital and General hospital	Medical reports	Alexnet	0.84	No
Son, 2021 [28]	Korea	Panoramic	420 Not mentioned	Not mentioned	Yolov4 Lat 6class	0.87	No

Discussion

The detection of mandibular fractures is a critical aspect of diagnosing maxillofacial trauma,[5] and the advent of AI has brought about significant advancements in this field. The development of an AI model for this purpose involves several key stages, including Data Annotation, Augmentation, Training, and Testing.[29] Each of these stages plays a crucial role in the performance of the model, and the reviewed literature presents various methods to optimize these processes.

Data Annotation is the process of labeling the data, which in the context of mandibular fracture detection, involves marking the fractures on the radiographic images. The importance of good annotation cannot be overstated, as it forms the foundation for training AI models; high-quality, accurate annotations ensure that the model learns correctly, thereby significantly improving its ability to detect and classify mandibular fractures effectively.[30] In the reviewed studies, the number of annotators varied from only using medical reports[27] to five oral and maxillofacial surgeons, as seen in the Warin et al., study.[21] The list of annotators is available in table 1. There are different ways to annotate the data, from simply indicating whether there is a fracture (classification models such as DenseNet and ResNet) to marking the exact location of the fracture on the x-ray image with bounding boxes (object detection models such as YOLO and Faster R). Some studies even used fracture lines for annotation, which are particularly useful in image segmentation models like U-Net. However, the choice between bounding boxes and annotation lines depends on the specific requirements of the study and the nature of the fractures being analyzed. For instance, in cases of condyle fractures that exhibit signs of dislocation in X-ray images but lack a visible fracture line, the use of image segmentation models like U-Net may be limited. In such scenarios, alternative annotation methods may be more effective.[25]

Data augmentation is another critical step in developing an AI model. It refers to the process of artificially expanding the size and diversity of a dataset to improve the performance and generalizability of machine learning models.[31] Various augmentation methods have been proposed in the literature, ranging from simple techniques to more complex ones. Simple methods, such as random cropping, color jittering, affine transformations, Gaussian noise addition, and random horizontal flipping, are often used, especially to increase the size of smaller datasets. These techniques introduce variability into the dataset, thereby helping to prevent overfitting and improve the model's ability to generalize to unseen data.[22]

More complex methods are such as gamma modulation, luminance adaptation transform, and extended multi-anchor boxes. Luminance adaptation transforms, including MLAT and SLAT, adjust the brightness and contrast of images to better highlight features of interest. These methods have been shown to be effective in some studies.[25, 28] For instance, radiographs often have varying levels of brightness and contrast, with some areas appearing dark or saturated. These variations can obscure fractures and reduce the performance of fracture detection models. However, if suitable image processing techniques, such as local tone improvement processing, are applied to the radiographs, fractures may be more clearly revealed.

After annotation and augmentation, the images are standardized to specific resolutions, ranging from 224x224 to 608x608 pixels. The impact of image resolution on model accuracy is a topic of ongoing debate in the field of computer vision. While higher resolution images generally provide more detailed information, which could potentially enhance model accuracy, they also require more computational resources and can lead to longer training times. Therefore, the optimal resolution may depend on the specific application and the available computational resources. In the context of literature, it is worth noting the work of Nishiyama et al.,[27] who utilized a resolution of 900x900 pixels. However, it is important to consider that, in their study, the actual size of the image was relatively small due to only cropping the condyle area.

Table 2. Classification of AI models used in mandibular fracture detection based on their functionalities: object detection, image classification, and image segmentation

Object Detection Models	Image Classification Models	Image Segmentation Models
These models are generally used for detecting the presence and location of multiple objects within an image.	These models are used for classifying an image into one of several pre-defined classes.	These models are used for image segmentation tasks, which involve dividing the image into multiple regions that correspond to objects or parts of objects.
YOLO (You Only Look Once) YOLO is a real-time object detection system that applies a single neural network to an entire image in one pass, predicting object locations and classifications simultaneously. It is known for its speed and efficiency.	DenseNet (Densely Connected Convolutional Networks) DenseNet operates like a well-connected team, where each member (or layer in the network) communicates directly with every other member. This direct connection promotes a more efficient exchange of information, enhancing the model's ability to learn from the data.	U-Net U-Net is a convolutional neural network used for examining each pixel of an image and determining what it represents. It has a unique encoder-decoder structure that is particularly effective for medical imaging applications, where an accurate understanding of each part of an image can be vital.
R-CNN (Regions with Convolutional Neural Networks) R-CNN is an object detection method that first identifies potential object regions, and then classifies these using a convolutional neural network (CNN). Although effective, it can be computationally intensive and slow.	ResNet (Residual Networks) ResNet uses a clever shortcut system to learn from complex data. It is like a traveler who knows when to take the main route for detailed views and when to take shortcuts to reach the destination more efficiently. This method greatly improves the efficiency and performance of the model in learning complex tasks.	
	AlexNet AlexNet is a pioneering convolutional neural network that was a game-changer in the realm of image classification. It is simpler and less computationally demanding than later models like ResNet and DenseNet, making it ideal for resource-limited settings.	

Following the preparation of the dataset, the next step is model training. A significant finding from the literature review is the high accuracy of AI algorithms in identifying mandibular fractures. Various methods have been proposed for diagnosing these fractures. Table 2 describes AI models used in mandibular fracture detection based on their primary functionalities.

The highest reported results were found in the Warin et al., study,[21] which achieved a perfect F1 score of 100%. However, it should be noted that these results were obtained from binary classification models, ResNet and DenseNet, which only indicate the presence or absence of a fracture, without providing any information about the fracture's location. In contrast, the study conducted by Nishiyama et al.,[27] manually cropped the condyle area to specifically diagnose fractures. The accuracy achieved was 0.84, which is acceptable considering the small dataset and the use of a relatively weaker model, AlexNet. AlexNet, while pioneering, is considered less powerful compared to newer models due to its simpler architecture and fewer layers, which may limit its ability to capture complex features.[32]

To address the need for automatic fracture localization, object detection models have been suggested. These models, such as Fast R-CNN and YOLO, detect and delineate the fracture area in bounding boxes and classify the location. While Fast R-CNN operates in two stages and takes more time, it often yields higher accuracy. YOLO, on the other hand, performs detection and classification simultaneously, offering a speed advantage.[21] Image segmentation models like U-Net have also been employed. U-Net is not an object detection model per se, but it detects fractures as lines on the label during training. However, it can be challenging to label dislocated fractures, such as condyle fractures, using this approach.[25]

Son et al. proposed a combination of YOLO and U-Net and included an auxiliary segmentation model to remove the tooth area in panoramic images, thereby reducing false fracture detections.[25] Another innovative method, suggested by Shahbazi et al. and Wang et al., involved separating the mandibular anatomy using segmentation models like U-Net, and then determining the presence of a fracture with a binary classification model. This two-step approach allows for detailed anatomical understanding and precise fracture detection.[22,24]

The performance of the AI models can vary significantly based on factors such as the dataset used, the architecture of the model, and the chosen hyperparameters, making direct comparisons challenging.[33] However, some studies have utilized multiple AI architectures on the same datasets, allowing us for some comparison.

Son et al.,[25] presented a comprehensive comparison among YOLOv4, Masked R-CNN, and U-Net models. The YOLOv4 LAT/U-Net hybrid model achieved the highest F1 score of 0.908, outperforming the other models. This was followed by YOLOv4 LAT with an F1 score of 0.875, then YOLOv4 with a score of 0.844, and U-net with an F1 score of 0.83. Meanwhile, Mask RCNN achieved the lowest F1 score among the compared models, with a score of 0.63. This comparison further illustrates the importance of model selection, showing that hybrid models may offer improved performance.

In the study conducted by Warin et al.,[23] the performances of classification models, ResNet-152 and DenseNet-169 were evaluated using axial view CT scans. Though both models exhibited modest performance, DenseNet slightly surpassed ResNet, with an F1 score of 0.56 as compared to 0.45. In the same study, the performances of the object detection models, YOLOv5 and Faster R-CNN models were also assessed. These models exhibited higher performance than the classification models mentioned previously, recording F1 scores of 0.81 and 0.83, respectively.

Another study by Warin et al.,[21] presented a comparative analysis between DenseNet-121 and ResNet-50, using panoramic images for evaluation. Contrary to the previous study, both models here achieved perfect F1 scores of 100. Furthermore, the same study compared the performances of YOLOv5 and Faster R-CNN, yielding F1 scores of 89.07 and 90.67, respectively. Interestingly, in this latter case, classification models outperformed the object detection models. This enhanced performance may be attributed to a multitude of factors such as the distinct nature of the imaging technique used or the specific architecture of the models. These findings underscore the dynamic potential of AI models and the need for tailored model selection based on specific tasks and conditions.

Two studies[21,22] have reported that AI algorithms can achieve performance levels comparable to, or even surpassing, those of human experts in the detection of mandibular fractures. This is a significant advancement as it suggests that AI could potentially reduce the workload of radiologists and improve the efficiency of fracture detection. Furthermore, the use of AI could potentially lead to faster diagnosis and treatment, improving patient outcomes. However, it is important to note that these results were obtained under specific conditions and may not generalize to all clinical settings. Therefore, further research is needed to validate these findings in larger and more diverse datasets.

The use of transformer models in the detection of mandibular fractures represents a promising avenue of research, as demonstrated in the study by Vinayahalingam et al.[26] Transformers, originally developed for natural language processing tasks, have shown great potential in various fields, including medical imaging. Transformers are a type of model that operate on the principle of self-attention, allowing them to weigh and prioritize different parts of input when making predictions.[34] Unlike Convolutional Neural Networks (CNNs), which process data in a hierarchical manner and often require a fixed input size, transformers can handle inputs of varying sizes and consider the entire context of the input at once. This makes them particularly suited to tasks where the relationship between different parts of the input is important, such as in the detection and localization of fractures in medical images.[35] In Vinayahalingam's study, a transformer model was trained to detect mandibular fractures, effectively replacing the traditional CNN-based approach. Resulting in an F1 score of 94.7 that showed that the transformer was able to accurately identify and localize fractures, demonstrating the potential of this approach. However, the use of transformers in medical imaging is still a relatively new field and further research is needed to fully understand their potential and limitations.

Among various imaging modalities employed for mandibular fracture detection, panoramic radiographs are the most commonly used due to their wide coverage of the oral and maxillofacial region and relative ease of acquisition.[6] However, some studies have utilized different X-ray views of the mandible. For instance, Warin et al.,[23] employed axial views of CT scans for the detection of mandibular fractures. Similarly, Wang et al.,[24] used panoramic images generated from CT scans. While these represent different views of the mandible, the basic principle for detection remains similar across these modalities. It is important to note, however, that in the case of panoramic images generated from CT scans, there is no superimposition of structures, which is a characteristic feature of traditional panoramic images.[36] This lack of superimposition may limit the generalizability of models trained on these generated panoramic images when applied to normal panoramic images. Therefore, while these alternative imaging modalities provide valuable additional perspectives, their differences from conventional panoramic images must be considered.

Choosing the most suitable method from the models mentioned in the article requires careful consideration of various factors. Firstly, the specific clinical requirements and objectives should be considered. Different methods may excel in different aspects, such as fracture detection, localization accuracy, or the ability to handle specific imaging modalities. Understanding the clinical needs and constraints will help identify the method that aligns best with the desired outcomes. Secondly, the available resources and computational capabilities should be considered. Some methods may require significant computational power or specialized hardware, which may not be feasible in certain clinical settings. Evaluating the available resources will help ensure practical implementation. Additionally, the type and quality of the available imaging data should be assessed. Some methods may perform better on certain types of images or require specific data preprocessing techniques. Evaluating the compatibility between the method and the imaging data will help determine its suitability. Lastly, the expertise of the users should be considered. Some methods may require advanced knowledge and skills to operate effectively. Considering the expertise of the clinical staff and their familiarity with different AI models will facilitate the successful implementation and utilization of the chosen method. By carefully evaluating these factors, healthcare professionals can select the most suitable method that addresses their clinical needs while considering the available resources and expertise.

A crucial aspect to consider is the interpretability of AI algorithms. While these algorithms can make accurate predictions, their decision-making processes are often opaque, which is a challenge known as the "black box". This lack of transparency can limit the trust and acceptance of AI among clinicians, who are accustomed to evidence-based practice and may be hesitant to rely on a tool whose workings they do not fully understand.[37,38] Furthermore, this opacity makes it difficult to troubleshoot and refine the model when it makes errors, as the underlying cause of the error is not clear.[39] Several solutions have been proposed to address this issue, including the development of model-agnostic interpretation methods, such as LIME (Local Interpretable Model-Agnostic Explanations), and the use of attention mechanisms, which can highlight the parts of the input that the model considers most important in making a decision.[40] However, the interpretability of AI algorithms is a critical area for future research.

One of the primary challenges in the application of AI for detecting mandibular fractures lies in the dataset. The quality, diversity, and size of the dataset used to train the AI model significantly influence its performance. [41]

However, the process of accurately annotating images for training these models is labor-intensive and requires expert knowledge, which can limit the availability of large, well-annotated datasets.[42] Multicenter datasets, which are collected from multiple institutions, offer a broader range of patient demographics, imaging protocols, and disease presentations, thereby enhancing the generalizability of the AI models trained on them. Large datasets allow the model to learn from a wider variety of cases, reducing the likelihood of overfitting and improving the model's ability to generalize to new cases.

Overfitting is a common problem in machine learning where a model performs well on the training data but poorly on new, unseen data.[43] Overfitting typically occurs when a model is too complex relative to the amount and noise level of the training data. It learns the noise in the training data, mistaking it for useful information, and as a result, performs poorly on new data.[44] This is a particular concern in medical imaging, where datasets are often small relative to the complexity of the task, and the data can be noisy due to variations in imaging techniques and patient anatomy.[45]

External validity, or the extent to which the results of a study can be generalized to other situations and populations, is a critical consideration in the evaluation of AI models. However, it is often overlooked in the literature, leading to an overestimation of model performance.[46] Among the reviewed literature only Nishiyama's study conducted external validation. [27] Upon testing their model on an external dataset, they reported a decrease in performance. The model's performance accuracy dropped from 80.4 to 59, indicating that the model did not generalize well to new, unseen data. This underscores the importance of testing AI models on external datasets, which are separate from the data used for training and internal validation. Such external validation provides a more realistic estimate of how the model will perform in real-world settings.

Conclusions

In conclusion, the application of artificial intelligence (AI) in detecting mandibular fractures shows promise in improving the accuracy and efficiency of fracture detection. The reviewed studies demonstrate that AI models, particularly Convolutional Neural Networks (CNNs) and transformers, can achieve high levels of effectiveness in detecting mandibular fractures, with F1 scores ranging from 45% to 100%. In some cases, AI models match or surpass the diagnostic performance of human clinicians. The use of AI has the potential to reduce the workload of radiologists, improve efficiency in fracture detection, and lead to faster diagnosis and treatment. However, further research is needed to validate these findings in larger and more diverse datasets and to address challenges, such as the interpretability of AI algorithms and the availability of well-annotated datasets. Moreover, efforts should be made to optimize data annotation, augmentation, and model selection processes to enhance the accuracy and generalizability of AI models. Despite these challenges, AI represents a promising tool that can assist clinicians in improving the diagnosis and management of mandibular fractures, ultimately leading to better patient care and outcomes.

Acknowledgment

The authors take this opportunity to thank the Department of Clinical Biochemistry, Faculty of Medicine, Rafsanjan University of Medical Sciences, Rafsanjan, Iran for their financial and technical support.

Competing interests

The authors declare that they have no competing interests.

Abbreviations

Coronavirus disease 2019: COVID-19;

Severe acute respiratory syndrome coronavirus 2: SARS-CoV-2;

World Health Organization: WHO

Public health emergency of international concern: PHEIC

Authors' contributions

All authors read and approved the final manuscript. All authors take responsibility for the integrity of the data and the accuracy of the data analysis.

Funding

None.

Role of the funding source

None.

Availability of data and materials

The data used in this study are available from the corresponding author on request.

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki. Institutional Review Board approval (code: IR.RUMS.REC.1396.119) was obtained (April 2020). The present study did not interfere with the process of diagnosis and treatment of patients and all participants signed an informed consent form.

Consent for publication

By submitting this document, the authors declare their consent for the final accepted version of the manuscript to be considered for publication.

References

1. Pickrell BB, Serebrakian AT, Maricevich RS. Mandible Fractures. Semin Plast Surg. 2017;31(2):100-7. doi:10.1055/s-0037-1601374 PMid:28496390 PMCid:PMC5423793
2. Zamboni RA, Wagner JCB, Volkweis MR, Gerhardt EL, Buchmann EM, Bavaresco CS. Epidemiological study of facial fractures at the Oral and Maxillofacial Surgery Service, Santa Casa de Misericordia Hospital Complex, Porto Alegre - RS - Brazil. Rev Col Bras Cir. 2017;44(5):491-7. doi:10.1590/0100-69912017005011 PMid:29019579
3. Nardi C, Vignoli C, Pietragalla M, Tonelli P, Calistri L, Franchi L, et al. Imaging of mandibular fractures: a pictorial review. Insights Imaging. 2020;11(1):30. doi:10.1186/s13244-020-0837-0 PMid:32076873 PMCid:PMC7031477
4. Amrish B, Virendra S, Ruchi S. Management of Mandibular Fractures. In: Mohammad Hosein Kalantar M, editor. A Textbook of Advanced Oral and Maxillofacial Surgery. Rijeka: IntechOpen; 2013. p. Ch. 14.
5. Panesar K, Susarla SM. Mandibular Fractures: Diagnosis and Management. Semin Plast Surg. 2021;35(4):238-49. doi:10.1055/s-0041-1735818 PMid:34819805 PMCid:PMC8604616
6. Mehta N, Butala P, Bernstein MP. The imaging of maxillofacial trauma and its pertinence to surgical intervention. Radiol Clin North Am. 2012;50(1):43-57. doi:10.1016/j.rcl.2011.08.005 PMid:22099486
7. Sheng K. Radiological investigation of acute mandibular injury. Natl J Maxillofac Surg. 2022;13(2).doi:10.4103/njms.NJMS_27_19 PMid:36051802 PMCid:PMC9426694
8. Reiter MJ, Schwope RB, Theler JM. Postoperative CT of the Mandible Following Trauma: Review of Normal Appearances and Common Complications. Acad Radiol. 2019;26(5):686-98. doi:10.1016/j.acra.2018.06.009 PMid:30072290
9. Pinto A, Reginelli A, Pinto F, Lo Re G, Midiri F, Muzj C, et al. Errors in imaging patients in the emergency setting. Br J Radiol. 2016;89(1061):20150914. doi:10.1259/bjr.20150914 PMid:26838955 PMCid:PMC4985468
10. Canoni-Meynet L, Verdot P, Danner A, Calame P, Aubry S. Added value of an artificial intelligence solution for fracture detection in the radiologist's daily trauma emergencies workflow. Diagn Interv Imaging. 2022;103(12):594-600. doi:10.1016/j.diii.2022.06.004 PMid:35780054
11. Chen X, Wang X, Zhang K, Fung K-M, Thai TC, Moore K, et al. Recent advances and clinical applications of deep learning in medical image analysis. Med Image Anal. 2022;79:102444. doi:10.1016/j.media.2022.102444 PMid:35472844 PMCid:PMC9156578
12. Krithika Alias AnbuDevi M, Suganthi K. Review of semantic segmentation of medical images using modified architectures of UNET. Diagnostics (Basel). 2022;12(12). doi:10.3390/diagnostics12123064 PMid:36553071 PMCid:PMC9777361
13. Alskaf E, Dutta U, Scannell CM, Chiribiri A. Deep learning applications in coronary anatomy imaging: a systematic review and meta-analysis. J Med Artif Intell. 2022;5:11. doi:10.21037/jmai-22-36 PMid:36861064 PMCid:PMC7614252
14. Koh D-M, Papanikolaou N, Bick U, Illing R, Kahn CE, Kalpathi-Cramer J, et al. Artificial intelligence and machine learning in cancer imaging. Commun Med. 2022;2(1):133. doi:10.1038/s43856-022-00199-0 PMid:36310650 PMCid:PMC9613681
15. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights Imaging. 2018;9(4):611-29. doi:10.1007/s13244-018-0639-9 PMid:29934920 PMCid:PMC6108980
16. Anaya-Isaza A, Mera-Jiménez L, Zequera-Diaz M. An overview of deep learning in medical imaging. Inform Med Unlocked. 2021;26:100723. doi:10.1016/j.imu.2021.100723
17. Wang T, Lei Y, Fu Y, Wynne JF, Curran WJ, Liu T, et al. A review on medical imaging synthesis using deep learning and its clinical applications. J Appl Clin Med Phys. 2021;22(1):11-36. doi:10.1002/acm2.13121 PMid:33305538 PMCid:PMC7856512
18. Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. Electron Mark. 2021;31(3):685-95. doi:10.1007/s12525-021-00475-2
19. Kumar Y, Koul A, Singla R, Ijaz MF. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. J Ambient Intell Humaniz Comput. 2023;14(7):8459-86. doi:10.1007/s12652-021-03612-z PMid:35039756 PMCid:PMC8754556
20. Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, et al. On evaluation metrics for medical applications of artificial intelligence. Sci Rep. 2022;12(1):5979. doi:10.1038/s41598-022-09954-8 PMid:35395867 PMCid:PMC8993826
21. Warin K, Limprasert W, Suebnukarn S, Inglam S, Jantana P, Vicharueang S. Assessment of deep convolutional neural network models for mandibular fracture detection in panoramic radiographs. Int J Oral Maxillofac Surg. 2022;51(11):1488-94. doi:10.1016/j.ijom.2022.03.056 PMid:35397969
22. Shahnavazi M, Mohamadrahimi H. The application of artificial neural networks in the detection of mandibular fractures using panoramic radiography. Den Res J. 2023;20:27.
23. Warin K, Limprasert W, Suebnukarn S, Paipongna T, Jantana P, Vicharueang S. Maxillofacial fracture detection and classification in computed tomography images using convolutional neural network-based models. Sci Rep. 2023;13(1):3434. doi:10.1038/s41598-023-30640-w PMid:36859660 PMCid:PMC9978019
24. Wang X, Xu Z, Tong Y, Xia L, Jie B, Ding P, et al. Detection and classification of mandibular fracture on CT scan using deep convolutional neural network. Clin Oral Invest. 2022; 26(6):4593-601. doi:10.1007/s00784-022-04427-8 PMid:35218428
25. Son DM, Yoon YA, Kwon HJ, Lee SH. Combined Deep Learning Techniques for Mandibular Fracture Diagnosis Assistance. Life. 2022;12(11). doi:10.3390/life12111711 PMid:36362866 PMCid:PMC9697461
26. Vinayahalingam S, van Nistelrooij N, van Ginneken B, Bressem K, Tröltzsch D, Heiland M, et al. Detection of mandibular fractures on panoramic radiographs using deep learning. Sci Rep. 2022; 12 (1):19596. doi:10.1038/s41598-022-23445-w PMid:36379971

PMCid:PMC9666517
27. Nishiyama M, Ishibashi K, Ariji Y, Fukuda M, Nishiyama W, Umemura M, et al. Performance of deep learning models constructed using panoramic radiographs from two hospitals to diagnose fractures of the mandibular condyle. Dentomaxillofac Radiol. 2021;50(7):20200611. doi:10.1259/dmfr.20200611 PMid:33769840 PMCid:PMC8474128
28. Son DM, Yoon YA, Kwon HJ, An CH, Lee SH. Automatic detection of mandibular fractures in panoramic radiographs using deep learning. Diagnostics. 2021;11(6):1-19. doi:10.3390/diagnostics11060933 PMid:34067462 PMCid:PMC8224557
29. Sarker IH. AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems. SN Comput Sci. 2022;3(2):158. doi:10.1007/s42979-022-01043-x PMid:35194580 PMCid:PMC8830986
30. Paullada A, Raji ID, Bender EM, Denton E, Hanna A. Data and its (dis)contents: A survey of dataset development and use in machine learning research. Patterns. 2021;2(11):100336. doi:10.1016/j.patter.2021.100336 PMid:34820643 PMCid:PMC8600147
31. Mumuni A, Mumuni F. Data augmentation: A comprehensive survey of modern approaches. Array. 2022;16:100258. doi:10.1016/j.array.2022.100258
32. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data. 2021;8(1):53. doi:10.1186/s40537-021-00444-8 PMid:33816053 PMCid:PMC8010506
33. Wu J, Chen X-Y, Zhang H, Xiong L-D, Lei H, Deng S-H. Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimizationb. J Electronic Sci Technol. 2019;17(1):26-40.
34. Li J, Chen J, Tang Y, Wang C, Landman BA, Zhou SK. Transforming medical imaging with Transformers? A comparative review of key properties, current progresses, and future perspectives. Med Image Anal. 2023;85:102762. doi:10.1016/j.media.2023.102762 PMid:36738650
35. He K, Gan C, Li Z, Rekik I, Yin Z, Ji W, et al. Transformers in medical image analysis. Intell Med. 2023;3(1):59-78. doi:10.1016/j.imed.2022.07.002
36. Papakosta TK, Savva AD, Economopoulos TL, Matsopoulos GK, Gröhndal HG. An automatic panoramic image reconstruction scheme from dental computed tomography images. Dentomaxillofac Radiol. 2017;46(4):20160225. doi:10.1259/dmfr.20160225 PMid:28112548 PMCid:PMC5594999
37. Juan Manuel D, Karin Rolanda J. Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. J Med Ethics. 2021;47(5):329.
38. von Eschenbach WJ. Transparency and the Black Box Problem: Why We Do Not Trust AI. Philos Technol. 2021; 34 (4): 1607-22. doi:10.1007/s13347-021-00477-0
39. Fraternali P, Milani F, Torres RN, Zangrando N. Black-box error diagnosis in Deep Neural Networks for computer vision: a survey of tools. Neural Comput Appl. 2023;35(4):3041-62. doi:10.1007/s00521-022-08100-9
40. Srinivasu PN, Sandhya N, Jhaveri RH, Raut R. From Blackbox to Explainable AI in Healthcare: Existing Tools and Case Studies. Mobile Information Systems. 2022;2022:8167821. doi:10.1155/2022/8167821
41. Hung KF, Ai QYH, Leung YY, Yeung AWK. Potential and impact of artificial intelligence algorithms in dento-maxillofacial radiology. Clin Oral Invest. 2022;26(9):5535-55. doi:10.1007/s00784-022-04477-y PMid:35438326
42. Tajbakhsh N, Jeyaseelan L, Li Q, Chiang JN, Wu Z, Ding X. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. Med Image Anal. 2020; 63:101693. doi:10.1016/j.media.2020.101693 PMid:32289663
43. Ying X. An Overview of Overfitting and its Solutions. Journal of Physics: Conference Series. 2019;1168(2):022022. doi:10.1088/1742-6596/1168/2/022022
44. Montesinos López OA, Montesinos López A, Crossa J. Overfitting, Model Tuning, and Evaluation of Prediction Performance. In: Montesinos López OA, Montesinos López A, Crossa J, editors. Multivariate Statistical Machine Learning Methods for Genomic Prediction. Cham: Springer International Publishing; 2022. p. 109-39. doi:10.1007/978-3-030-89010-0_4
45. Mutasa S, Sun S, Ha R. Understanding artificial intelligence based radiology studies: What is overfitting? Clin Imaging. 2020;65:96-9. doi:10.1016/j.clinimag.2020.04.025 PMid:32387803 PMCid:PMC8150901
46. Yu AC, Mohajer B, Eng J. External Validation of Deep Learning Algorithms for Radiologic Diagnosis: A Systematic Review. Radiology: Artif Intell. 2022;4(3):e210064. doi:10.1148/ryai.210064 PMid:35652114 PMCid:PMC9152694

How to Cite this Article:
Sehat MS, Hariri A, Soltani P, Khademi A. Artificial intelligence in detecting mandibular fractures: A review of literature. Arch Trauma Res. 2023;12(2):102-111. doi: 10.48307/ATR.2023.176358