TURKISH JOURNAL OF EMERGENCY MEDICINE

Invited Review Article

Access this article online

Quick Response Code:

Website:
https://turkjemergmed.com/

DOI:
10.4103/tjem.tjem_45_25

Mapping artificial intelligence models in emergency medicine: A scoping review on artificial intelligence performance in emergency care and education

Göksu Bozdereli Berikol[1], Altuğ Kanbakan[1]*, Buğra Ilhan[2], Fatih Doğanay[3]

[1]Department of Emergency Medicine, Ufuk University School of Medicine, Ankara, [2]Department of Emergency Medicine, Kırıkkale University School of Medicine, Kırıkkale, [3]Department of Emergency Medicine, University of Health Sciences School of Medicine, İstanbul, Türkiye
**Corresponding author*

Abstract:

Artificial intelligence (AI) is increasingly improving the processes such as emergency patient care and emergency medicine education. This scoping review aims to map the use and performance of AI models in emergency medicine regarding AI concepts. The findings show that AI-based medical imaging systems provide disease detection with 85%–90% accuracy in imaging techniques such as X-ray and computed tomography scans. In addition, AI-supported triage systems were found to be successful in correctly classifying low- and high-urgency patients. In education, large language models have provided high accuracy rates in evaluating emergency medicine exams. However, there are still challenges in the integration of AI into clinical workflows and model generalization capacity. These findings demonstrate the potential of updated AI models, but larger-scale studies are still needed.

Keywords:

Artificial intelligence, emergency medicine, image processing, large language models, machine learning, signal processing

Submitted: 04-02-2025
Revised: 07-03-2025
Accepted: 08-03-2025
Published: 01-04-2025

ORCID:
GBB: 0000-0002-4529-3578
AK: 0000-0002-1063-3018
BI: 0000-0002-3255-2964
FD: 0000-0003-4720-787X

Address for correspondence:
Dr. Altuğ Kanbakan, Department of Emergency Medicine, Ufuk University School of Medicine, Ankara, Türkiye.
E-mail: drkanbakan@outlook.com

EMAT EMERGENCY MEDICINE ASSOCIATION OF TURKEY

Introduction

Artificial intelligence (AI) is a rapidly advancing, game-changing technology in health care. Emergency medicine, as a young and rapidly updating field with its sub-branches open to technologies, provides an ideal foundation for AI applications. AI studies have been increasing logarithmically in recent years and are being applied with different methods in many areas of emergency medicine. The use of AI in areas such as triage, diagnosis, outcome prediction, and research on this topic is rapidly increasing. The performance of applications of AI models generally varies depending on the models and usage areas.

Although there are a large number of reviews in the literature focusing on specific areas of the use of AI in emergency medicine, most of the existing studies remain limited in scope. Furthermore, these studies of AI inherently become outdated over time. This scoping review aims to investigate the current areas of use of AI in emergency medicine and investigate their performance in these areas by categorizing them under AI applications.

How to cite this article: Berikol GB, Kanbakan A, Ilhan B, Doğanay F. Mapping artificial intelligence models in emergency medicine: A scoping review on artificial intelligence performance in emergency care and education. Turk J Emerg Med 2025;25:67-91.

Methods

This scoping review was conducted according to the PRISMA Scoping Review guidelines. There is an unprecedented rise in AI models, and models are frequently updated, with older versions becoming obsolete. While traditional machine learning (ML) models are being replaced by ensemble methods and deep learning (DL), previous versions of large language models (LLM) are disappearing from use as new versions are released. AI studies are also experiencing logarithmic increases at this rate all over the world. For these reasons, articles published between January 1, 2024, and January 1, 2025, were scanned in order to provide an up-to-date compilation. Case reports, reviews, comments and letters, and studies not related to AI and emergency medicine are excluded. PubMed and Web of Science (WoS) databases are searched within this scope using boolean search operators.

Our research question is determined as "In which areas (triage, diagnosis, prediction, etc.) are AI-supported systems more effective in the field of emergency medicine" and the search was made with Boolean search strategies and includes keywords and boolean operators optimized in accordance with the research questions in WOS and Pubmed Databases.

WOS: ALL=((TOPIC: ("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Image Processing" OR "Large Language Model" OR "Natural Language Processing" OR "Signal Processing")) AND (TOPIC: ("Emergency Medicine" OR "Emergency Department" OR "Triage" OR "Prehospital"))).

Pubmed: (("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Image Processing" OR "Large Language Model" OR "Natural Language Processing" OR "Signal Processing" OR "Big Data") AND ("Emergency Medicine" OR "Emergency Department" OR "Triage" OR "Prehospital")).

During the searches, articles written in English and published between January 2024 and January 2025 were scanned and studies that met these criteria were evaluated for eligibility and the selection process was shown in the PRISMA flowchart [Figure 1]. In this scoping review, the selection process of studies was carried out in three stages: "Title screening," "abstract screening" and "full text screening." First, the titles of the studies obtained as a result of the search in the databases were scanned, and those that were not directly related to the research questions were eliminated. The abstracts of the studies that passed the title screening and, in the last stage, the full texts of the studies that passed the abstract review were evaluated, and those that fully met the criteria were included in the review. Search results in both databases were evaluated by two independent researchers, with disagreements regarding selection resolved by a third researcher [Figure 1]. Ineligible study design, studies not related to the use of AI in emergency medicine, studies that include AI but not in the context of emergency medicine, bioinformatics studies, theoretical models, animal models, studies with very small sample size ($n < 10$), retracted, preprinted and whose results are not reported are excluded from the review.

Data collected from studies included in this scoping review are investigated for population, intervention, comparison and outcome, type of AI application (e.g., triage, diagnosis, outcome prediction), AI methods used (e.g. ML, image processing, signal processing), and the performance metrics associated with each AI model.

Results

We reviewed a total of 1360 studies on the use of AI in emergency medicine. The distribution according to the reasons for exclusion is shown in the flowchart. The concepts are investigated under two essential categories in emergency medicine: Emergency patient care and Emergency Medicine Education. In emergency patient care, the AI models are evaluated as its subtitles: image processing (n = 36), text mining (n = 43), signal processing (n = 11), and data mining with structured big data (n = 85). On the other hand, there were 12 studies involving emergency medicine education. In total, 187 studies were included in the review.

This scoping review categorizes AI applications in emergency medicine into two main domains: emergency patient care and emergency medicine education.

Emergency Patient Care

Image processing

Artificial ıntelligence-assisted image processing procedures

The effectiveness of AI-based image processing analyses depends on both the preprocessing techniques applied to the images and the efficiency of the selected method. Medical images can vary based on the type of machine used, how the image is taken, and differences between patients.[1-3] DL models achieve higher accuracy with large datasets, prompting researchers to integrate multiple datasets to enhance performance. However, combining data from different sources can make training AI models more difficult.[4,5]

Therefore, standardizing datasets and minimizing image variations are essential for AI models to accurately learn specific patterns.[6] During the preprocessing stage,

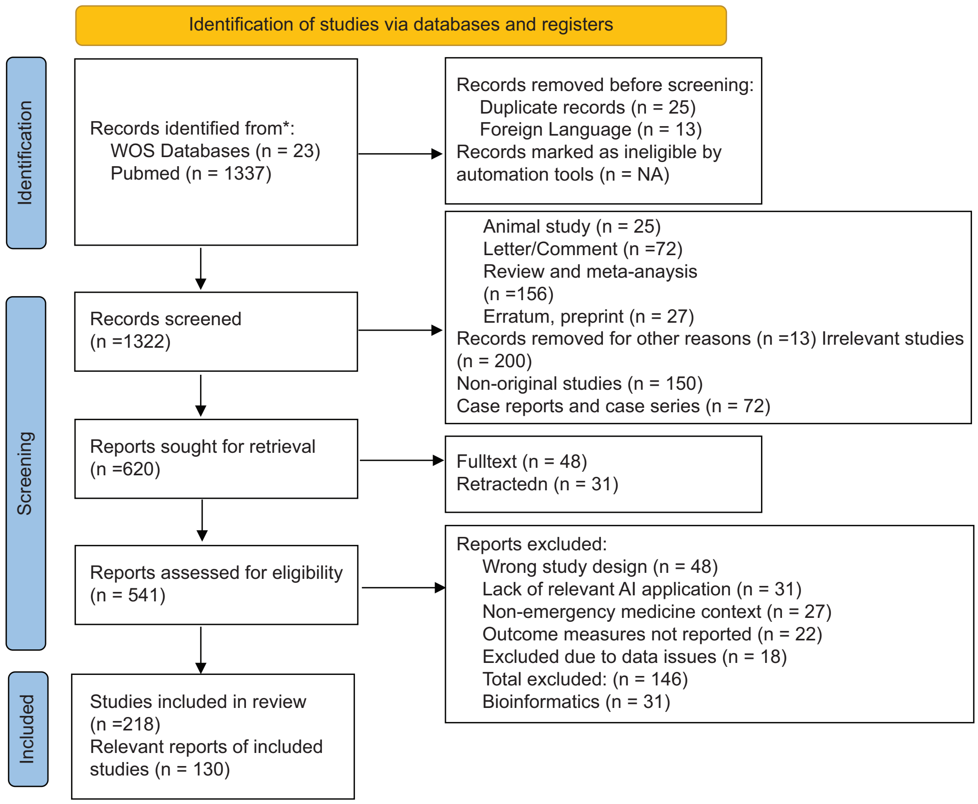

Figure 1: Flowchart diagram of the selection process in the review

techniques such as image normalization, denoising, and data augmentation enhance the generalization capability of models.[7] In addition, accurate labeling and annotation are fundamental factors that determine the success of model training. Since manual labeling is a time-consuming process, semi-automatic labeling, and AI-assisted annotation tools offer significant advantages at this stage.[8,9] These optimizations improve the accuracy and reliability of analysis algorithms, ultimately leading to more dependable results in clinical applications.

One of the most critical stages in image processing is the application of image enhancement techniques.[10] Raw images often exhibit low contrast, noise, or artifacts, making direct analysis challenging. To improve image quality, techniques such as contrast enhancement, edge detection, filtering, thresholding, and segmentation are commonly employed.[11-13] In medical imaging, the effective application of these techniques is particularly crucial for obtaining clearer and more accurate diagnostic results.

Following these preprocessing steps, the model design and training process begins. The performance of AI models depends on the choice of algorithms, the quality of the training dataset, and the model's learning process.[14] In medical image processing, both supervised learning methods (e.g., convolutional neural network [CNN], ResNet, VGG) and unsupervised learning techniques (e.g., autoencoders, GANs) are widely utilized. During model training, errors are minimized through loss function calculations, model parameters are optimized using the backpropagation algorithm, and the efficiency of the process is enhanced with GPU-accelerated computations.[15,16]

After the training process, the model must be validated and tested to ensure its reliability. Performance metrics such as accuracy, precision, recall, F1 score, structural similarity index (SSIM), peak signal-to-noise ratio, and area under the curve (AUC) are commonly used to assess the model's overall effectiveness. In addition, cross-validation methods are applied to prevent overfitting and ensure that the model performs consistently across different datasets.[17]

In the final stage, the model undergoes optimization and deployment. Techniques such as learning rate adjustments, regularization methods, and model compression (e.g., quantization, pruning) are implemented to enhance efficiency in real-world applications. As AI-based image processing solutions continue to be adopted across various domains, ongoing optimization and refinement of these models remain crucial.[18] The advancement of AI-driven image-processing technologies is driving revolutionary transformations in sectors such as medical diagnostics, autonomous systems, and industrial automation.

Preprocessing procedures for images: Artificial ıntelligence-assisted medical ımage processing approach

The preprocessing phase enhances the accuracy and reliability of analysis algorithms by reducing image variations and improving overall image quality. Preprocessing steps tailored to different medical imaging techniques play a critical role in determining the success of AI models. Employing preprocessing techniques that align with the specific characteristics of an image can enhance model performance and contribute to more accurate clinical decision-making.[7,17]

X-ray is a two-dimensional imaging technique represented by a single image frame. In contrast, computed tomography (CT) and magnetic resonance imaging (MRI) are three-dimensional imaging methods that capture multiple cross-sections of a specific anatomical structure at a given point in time. Ultrasound (USG) is a dynamic imaging technique that records video sequences composed of multiple frames over a specific time interval.

While basic preprocessing techniques are applied to individual frames in X-ray images, they are implemented for each section in CT and MRI scans and for each frame in USG videos. However, due to the unique characteristics of each imaging modality, specialized preprocessing techniques have been developed for different imaging types, including X-ray, CT, MRI, and USG, to optimize AI model performance.[13]

X-ray images are often affected by low contrast, low resolution, noise, and artifacts. To enhance the effectiveness of AI models, preprocessing techniques such as noise reduction, contrast enhancement, and edge detection are applied. Commonly used noise reduction methods include Gaussian Blur, Median Filtering, and Wavelet Denoising.[19] Histogram equalization and Contrast Limited Adaptive Histogram Equalization improve the visibility of bone structures,[20] while normalization techniques contribute to model performance and robustness.[17]

In CT and MRI images, each slice must be processed individually to maintain data consistency. Resizing is commonly performed to standardize image dimensions across datasets.[15,21] Various methods are employed to remove metal artifacts, including metal artifact reduction algorithms and iterative reconstruction techniques. In MRI images, intensity variations caused by magnetic field inhomogeneities can negatively impact model learning. To correct these variations, N4ITK bias field correction is frequently applied during preprocessing. In addition, normalization techniques and 3D CNNs improve data processing efficiency and model accuracy.[22-24]

USG videos require specialized preprocessing due to high noise levels and variable contrast. Motion analysis and optical flow algorithms are used to identify key frames.[25] Speckle noise, a common issue in USG imaging, is reduced using techniques such as the Wiener filter and anisotropic Diffusion.[26] Image quality can be further enhanced with super-resolution techniques and histogram equalization.[27] For time-series analysis of USG videos, long short-term memory networks and 3D CNN-based approaches are often employed.[28]

X-ray

During the review period, a total of 9 X-ray studies were evaluated based on predefined inclusion and exclusion criteria[29-37] [Supplementary Table 1] (https://turkjemergmed.com/pages/2025-2-issue-supplementary-files). These studies primarily focused on the analysis of bone structures and chest X-ray images. Among the reviewed studies, the research by Wang *et al*. stands out due to its use of the largest dataset. In this study, an EfficientNetV2-based model was developed using 3498 chest radiographs along with external datasets, achieving an AUC of 0.878 in detecting pulmonary tuberculosis in the test set.[31]

The highest diagnostic performance was reported in the study by Ghatak *et al*.,[37] where the Annalise Enterprise CXR AI model was used to detect vertebral compression fractures in 596 chest radiographs (272 positive and 323 negative cases). This AI model demonstrated strong performance in the automated diagnosis of vertebral compression fractures, achieving an AUC of 0.955.

Conversely, the study with the lowest performance was the external dataset validation conducted by Wang *et al*.[31] The objective of this study was to develop and validate a DL-based computer-aided diagnosis (CAD) algorithm for detecting pulmonary tuberculosis in emergency department settings. The study compared the performance of the EfficientNetV2-based CAD algorithm with radiologists' clinical reports. The findings indicated a decrease in model performance when tested on the Montgomery (AUC: 0.838) and Shenzhen (AUC: 0.806) datasets, highlighting the limitation of using single-center data in terms of generalizability.[31]

Overall, AI-based analyses in X-ray imaging have been shown to enhance diagnostic accuracy, assist in fracture detection, and improve the identification of pulmonary diseases. However, challenges such as artificial dataset augmentation, studies conducted in limited clinical settings, and issues related to model generalizability remain key limitations.[38]

Computed tomography

A total of 11 CT studies were evaluated based on inclusion and exclusion criteria during the publication review period.[39-49] These studies assessed the effectiveness of AI applications across a wide range of clinical conditions, including acute pancreatitis, ureteral stones, skull fractures, intracranial hematomas, cervical fractures, and aortic dissection.

The study with the largest dataset was conducted by Ruitenbeek *et al*., which included cervical spine CT images from 2973 patients and evaluated the impact of the AIDoc Medical AI algorithm on cervical fracture detection.[48] The AI-assisted workflow improved diagnostic efficiency by achieving an accuracy of 94.8% and reducing the average diagnosis time for fracture cases by 16 min.

In terms of performance, AUC and accuracy metrics varied between 0.788 and 0.993. The highest AUC value (0.993) was reported in the study by Zhang *et al*., which focused on the classification and severity assessment of acute pancreatitis.[44] The model, trained on a dataset of 190 patients, demonstrated high accuracy in pancreatic segmentation and successfully detected complications such as peripancreatic necrosis and edema.

The lowest-performing model was developed by Choi *et al*. for the detection of cerebral hemorrhage. The DLHD algorithm, evaluated on 111 brain CT images, achieved the lowest AUROC value of 0.788. The study indicated that while the model improved sensitivity, it also reduced specificity and exhibited a high false-positive rate.[45]

While AI-based analyses of CT images provide significant advancements in early diagnosis and rapid intervention, challenges such as the lack of large-scale multicenter validation and difficulties in adapting models to different imaging protocols remain critical considerations for clinical integration.

Magnetic resonance imaging

During the publication review period, a total of three MRI studies were evaluated based on inclusion and exclusion criteria.[50-52] These studies primarily investigated the diagnostic efficacy of acute ischemic stroke (AIS) detection, mortality prediction, and ultrafast brain MRI protocols. The sample sizes varied, with the largest dataset belonging to a study that developed a DL-based model for mortality prediction in ischemic stroke patients, utilizing data from 2710 individuals.[50]

In terms of performance, AUC and accuracy metrics ranged between 0.852 and 0.95. The highest AUC value (0.95) was reported in the study by Kim *et al*., which developed a 3D CNN model for AIS detection.[50] In addition, Lang *et al*. evaluated a 2-min ultrafast brain MRI protocol designed for rapid imaging in emergency settings, demonstrating a diagnostic agreement of 98.5%.[51]

Overall, AI-supported MRI analysis has been shown to enhance diagnostic accuracy in emergency situations, expedite patient management, and support clinical decision-making. However, challenges such as single-center study designs, demographic imbalances, and limitations in model generalizability remain key considerations in the broader implementation of these models.

Ultrasonography

During the publication review period, a total of four USG studies were evaluated based on inclusion and exclusion criteria.[53-56] These studies explored the effectiveness of AI applications in various clinical settings, including cardiac function assessment, carotid artery compressibility analysis, acute gallbladder pathologies, foreign-body detection, and the determination of return of spontaneous circulation (ROSC) during cardiopulmonary resuscitation (CPR).

In terms of performance, AUC and accuracy metrics ranged from 0.81 to 0.99. The highest accuracy value (99.1%) was reported in the study by Holland *et al*., which utilized U-Net and YOLOv7-based AI models for foreign-body detection in USG images containing 12,144 annotations.[56] The study highlighted that these models could expedite decision-making, particularly in remote areas or settings with limited access to experts. However, the labeling process was noted to be resource-intensive, and *in vivo* validation remained limited.

The lowest AUROC value (0.81) was recorded in the study by He *et al*., which evaluated cardiac function using point-of-care echocardiography (point-of-care USG [POCUS]) in the emergency department.[53] The EchoNet-POCUS model achieved an AUROC of 0.92 for cardiac function assessment but only 0.81 for video quality. While the model accelerated bedside assessment and reduced operator dependency, its lack of multicenter validation and challenges in adapting to different imaging devices were cited as limitations.

Park *et al*. introduced RealCAC-Net, an AI-based model designed to determine ROSC during CPR by analyzing carotid artery compressibility.[54] Trained on 11,958 images for training and 15,080 for testing, the model demonstrated superior performance over traditional manual palpation, achieving 96% accuracy and a 97% F1 score. This system has the potential to enhance in-hospital resuscitation management by supporting decision-making during CPR. However, concerns regarding its generalizability across different devices and patient populations were noted.

Ge *et al*. investigated the use of AI in diagnosing acute gallbladder pathologies.[55] A DL model, trained on 266 USG images from 186 patients, distinguished normal from abnormal gallbladder cases with 91% accuracy and categorized urgent versus nonurgent cases with 82% accuracy. The study aimed to facilitate rapid triage in gallbladder pathologies, potentially reducing reliance on specialist radiologists.

While AI-based analyses of USG images significantly contribute to rapid diagnosis and patient management, challenges such as the lack of multicenter validation, device dependency, and sensitivity to imaging quality must be addressed for broader clinical integration.

Alternative image analysis and artificial ıntelligence applications beyond standard medical ımaging methods

During the publication review period, a total of nine alternative image analysis studies were evaluated based on inclusion and exclusion criteria.[57-65] These studies explored the effectiveness of AI applications across various clinical settings, including the detection of retinal diseases using fundus photographs, fracture identification with infrared thermal images, anemia screening via conjunctival photographs, stroke detection from facial images, and cardiac function analysis using electrocardiography (ECG) images.

The study with the largest dataset was conducted by Song *et al*., focusing on the automatic detection of posterior segment pathologies using 90,250 robotic alignment optical coherence tomography images.[62] The model, named RobOCTNet, demonstrated high efficacy as a triage tool in ophthalmology emergency settings, achieving an AUC of 1.00 in internal validation and 0.91 in external testing. However, the study highlighted limitations, such as the model's training on a relatively small volumetric dataset and its lack of real-world clinical integration.

In terms of performance, AUC and accuracy metrics ranged from 0.75 to 1.00. The highest AUC value (1.00) was reported in the study by Song *et al*.[62] Conversely, the lowest accuracy (75.4%) was observed in the study by Zhao *et al*., which focused on anemia detection using conjunctival photographs.[61] The smartphone application eMoglobin was utilized to detect anemia by analyzing conjunctival images, achieving an AUC of 0.92 at an HBc threshold of 7 g/dL. However, the study noted the model's limited sensitivity in detecting mild anemia cases.

In addition, Biousse *et al*. reported an AUC of 0.97 in the detection of papilledema using the BONSAI-DLS model with nonmydriatic fundus photographs.[57] Another notable study by Wang *et al*. developed an AI model for diagnosing AIS from facial images of stroke patients. This model, based on EfficientNet and ResNet50, achieved an AUC of 0.91 in cross-validation and 0.82 in independent tests.[60]

Furthermore, Maxin *et al*. introduced a model aimed at distinguishing ischemic from hemorrhagic stroke through a combination of pupillometry and ML. This model, which demonstrated an accuracy of 91.5%, has the potential to serve as a valuable decision-support tool in prehospital stroke management.[64]

Although these alternative imaging modalities and AI-supported analyses hold promise for enhancing rapid diagnosis and patient management, challenges such as the need for broader multicenter validation, dataset balancing, and clinical adaptation remain key considerations for their widespread implementation.

Text mining

Text mining is used to analyze unstructured medical records, such as triage notes and discharge summaries, to identify important patterns.[66] This prediction and feature extraction requires a certain preprocessing and analysis using NLP, an AI method that helps computers analyze and understand written text.[66]

In recent years, two concepts, LLM and NLP, have rapidly gained popularity, leading to a surge in publications and applications. Given the abundance of verbal and unstructured data in emergency medicine, these concepts have found extensive use in the field. Particularly in reporting, research is increasingly focused on models for epicrisis summarization, feature extraction from triage and anamnesis notes, and predictive analysis.

Forty-three studies conducted on NLP (n = 26)[67-92] and LLM[76,91,93-107] (n = 17) in emergency medicine are included in the review [Supplementary Table 2] (https://turkjemergmed.com/pages/2025-2-issue-supplementary-files).

Natural language processing

The majority of NLP-based studies are designed to retrospectively analyze unstructured text data, including triage notes, medical history, and emergency department notes from the electronic health records to predict emergency department patient triage,[71] diagnosis,[69,74,83] need for intervention[70,80,82] and outcome.[67,69,71-73,75,77,79,81,82,83]

Mostly used methods are transformer-based DL, Bidirectional Encoder Representations from Transformers (BERT),[67,70-72,76,79,81] Term Frequency-Inverse Document Frequency (IDF),[74,75,77,78,80] Bag of Words.[75,78] The ML models used are mostly ensemble methods, including the types of boosing algorithms as Categorical Boosting, Light Gradient Boosting Machine, Extreme Gradient Boosting (XGB), Logistic regression, and Deep Neural Networks.

Especially in NLP studies, the use of structured data together with unstructured data has a significant impact on AUC values. One study showed an increase in the AUC values when using both structured and unstructured data on the prediction of ED dispositions with the chief complaint, vital signs, and demographics.[78]

The highest patient population was seen with 1,391,988 patient records by Patel *et al*., where BioClinical BERT was used in the hospitalization decision-making from triage notes.[75] Within the duration of this review, NLP has been applied to diagnostic predictions such as syncope detection (AUC = 0.95), febrile convulsion prediction (F1 = 0.921), serious infection prediction (AUC = 0.913) and COVID-19 prediction (F1 = 0.796).

Evaluating the performances of the predictive analyses in terms of the need for intervention, Chai *et al*. found the highest AUC value of 0.89 in 38,214 patients for predicting the surgery indication, while Weidman *et al*. reported the highest performance as an AUC of 0.79 using histogram gradient boosting with TD-IDF in the predicting life-saving intervention, laboratory, and imaging needs on 12,913 patients just at the prehospital area.[80,82]

These data show that NLP methods show moderate and high performance in the prediction of diagnosis and outcome. The large data differences between studies indicate that the methods and the performance comparisons vary on different data sets and structured data integration. The NLP methods have been effective in many areas, from diagnose in the emergency department to predicting sociodemographical processes. These findings reveal that NLP-based methods are largely studied, however, further optimization and transparent tuning processes are required. Further testing and optimization of NLP-based clinical decision support systems is critical for clinical applications.

Large language models

The development of LLM has accelerated significantly in the last 2 years. Initially, LLM were trained on large datasets to predict text, then improved with human feedback. These models have become capable of performing various language-based tasks and have acquired skills such as few-shot learning. LLM models can summarize medical records, suggest possible diagnoses, and have demonstrated strong performance on medical examinations. However, the security challenges, risk of generating misinformation, and hallucinations are still an issue. Significant improvements have been made to make these models much more secure than previous models.[108]

One of the most studied models, GPT-1, one of the first versions of Open AI, was released in 2018 and worked with limited training data and performed well on many NLP tasks.[109] However, as the model size increased, it was able to perform better on more complex tasks. GPT-2 was released in 2019 with 1.5 billion parameters, making it much more powerful and successful on general language tasks. Then, GPT-3 was released in 2020 with 175 billion parameters and undertook many NLP tasks. Finally, GPT-4 has much more powerful features and attracted attention with its ability to accept multimodal data inputs. The development of these models has been made possible especially by the combination of large data sets and powerful processing resources. In a noticeably short time, their accuracy rates are rapidly increasing, thus decreasing the use of older versions. However, the use of these technologies in critical areas such as medicine still faces many challenges in obtaining accurate and reliable results.[109]

LLMs are recently been studied within the emergency department data. The studies used different LLM, such as GPT and Bard and examined how these models perform in clinical decision-support processes. Most LLM studies focused on GPT-3.5 and GPT-4, with comparing them to Bard and other specialized models. Among the methods used, the application of LLMs in critical areas such as physician decision support systems, patient triage, and disease diagnosis was prominent.

When the studies are examined, more prospective observational and prospective cohort studies are encountered, and it is seen that they contribute to the solution of various clinical problems such as triage,[91,92,94-96,101] diagnosis,[92,95,98,110] appropriate test selection and outcome (admisssion[105] and mortality) prediction. The most commonly used LLM was GPT 4 (n = 10), followed by GPT 3.5 (n = 6), BERT, Copilot, and Llama2.

Although the performance is found to be higher than NLP studies, the number of patient data was lower at 45 (AUC 0.87)[94] to predict outpatient triage and higher as 864.089 to predict hospitalization with BERT with XGBoost, resulting in an AUC up to 0.87.[105]

These differences in data size indicate that testing LLM-based models with larger datasets may yield more reliable results in clinical practice. When the performance results of the studies were examined, it was determined that LLMs were generally successful, but some models fell short of expectations. These differences reveal that LLM need to be tested and optimized further before they can be fully integrated into clinical use. Although LLMs generally outperform traditional NLP methods, data availability and sample sizes vary widely. For example, studies using LLMs in 864,089 records for hospitalization prediction with BioClinicalBERT and XGBoost with an AUC of 0.87, while another large study with 484,094 patients used NLP with GB showed an AUC of 0.92 in ICU admissions.[89,105] These differences indicate that LLM-based models require larger datasets for reliable clinical integration. The accuracy changes (AUROC = 0.65) among diagnosis, highlighting reliability concerns in stroke screening processes with GPT 3.5,[100] which is one of the older versions of GPT models. One of the most successful methods in diagnosis was found to be a Multilingual BERT by Levra *et al*., which predicts syncope from emergency department notes with symptom extraction and F1 scores of 0.98.[69]

Signal processing

Signal processing and AI-assisted analysis techniques are increasingly playing a role in emergency medical decision-making processes. The studies examined in this review focus on the processing of physiological signals, such as ECG and brain imaging, using AI algorithms to increase clinical diagnostic accuracy [Supplementary Table 3] (https://turkjemergmed.com/pages/2025-2-issue-supplementary-files).

When the geographical distribution of the reviewed studies is examined, South Korea is represented by 50.0%,[111-116] China by 16.7%,[117,118] international by 8.3%, Taiwan by 8.3%,[119] Europe (France and Spain, Germany) by 16.7%.[120,121]

When evaluated in terms of intervention features, one of the examined studies (8.3%) focused on stroke detection.[120] The number of studies, including ECG analysis, was 3 (25.0%), two of which (16.7%) were directly aimed at the diagnosis of myocardial infarction (MI).[111-113] The number of studies on arrhythmia detection (n = 1, 8.3%),[119] optimizing the CPR process was (n = 2, 16.7%),[114,115] prediction of admission to the intensive care unit and early warning systems (n = 2, 16.7%)[116] and SARS-CoV-2 detection (n = 1, 8.3%).[121]

When the studies are analyzed in terms of AI methods, CNNs is the most widely used method in signal processing and medical decision support systems. Four of the studies (33.3%) adopted CNN-based approaches. In addition, two (16.7%) studies performed signal analysis using the transformer architecture. DL techniques were generally applied in two (16.7%) studies. Advanced modeling methods such as LightGBM were also used by 3 (25.0%) studies.

Electrocardiography

Herman *et al*. developed a DL-based model to evaluate the performance of myocardial infarction (OMI) detection on 12-lead ECG data in patients with suspected acute coronary syndrome. The study determined that the model showed two-fold higher sensitivity compared to STEMI criteria, but lower specificity.[122] It was suggested that the model has the potential to improve patient outcomes by supporting early diagnosis and revascularization decisions in prehospital and emergency departments. Lee *et al*. developed a DL model that can extract digital STEMI biomarkers from printed ECG outputs to improve prehospital telecardiology. It was determined that the model achieved similar sensitivity and specificity levels with expert consensus.[112] Jang *et al*. proposed an AI-supported ECG analysis model in determining the etiology of dyspnea. The model provided higher diagnostic accuracy than the NT-proBNP test.[111]

Park *et al*. evaluated an AI-based Quantitative ECG system in the detection of acute coronary occlusion after OHCA.[113] The diagnostic performance of the model was compared with expert assessment and shown to be noninferior. Liu *et al*. developed a CNN model that classifies arrhythmias with single-lead ECG. The model achieved high accuracy with short-term ECG recordings.

Cardiopulmonary resuscitation

Han *et al*. developed a noninvasive blood pressure prediction model during CPR. The model achieved high correlation coefficients in estimating systolic blood pressure, diastolic blood pressure, and mean arterial pressure. This study provides significant contributions to the real-time evaluation of the CPR process.[114] Kim *et al*. showed that the AI-assisted CPR robot provided similar hemodynamic results to LUCAS 3. The study revealed that AI can create individualized CPR management.[115]

Stroke

Ou *et al*. created a multimodal DL model that combines video images and clinical data to provide early diagnosis of stroke patients. The model achieved higher accuracy compared to individual modalities.[117] Sen *et al*. developed a ML model that analyzes hemodynamic waveforms obtained from carotid arteries to detect large vessel occlusions in patients with AIS.[120] The study has the potential to contribute to the development of a low-cost, rapid prehospital screening tool that can be integrated with devices such as portable Doppler USG.

Alert systems

Zhang *et al*. developed a ML model using only non-invasive parameters to predict the need for invasive mechanical ventilation (IMV).[118] The model achieved a higher AUC value (0.935) than traditional risk-scoring methods. These findings form the basis of a system that can enable the prediction of IMV needs through early warning systems in prehospital and emergency department environments. Choi *et al*. developed a ML-based model to increase the effectiveness of early warning systems in intensive care patients.[116] It was shown that the model has higher sensitivity than traditional scoring systems and can accelerate emergency intervention processes.

Other

Woehrle *et al*. developed a breath analysis model using semiconductor-based electronic nose (E-Nose) technology to distinguish patients with SARS-CoV-2 pneumonia from uninfected individuals. The study shows that it has the potential to provide a rapid, noninvasive, and portable solution for the diagnosis of SARS-CoV-2 and similar respiratory diseases.[121]

These studies offer significant contributions to the integration of AI and signal-processing techniques into clinical decision-support systems in prehospital and hospital environments of emergency patient care. Such approaches in the field of signal processing have the potential to improve patient outcomes by optimizing early diagnosis and intervention processes.

Data mining on structured big data

Big data refers to large sets of patient records, lab results, and imaging reports that AI can analyze for patterns. Big data is large, fast-growing, and diverse, making it useful for AI-driven analysis in emergency care. However, raw big data is not inherently valuable; its true potential is realized through proper analysis and integration into clinical workflows. AI can analyze big data to predict ED overcrowding, patient deterioration, and other critical issues via extracting hidden patterns and relieving unknown associations.[123]

The health sector produces large amounts of data instantly, at high speed, and in variety. ML and DL methods are being used to improve health care, reducing human error regarding disease detection, diagnosis, prediction, drug discovery, precision medicine, and robotic surgery.[38] The digitalization of such data (transformation from hard copy to digital data) has paved the way for big data analytics applications in the health sector, which is promising.

Supervised learning is used for labeled (Survivor vs. nonsurvivor, admission vs. discharge, disease present vs. absent, etc.) data, and unsupervised learning is used for unlabeled data. While structured data (categorical and numerical data including laboratory results, demographics, vital signs, structured history data, etc.) is mostly used in prediction models such as mortality, risk stratification, and length of stay estimations, unstructured data is commonly applied in clustering and text-based AI applications. ML models using structured data are frequently used in medical research.

Despite the rapid adoption of AI in emergency medicine, significant challenges remain, including data quality issues, bias in predictive models, and integration barriers with existing clinical workflows. Emergency physicians should pioneer the use of new technologies in emergency medicine practice. These technologies should be seen as tools that enhance clinical decision-making and efficiency rather than as substitutes for the expertise and judgment of healthcare professionals. Studies have shown variable levels of success in AI-powered models. AI models predicting emergency department overcrowding have achieved AUC values ranging from 0.70 to 0.89,[124-128] indicating moderate-to-high predictive power but still requiring further validations and optimization. However, physician-AI collaboration holds promise for improving the quality of patient care and reducing medical errors and costs.

Data is generated when the patient first contacts the healthcare system, either remotely or face-to-face. Since almost all data are produced digitally today, it can be processed instantly, and decision-support systems can be started to operate. AI-supported systems and ML models are frequently used in medical research and for outcome and risk prediction [Supplementary Table 4] (https://turkjemergmed. com/pages/2025-2-issue-supplementary-files).

Prehospital

Nine studies related to prehospital patient care were reviewed.[129-137] The studies evaluated the performance of AI and ML-powered models for decision-making of transfer and termination of resuscitation (TOR), predicting short- and long-term mortality, bed availability before transfer, and determining factors that cause transfer delays.[129-137]

Although the sample sizes of the studies varied, the largest data set was the study by Kajino *et al*., which evaluated the effectiveness of AI-supported decision support systems in the TOR.[133] The study reported an AUC of 0.96, which is a highly accurate predictive model for TOR.

It was observed that the performance of the prediction models was evaluated in seven studies on structured data. In these studies, AUC, mortality rates, accuracy, and specificity were used in the performance evaluation. Farhat *et al*. developed XGBoost and RF models for transport decision-making, reaching 95% and 97% specificity values, respectively.[134] Kajino *et al*.'s AI-supported models achieved an AUC of 0.96 in neurologically survival favorable survival prediction in OHCA regarding TOR decision-making.[133] This study shows the potential of AI on one of the decision points in prehospital cardiac arrest management.

Besides its use for resuscitation; AI is also involved in resource management studies. Xu *et al*. showed a real-time simulation-based application integrating live data from 48 hospitals to optimize dispatch with prehospital bed availability predictions, potentially reducing transport delays and improving patient outcomes.[136] Furthermore, ML models are used in survival predictions of trauma patients using Survival Tree and Random Forest algorithms, effectively predicting 8-h and 24-h survival probabilities in severe trauma patients.[137]

Overall, prehospital AI models have shown similar or more successful results than traditional methods. However, concerns regarding real-time implementation in prehospital area, interpretability of the models, and physician reliance on AI recommendations still remain unsolved and require external validation and prospective trials to assess real-world applicability.

Triage

Triage is one of the most critical concepts in emergency medicine. Due to its nature, it involves sorting and prioritizing patients, making it inherently complex and filled with numerous gray areas. Various triage models have been developed to differentiate those who require urgent medical care, particularly in situations where resources are limited or demand surges. Among these models, five level triage systems such as the Canadian Triage and Acuity Scale (CTAS) and Manchester Triage System, which are complaint based, as well as the Emergency Severity Index (ESI), which is algorithm based and focuses on resource utilization, have been widely used.[138-140] Beyond these, additional scoring systems have been developed to assess urgency at different levels.

One of the most critical challenges in triage is the issue of overtriage and undertriage. Undertriage can lead to delays in providing timely emergency care to patients, while overtriage results in unnecessary resource utilization.[141] Moreover, triage accuracy is influenced by several factors, including the experience of the triage team, the discrepancy between supply and demand, and other factors.

Given its many gray areas, triage has become a significant area of research in ML applications, with numerous studies focusing on integrating AI to enhance decision-making and improve triage accuracy. As a result of the inclusion and exclusion criteria, 15 articles related to triage were reviewed. The studies mostly focus on validation studies of ML models developed for identifying low-acuity and high-acuity patients. AI-driven triage models have been applied in pediatric and adult patient groups with decision-making in trauma, major incidents, CBRN cases, and incorrectly classified patients (overtriage and undertriage).[138,142-155]

It was determined that the sample sizes (studies conducted on real cases) are quite large, reinforcing the generalizability of the findings. The largest dataset sample, consisting of 1,833,908 ED patients, was studied by Look et al. to address class imbalance in ED classification models.[147] It has been observed that model performances are generally determined by AUC values, which range from 0.75 to 0.91. In addition to AUC values, accuracy, F1 score, sensitivity, and over/undertriage rates were also used to evaluate model performances.

While Chen *et al*. introduced the Low Acuity Visit Algorithms model, which effectively identified nonurgent patients using logistic regression and random forest classifiers.[146] Yu *et al*. conducted an external validation study using the AutoScore framework to predict 2-day mortality among ED patients, showing improved interpretability and robustness.[148] Evaluation of the performance of the modelLook *et al*. developed an AutoScore-Imbalance framework to improve class imbalance in triage models, achieving AUC values between 0.75 and 0.91 with a higher sample size.[147]

The models are also compared with traditional models as Grant *et al*. demonstrated that ML models outperformed the CTAS in predicting the need for early critical care within 12 h, utilizing DL and gradient-boosted trees.[153] Nanini *et al*. developed an ML model for hypoxemia severity triage in CBRNE emergencies, leveraging XGBoost and LightGBM with sensitivity values above 85%.[151] Defilippo *et al*. employed graph neural networks (GNNs) in 6962 patients with decision-making efficiency and interpretability more than traditional models with almost 10% of accuracy.[149]

For misclassification and errors, two articles suggested AI solutions for reducing over and undertriage. Wyatt *et al*. explored AI's ability to identify subgroups of misclassified patients (overtriage/undertriage) in a multicenter study, revealing that XGBoost performed better in reducing overtriage errors than random forest models.[150] Xu *et al*. developed ML-derived triage tools for major incidents, improving resource allocation and triage efficiency in mass casualty scenarios.[155]

Emergency department overcrowding

Emergency department overcrowding is another complicated issue that requires effective solutions. Although triage systems are designed to classify patients based on limited resources and prioritize those in urgent need of medical attention, they may become insufficient in the excessive demand. Overcrowding, often driven by unnecessary visits, leads to prolonged waiting times in the ED. As a result, the factors contributing to ED overcrowding and its consequences have become key subjects in predictive analyses involving ML.[125,156]

Eight articles were included in the review and were related to overcrowding. Studies were examined to evaluate ED overcrowding, ED visits and revisits, ED length of stay, and factors affecting ED length of stay prediction.[124-128,157-159] Study populations were sufficient to measure the models' performances with AUC, c-index, F1 score, and MAPE values. In the study by Davoudi *et al*., the ML models they developed in predicting the risks of ED visits and hospitalization in 9340 home healthcare patients with heart failure reached an AUC value of 0.89.[124] Haraldsson *et al*. applied a time-to-event ML model for real-time ED overcrowding prediction, using XGB, RF, DL survival analysis techniques with C-index of 0.78.[125] Porto *et al*. leveraged feature engineering with XGBoost, LightGBM, and SVM models, achieving AUC values between 0.78 and 0.88 in ED patient arrival forecasting. In the length of stay prediction.[128] Canellas *et al*. introduced an interpretable ML model for prolonged ED LOS classification, combining random forest, logistic regression, and XGBoost, with an AUC range of 0.75–0.85 in 135,044 patients.[157] Aziz *et al*. developed an ensemble-based (RF and GB) classification system for LOS estimation, outperforming traditional logistic regression models, however, with an AUC of 0.69 (RF), 0.72 (GB).[127]

Other emergency overcrowding studies are focused on patient flow optimization and forecasting models. Peláez-Rodríguez *et al*. utilized clustering and multi-model regression techniques to forecast ED visits with improved short- and long-term accuracy.[126] Lehan *et al*. examined factors contributing to pediatric urgent care demand, employing random forest and linear regression models in 164,660 patient data.[159] Saggu *et al*. implemented DL techniques (GNN, RNN, XGBoost, and Decision Trees) to predict 30-day ED revisits, showing promising low results in early risk identification with 0.65–0.70 AUC results.[158]

According to the overall results of the study, it can be said that ML algorithms show performances between 0.75 and 0.91in evaluating and predicting ED crowding but promising improvements.

Diagnosis and management

ML methods are being studied to enhance and accelerate diagnostic processes in the emergency department, as well as to improve disease management. The seventeen articles regarding diagnosis and management included in the current review were assessed. It was determined that the studies mostly evaluated ML models in the prediction of different diagnoses in the ED, in addition to sepsis, rhythm recognition, and distinguishing challenging diagnoses.[160-176]

Focusing on sepsis and infections, the overall sample size of the studies was sufficient. The largest sample size was the study by Song *et al*., which evaluated the performance of ML models in sepsis diagnosis[170] in a large-scale dataset. Their ML models demonstrated AUC values between 0.68 and 0.93, with XGBoost outperforming other models. Aygun *et al*. introduced an interpretable XGBoost-based sepsis risk model, incorporating Shapley values for feature explanation.[167] Besides sepsis prediction, Chiu *et al*. developed the most successful model for bacteremia prediction with laboratory results, combining ensemble learning, resulting in the highest reported AUC value (0.93) in this diagnosis processes.[176] Flores *et al*. applied random forest and neural networks to urinary tract infection diagnosis, showing that ML-enhanced clinical decision support systems improved diagnostic accuracy compared to traditional methods (AUC 0.81–0.88).[161]

Toprak *et al*. developed the ARTEMIS-POC AI model, which uses high-sensitivity cardiac troponin I data to rule out MI, achieving high NPV (99.96%) and sensitivity (99.68%).[169] Holmstrom *et al*. implemented XGB models to differentiate pulseless electrical activity from ventricular fibrillation, aiding sudden cardiac arrest diagnosis (AUC 0.68–0.72).[166] Chang *et al*. used synthetic minority oversampling techniques (SMOTE) and multiple ML models (RF, SVM, KNN, LR) to predict acute MI risk in chest pain patients, increasing diagnostic sensitivity (AUC 0.63–0.82).[164] Besides AMI, Yilmaz *et al*. leveraged explainable AI models (XGBoost, LASSO, SHAP analysis) to assess hematological indicators in acute heart failure diagnosis, achieving strong interpretability and accuracy.[165]

The reviewed studies reported AUC values ranging from 0.68 to 0.93, with XGBoost and random forest models often outperforming traditional statistical models. However, there was significant variability in model performance based on dataset characteristics, feature selection methods, and validation techniques.

Outcome and risk prediction

Beyond predictions in prehospital processes, triage, and diagnosis, another crucial role of AI in emergency medicine is patient management and survival. The prediction models may guide emergency physicians in clinical decisions, improving patient outcomes, and optimizing resource use. However, their effectiveness depends on careful model development, validation, and consideration of methodological challenges to ensure accurate and clinically useful predictions. Treatment effects may impact the ability to identify high-risk patients and direct intervention.[177]

Thirty articles related to outcome and risk prediction were examined for inclusion in the current review. The majority of the studies evaluated the performance of ML models in mortality prediction and risk stratification, achieving AUC values ranging from 0.75 to 0.97 regarding hospitalization and ICU admission, and long-term risk prediction, and early clinical deterioration prediction.[89,90,171,178-204]

Several studies, such as Rahmatinejad *et al*. and Jawad *et al*., demonstrated the superiority of ensemble learning models over traditional logistic regression in mortality prediction, achieving AUROC values above 0.83.[178,200] Ding *et al*. and Shashikumar *et al*. successfully implemented XGBoost and DL models for intubation and physiological deterioration detection, showing high sensitivity and specificity.[181,185] In addition, Richards *et al*. developed an ML-based Coagulation Risk Index, outperforming traditional INR-based assessments with an AUROC of 0.97.[180]

Despite these advances, several challenges remain, including data imbalance issues, as observed in Park *et al*., which required external validation due to dataset variability.[188] Similarly, Hinson *et al*. highlighted the need for prospective validation, as most models were trained on retrospective datasets, limiting real-world implementation.[195] Gauss *et al*. further emphasized interpretability concerns, noting that while SHAP-based feature explanations improved model transparency, DL models in hemorrhage prediction of trauma patients.[198]

The sample sizes and method selection of the studies were compatible with the data sets. Across these studies, AUC values ranged from 0.75 to 0.97, with ensemble learning models (XGBoost, Random Forest, AdaBoost) and DL techniques outperforming traditional logistic regression-based models. However, some studies had dataset imbalance issues, requiring data augmentation (e.g. SMOTE)[193,201] and multi-site validation to improve reliability, while some models used explainable AI techniques (SHAP, LIME),[171,181,188,195,198] DL models remain black-box systems, posing barriers to clinician adoption. The developed ML models achieved more successful results than classical methods.

Patient safety

Six studies included in the review were evaluated. It was determined that the studies were on predicting ED revisits, anticoagulation type, pressure injury risk, medication-associated ED visits, and leaving against medical advice (AMA) patients.[205-210]

Wei *et al*. developed ML-based pressure injury prediction models using logistic regression, decision trees, and neural networks, achieving AUC values ranging from 0.944 to 0.959, indicating high predictive accuracy.[205] Seger *et al*. introduced the FeelBetter ML system to stratify medication-related risks, reporting odds ratios (ORs) of 7.9 for ED visits and 17.3 for hospitalizations, demonstrating its potential in identifying high-risk patients before adverse events occur.[206]

Ahmed *et al*. studied a quality indicator by applying an XGBoost model with adaptive optimization to predict patients leaving AMA, achieving an AUC of 0.76 and a sensitivity of 82%.[207] Hsu *et al*. developed ML models for predicting 72-h unscheduled return visits, comparing logistic regression, random forest, and DL models.[209]

Fujiwara *et al*. created an ML-based model to predict anticoagulant use in elderly trauma patients, with AUC values of 0.88 for direct oral anticoagulants (DOACs) and 0.96 for Vitamin K antagonists (VKAs), demonstrating high accuracy in medication selection.[208]

Across these studies, AUC values ranged from 0.71 to 0.96, with random forests, XGBoost, and logistic regression being the most frequently used models. ML systems are also promising for medication safety, and emergency return visits, potentially improving patient outcomes.

Emergency Medicine Education

With the frequent use of AI and LLM in daily life, the use of AI in medical education is also on the agenda. Studies on the use of AI in medical education have been increasingly on the rise over the past 20 years.[211,212]

In the development of medical education, determining the learning styles and habits of medical students and trainees undergoing specialization training, and developing educational approaches in line with these identified needs, holds significant importance.[213] The standout feature of AI in the integration into medical education is its potential to offer personalized, adaptive learning experiences.[212] By providing content and feedback tailored to medical students' individual learning styles and habits, AI-powered personalized learning systems can optimize study efficiency, such as literature search and study planning. In this way, students can devote the time saved to in-depth learning of medical concepts and practices.[213]

After the systematic search, sixteen studies on the use of AI in emergency medicine education were found [Supplementary Table 5] (https://turkjemergmed.com/pages/2025-2-issue-supplementary-files). The full texts of three studies could not be accessed, and only one study was excluded from the review due to foreign language (German). The thirteen studies included in the review were classified according to the possible areas of use of AI in medicine and specifically in emergency medicine education [Supplementary Table 5] (https://turkjemergmed.com/pages/2025-2-issue-supplementary-files). AI models are widely used in the field of EM education. Since OpenAI's ChatGPT announced in 2022, the studies in this domain progressively increased in educational use. Thus, nearly all the models used in this review are LLMs, we categorized the studies according to educational use. A total of five studies focused on Evaluation and Feedback Systems, two studies on Simulation-Based Learning, Serious Games and Gamification, one study on Educational Content Development and Effectiveness Analysis, two studies on Skills Assessment and Video Analysis, one study on Planning and Management of Educational Programs, and one study on NLP and Educational Evaluations.

Simulation-based learning, serious games, and gamification

The first of the studies classified under the title of Simulation-Based Learning, Serious Games and Gamification is Aster *et al*.'s work on developing an emergency department simulation game called Digitale Virtuelle Notaufnahme (DIVINA) to improve medical students' clinical reasoning skills and investigating the usability and user experience of this game.[214] The game was developed in a multidisciplinary way with the collaboration of software developers, physicians, and students who are potential users. It is stated that a virtual patient generator, a chatbot used to take medical history, and virtual patient faces developed with AI were used for the game. The study shows that DL related generative tools such as Generative Adversarial Network (StyleGAN) can be used for visual representations of virtual patients to ensure data privacy.[214] The other study evaluated within the classification is the one conducted by Duggan *et al*., which investigates whether the gamified crowdsourcing labeling method is a suitable approach for creating POCUS datasets for ML models.[215] The other study evaluated within the same classification is the one conducted by Duggan *et al*., which investigates whether the gamified crowdsourcing labeling method is a suitable approach for creating POCUS datasets for ML models.[215] Although this study did not directly focus on medical education, its findings suggest that gamified crowdsourcing methods may contribute to the development of high-quality datasets, which are essential for ML-supported tools in POCUS training.

Assessment and feedback models

The first study under the classification of Evaluation and Feedback Systems is by Spadafore *et al*., which evaluates the quality of narrative assessment comments used to measure students' performance and progress in competency-based medical education using NLP.[216] In the study, it is stated that narrative comments are currently evaluated using the Quality of Assessment for Learning (QuAL); the aim is to evaluate this time-consuming method quickly and efficiently using a ML method like NLP. A total of 2500 evaluation comments from two emergency medicine residency programs were scored using QuAL by 50 raters, and this dataset was used to train the NLP model. The developed model reportedly predicts the QuAL score with high accuracy and effectively identifies comments lacking improvement suggestions.[216] The successful results of the study promise new methods for analyzing and evaluating student development. The authors' sharing of the model they developed as open source not only ensures the reproducibility of the results but also serves as an example for models to be developed for future emergency medicine education assessments. Shamim *et al*. conducted a study examining the use of AI in evaluating essay-type questions in medical education.[217] The authors manually evaluated and graded 10 short formative essays given to final-year dental students and compared the grading using Chat Generative Pre-training Transformer (ChatGPT) 3.5. Unfortunately, the authors did not share the results, stating that the responses were recorded and compared with manual grading, so there are no conclusions about the detailed analysis provided by ChatGPT and the reliability and consistency of the system. Moreover, the possible benefits directly to emergency medicine education could not be evaluated. It is seen that the authors additionally emphasized the potential of using AI in the evaluation of essay-type questions.[217] In another study evaluating the performance of ChatGPT as an example of LLM in emergency medicine residency exams in Qatar and comparing the performance of residents, AI performance on multiple choice question (MCQ) format exams was assessed.[218] Between October 2021 and September 2022, the results of five different examinations applied to emergency medicine residents (Post Graduate Year - PGY1 to PGY4) were collected, and the same MCQ questions from these exams were asked to ChatGPT 4.0 (paid version) in May 2023, and performance comparison was performed. In the study, it was found that ChatGPT achieved a higher mean score (25.8 ± 2.6) than all resident groups; the mean scores of the residents increased according to the PGY level (PGY1 18 ± 3.5, PGY2 19.4 ± 3.2, PGY3 21.1 ± 3.8 and PGY4 21.9 ± 4.2)[218] However, the limitations of the study include the fact that the data were collected from a single institution, only multiple-choice questions were used, short-answer questions or clinical skills exams were not included, and questions containing images were transcribed and evaluated. The research indicates that AI, specifically ChatGPT, exhibits significant theoretical competence in emergency medicine examinations. The authors emphasize its potential as a supplementary resource in medical education; however, additional research is required to assess its relevance in more complicated, practice-oriented training scenarios.[218] Another study is Misra *et al*.'s perspective-type study examining the integration of ChatGPT in the objective structured clinical examinations (OSCE) process.[219] It was emphasized that OSCEs are a time- and resource-intensive process for educators and that ChatGPT can create significant efficiency by contributing to the preparation of educational content and assessments. The study also included opinions on the potential uses of ChatGPT in OSCE rubrics, case preparation, and standardized patient (SP) creation.[219] In the study, an example of checklist preparation was created using ChatGPT, and random responses were given by the authors and ChatGPT was asked to evaluate the responses and give feedback.[219] However, there is no verification of the checklist, comparison with existing checklists, consistency and repetition of the assessment with real-life examples. In a study analyzing the competition levels of standardized letters of evaluation (SLOEs) used during Emergency Medicine residency applications, Schnapp *et al*. examined the potential of AI-based LLM (LLMs), specifically ChatGPT, in this process.[220] Analyses using ChatGPT-4o based Julius AI (Caesar Labs, Inc.) demonstrated a strong correlation with faculty members' rankings of SLOEs ($r = 0.96$).[220] However, the AI primarily relied on rating scales and often overlooked narrative data, even when given additional prompts to incorporate it.[220] Notably, when explicitly directed to focus on narrative elements, the model adjusted its assessment, though this led to a lower correlation with faculty consensus ($r = 0.89$).[220] This indicates that although LLMs perform well in structured, quantitative assessments, they may need clear direction to effectively incorporate qualitative elements. Their strength appears to lie in large-scale, objective data analysis rather than comprehensive human-like assessment.

Skills assessment and video analysis

The first of the studies we categorized as Skills Assessment and Video Analysis is the study by Wang *et al*. which examined the accuracy and reliability of ChatGPT-4o's assessment of CPR skills exams through video recordings.[221] In CPR skill examinations, due to the potential subjectivity in certain parameters (such as chest compression depth, and chest rise during ventilation) and the possibility of evaluators' attention being negatively affected during long exams, the authors have stated that they considered the use of AI to prevent potential human errors.[221] While evaluating the video, ChatGPT was asked to score different CPR skills such as patient assessment, chest compressions, rescue breaths, and repeated operations. The scores obtained were compared with those of the expert raters. In the study conducted on 103 students' skills test videos, it was reported that the ChatGPT-4o model gave scores closer to the evaluations of senior experts, and ChatGPT-4o had higher accuracy rates in the areas of patient assessment and rescue breathing.[221] Expert evaluators were also asked to rate the LLM scores on a Likert scale, and it was concluded that GPT-4o showed consistency with the evaluation results and was reliable.[221] The study, which suggests that the use of AI in objective video analysis can be useful, gives an idea that computer vision methods can be useful, especially that evaluation processes can be accelerated by giving consistent results. Another study on skill assessment and video analysis is the study by Huang *et al*. which examines the development of a training evaluation system called SmartCPR, which was developed using the human pose estimation technique in CPR training.[222] The system, developed with the MoveNet model in the open-source TensorFlow (Google LLC) library – integrating multiple ML and DL algorithms – is designed to run on Android-based phones. It evaluates compression cycle, depth, frequency, and position to provide real-time feedback.[222] In the study, in which a comparison was made with Resusci Anne QCPR (Laerdal Medical Corp.), it is seen that the performance and effectiveness of the system on real users were not measured, the technical features of the system were compared, and potential advantages were evaluated.[222] From the perspective of emergency medicine education, we can say that even if speculative, AI could be a tool that can be used in CPR training and could have beneficial aspects for learning processes. Especially through mobile devices, we can say that these systems could help make educational processes more accessible in the future.

Planning and management of educational programs

Eskandarani *et al*. address the use of AI in the process of creating annual rotation schedules for emergency medicine residents.[223] The challenges associated with organizing clinical rotations are reported to stem from the need to balance optimal patient care, adequate staffing, and the maximization of residents' educational experiences while also addressing time-sensitive curricular requirements.[223] While the authors emphasize the potential use of LLMs such as ChatGPT and AI agents like task-based AutoGPT, which leverage the APIs of ChatGPT models (e.g., 3.5, 4o) in the preparation of rotation programs, their study primarily describes a manually operated Excel (Microsoft Inc.) system as an example, without further elaborating on AI-based implementations.[223] Given the complexity of such planning scenarios, the use of Computer-Interpretable Guidelines (CIGs) may offer a more effective approach for AI-driven implementation.

Johnson *et al*. explore the application of NLP techniques in educational assessments to analyze the sentiments of residents and faculty members toward Entrustable Professional Activity (EPA) evaluations.[224] EPAs are assessment tools designed to determine residents' competence in patient care, and the study indicates that residents generally associate these evaluations with negative emotions.[224] Using Sentiment Analysis (SA), one of the NLP methods, the researchers aimed to quantitatively analyze the emotions of the residents and faculty members regarding this measurement tool and to determine the emotional differences between different groups (gender, specialty, etc.).[224] Participants from the fields that include pediatrics, general surgery, and emergency medicine were asked to answer standardized questions as well as open-ended questions about their feelings about the EPA assessment and the factors affecting it.[224] The authors report that 91 respondents answered the survey, 73 respondents answered the open-ended question, and data from a total of 66 participants (30 faculty and 26 residents) were considered usable.[224] Using the National Research Council Canada (NRC) Emotion Lexicon, the frequency of words categorized as positive in the texts was analyzed, and the differences between the specified groups were compared.[224] In the group evaluation, it was observed that the frequency of positive words used varied according to the specialty. It was reported that the highest use of positive words was observed in pediatrics, and the lowest use of positive words was observed in general surgery.[224] Of course, in the article, a definitive result cannot be obtained because the evaluation was made only on the frequency of words without sentence context. Nevertheless, it points to the usability of NLP methods in EPA assessment and emerges as an area of study to be repeated in other training processes.

Educational content development and effectiveness analysis

Karnan *et al*.'s study of the effectiveness of educational materials used for patients developed by AI, which we classified in Educational Content Development and Effectiveness Analyses, gives an idea about whether materials such as informed consent and discharge recommendations, which are frequently used in emergency medicine, can be developed by AI.[225] ChatGPT 3.5 and Google Gemini (Google Inc.) have compared patient education materials produced on topics such as mammography screening, claustrophobia during MRI, and MRI safe/unsafe items.[225] When the texts were evaluated for scientific reliability (Modified DISCERN score), originality (QuillBot-Learneo, Inc.), and ease of readability (Flesch-Kincaid Calculator), both LLMs showed similar average performance in terms of scientific reliability. The similarity percentage was 0.5% in texts generated by ChatGPT and 9.43% in those produced by Google Gemini. In addition, ChatGPT-generated texts had a higher ease of readability score, though the difference was not statistically significant ($P = 0.1102$, $P < 0.05$).[225] Although it is uncertain how the results of this study, conducted in April 2024, would be affected by the newly introduced models, its importance lies in the preparation of AI-generated documents that meet quality standards for both patient-related materials and other informational content. In addition, future studies should focus on assessing the extent to which AI-generated patient education materials align with established scientific knowledge, ensuring their accuracy and credibility in clinical practice.

Guidelines on medical education

A guideline that was not included in our review with our search query, but which we would like to mention because it is noteworthy, is the last of the Best Evidence Medical Education (BEME) guidelines[226] published by The International Association for Health Professions Education (AMEE), which provides a framework for creating more effective and efficient learning environments in medical education and adopts an evidence-based approach. In the 84th guideline of the BEME, which also provides an evidence-based and evidence-based approach to emerging AI studies and examines the role of AI in medical education, it states that the majority (48.6%) of studies involving AI-based medical education practices are on undergraduate medical education, followed by graduate medical education and continuing professional development (22.3% and 2.5%, respectively), and that the majority of publications (68.7%) are about articles and innovations. In these articles and innovations studies, again, the largest number of publications were about studies involving knowledge and attitudes about AI ($n = 51$, 26.7%), followed by assessment of learning ($n = 50$, 26.2%).[212] Assessment of learning includes assessment of clinical skills and surgical/procedural skills.[212] It has been reported that 32 studies focused on evaluating LLM performance in examinations, while 19 examined performance analytics, 11 investigated Virtual Patient Simulators, and 10 explored clinical guidelines for residents, such as Decision Support Systems on evaluating the studies referenced in the guideline from the perspective of emergency medicine, it is noted that there are two direct studies regarding emergency medicine and two indirect studies that assess procedural skills in laryngoscopy use.[227-230] Since the literature in the guideline is relatively limited in terms of emergency medicine, the aims and findings of the studies are briefly summarized.

The first of the studies, which is directly related to Emergency Medicine, evaluates whether ChatGPT can be used as a tool to teach bad news reporting skills to emergency physicians.[227] For this purpose, a detailed prompt was used with the ChatGPT-3.5 model, specifying the rules it needed to follow, and the SPIKES framework (Setting up, Perception, Invitation, Knowledge, Emotions with Empathy, and Strategy or Summary) was employed as the assessment method.[228] In the study, it was found that the model can design an appropriate scenario, give feedback to the user in the role of a physician, and evaluate user performance.[228]

In another directly related study, Yilmaz *et al*. evaluated whether comment data obtained through workplace-based assessment (WBA) using NLP and ML applications could assist educators in identifying trainees who are at risk.[228] This retrospective study examined WBA data from September 2012 to July 2018 to determine whether NLP and ML applications could assist educators in identifying at-risk trainees – those who failed to meet expected competency levels or adequately perform assigned tasks.[219] Detecting such trainees was highlighted as crucial for enhancing patient safety, assessing training program efficacy, and ensuring efficient resource utilization, though it also posed a substantial workload for faculty members. The free-text narrative comments written by the faculty members were converted into quantitative data using the bag-of-n-grams technique, which works by counting the frequencies of words or groups of words (n-grams), and these data were analyzed with ML models to identify trainees at risk.[228] These data were subsequently analyzed using ML models, with findings indicating that bigram-based models demonstrated 86.9% accuracy in detecting low-performing trainees, and were suggested as a potential decision-support tool for faculty in assessing trainee performance.[228]

Among the studies involving the use of laryngoscopy and assessment of procedural competencies, Choi *et al*. aimed to determine which of four different laryngoscopes (Macintosh, McGrath, Pentax Airway-Scope), including the A-LRYNGO, a channel-type video laryngoscope with an integrated AI-assisted glottis guidance system, was suitable for intubation training for medical students who were novices and inexperienced in the use of personal protective equipment (PPE).[229] In a randomized, simulation manikin study, the groups were compared based on intubation time, success rate, and posttest short questionnaire with a short posttest questionnaire, administered both before and after the intervention. In this study of 30 senior medical students, participants were tested twice: once after the lecture and again following the hands-on workshop, and the findings indicated that intubation success with channel-type video laryngoscopes increased after the hands-on workshop, while the AI-assisted video laryngoscope showed 93.1% accuracy.[229]

In the study by Zhao *et al*., which examined the use of automated systems in the evaluation of neonatal endotracheal intubation training, it was emphasized that current training is conducted on mannequins and assessed by expert instructors. However, due to the limited number of expert instructors, pediatric trainees have restricted opportunities for adequate practice.[230] They reported that the sensor-based, computer-aided systems used to overcome these limitations are inadequate in analyzing complex movements, recognizing critical directions, and providing accurate feedback.[230] In the study, kinematic multivariate time series (MTS) data – including rotation, position, and velocity – collected from electromagnetic sensors attached to laryngoscopes and mannequins were processed using a dilated CNN. Motion patterns were then visualized as heat maps through Class Activation Mapping.[230] Thus, the study aimed to provide meaningful feedback to trainees by identifying movements with significant impact. The performance of the CNN model, trained on 190 intubation attempt datasets from 44 subjects, was evaluated using the Leave-One-Out Cross-Validation method. The findings reported a high accuracy rate (92.2%) and reliable outcomes, highlighting the need for further studies to facilitate the integration of this model into computer-aided training systems.[230]

Discussion

In this scoping review, studies conducted in the last year on emergency department patient care and emergency medicine education have been examined. It explores the use of AI subfields such as image processing, natural language processing, signal processing, and text mining in various areas of emergency medicine, including triage, diagnosis, outcomes, risk analysis, and education. The findings suggest that studies on the application of AI subfields in emergency medicine show promising potential. However, each method has its own unique characteristics, specific areas of application, and inherent limitations.

AI has the potential to enhance medical imaging processes in emergency medicine. AI models can automate routine tasks, facilitate early disease detection, and accelerate decision-making by assisting radiologists and clinicians without formal radiology training. AI-supported imaging tools significantly reduce interpretation time and improve decision-making efficiency in emergency departments. Computer-aided detection (CADe) and diagnosis (CADx) systems automatically highlight pathologies such as fractures, lung diseases, and neurological disorders, thereby saving valuable time for physicians.

Additionally, AI-based imaging systems provide substantial support in regions with a shortage of experienced radiologists. Studies have demonstrated that AI-assisted radiographs enhance sensitivity and specificity in detecting conditions such as fractures, lung nodules, and ischemic strokes. These systems improve the efficiency of healthcare services by increasing diagnostic accuracy, particularly in resource-limited settings.

AI-driven segmentation and classification models streamline the diagnostic process by minimizing human errors in image interpretation. For instance, AI applications in USG imaging can rapidly assess cardiac function, aiding in the management of critically ill patients. With the increasing integration of automation, clinicians can make faster and more precise decisions, optimizing patient care pathways.

Furthermore, AI integrates medical imaging with patient data to provide comprehensive diagnostic insights. AI systems that function in conjunction with electronic health records (EHRs) can detect conditions such as acute heart failure and sepsis at an early stage, enabling the development of personalized treatment plans. These multimodal AI approaches play a crucial role in the future of medicine by offering a more holistic evaluation of patients' health conditions.

AI holds great potential for medical image processing, but several significant challenges remain in this field. Medical images vary due to factors such as low resolution, artifacts, and differences in imaging devices. While large, high-quality datasets are essential for AI models to achieve high accuracy, the lack of standardization across data from different institutions presents a major obstacle.

Moreover, AI models trained on specific datasets may not perform as expected when applied to diverse patient populations and imaging techniques. Variations in imaging devices and patient demographics can impact model accuracy and reliability. Challenges related to model robustness and generalizability remain key barriers to the widespread adoption of AI in clinical settings.

Another critical issue is the interpretability of DL-based systems, which are often perceived as "black boxes." The opacity of AI decision-making processes makes it difficult for clinicians to fully understand and trust these systems. Enhancing interpretability is essential to increase clinician confidence and facilitate the integration of AI into routine medical practice.

The integration of AI into existing clinical workflows also presents logistical challenges. AI tools that are not designed to seamlessly interact with hospital information systems often require additional infrastructure and significant computational resources, limiting their usability – particularly in smaller healthcare facilities. This is one of the factors delaying the widespread adoption of AI technology in medicine.

Furthermore, the implementation of AI-based medical imaging tools raises ethical concerns related to patient privacy, data security, and algorithmic bias. Regulatory bodies like the FDA require rigorous validation before approving AI-driven diagnostic tools for clinical use. While these regulatory measures enhance reliability, they also slow the transition of AI innovations from research to clinical practice.

Furthermore, this 1-year-review particularly reflects the rapid progress and competition in NLP and LLM. Although the most important problem of NLP is the complex structure of the language itself, this problem has been largely solved with the advancement of the concept of ontology in health data, but when language models were released for distribution in recent years, this issue also provided feature extraction and reasoning with higher and faster models.

Although studies on LLM, especially in emergency medicine, are primarily conducted on sample scenarios determined by experts to determine the accuracy of the LLM, studies using real patient data are also increasing today. This situation has led to the need for data to be entered correctly into electronic health records.

Although minimizing the need for structuring the data seems advantageous, the fact that the training rules of large databases of LLM can be affected by external factors necessitates the need to include customized tools for health data. In general, LLM-based studies offer significant potential in the emergency department environment. Models such as GPT-4 and BioClinicalBERT have been found to be higher performance than NLP studies.

Most of the studies have been conducted with retrospective data analysis; thus, the development of systems modeled with real-time data streams is important to make clinical applications more reliable. More prospective and larger-scale studies are needed to understand how LLMs can be used more effectively in medical decision-support systems.

ML methods are used to overcome existing standard clinical decision support systems and develop new prediction models. These prediction models have the potential to assist emergency physicians in decision-making. Considering the breadth and diversity of the field of emergency medicine, the use of ML models in emergency medicine practice is an opportunity that cannot be ignored. Hidden patterns that will contribute to emergency medicine patient care in meticulously obtained data sets can be revealed with ML models and used in patient care. Besides the issues mentioned on image and text processing; as a result of examining ML models with structured data, it was determined that most studies were aimed at making predictions in different datasets for various outcomes and diagnoses. However, it should not be forgotten that all these prediction models were created with data obtained from existing data sets. Ultimately, the ML model's performance also depends on the data in the dataset. Thus, the accuracy of the structured data, complete and error-free recording, and meticulous preprocessing are the main factors in the success or failure of the models.

Although AI-driven triage systems exhibit strong predictive power, concerns remain regarding model bias, and integration challenges. Many models are lack of adaptability to real-time environment, limiting their deployment in as triage in high-acuity emergency settings. Future studies should focus on external validation across diverse populations and interpretable AI models to enhance clinician acceptance is that each population is unique, and the results obtained are valid for that population. Its validity for different populations needs to be confirmed by external validation studies.

On the view of emergency medicine education, current research on the use of AI in emergency medicine education largely consists of proof-of-concept studies, often assessing AI models – particularly LLM – through standardized tests. The prevalence of small-scale, non-randomized, and single-institution studies limits the ability to draw broad conclusions, making it difficult to determine AI's actual role beyond initial feasibility testing. Like many emerging technologies, AI is frequently portrayed as a game-changing solution to a variety of challenges, including those in medical education. However, having a powerful tool at hand does not mean it should be applied indiscriminately – a perspective well summarized by the saying, "If you only have a hammer, you tend to see every problem as a nail." While AI-based tools, including LLMs and other ML approaches, have the potential for improving certain aspects of medical education, their adoption should be driven by solid evidence and genuine educational needs, rather than a default inclination to incorporate AI into every possible domain.

In conclusion, AI models are evolving and gaining significant potential across multiple areas in emergency medicine, such as triage, diagnosis, and outcome prediction. However, mostly faced challenges such as data variability, model generalizability, and integration into clinical workflows. Rapid updating of versions requires that the results in the literature progress at the same pace. With the continuous refinement of models, better data quality shows promising results within emergency care practice and emergency medicine education.

Acknowledgment(s)
Fatih Doganay, acknowledges a Postdoctoral grant from the Scientific and Technical Research Council of Turkey (TUBITAK, 2219 - International Postdoctoral Research Scholarship Programme, 1059B192302217).

Author contributions statement
AK, FD, BI: Conceptualization, methodology, investigation, resources, data curation, and writing – Original draft. GBB: Conceptualization, methodology, investigation, resources, data curation, writing – Original draft, review and editing, and supervision. All authors approved the last version of the manuscript.

Conflicts of interest
None Declared.

Ethical approval
Not applicable.

Funding
None.

References

1. Pinto-Coelho L. How artificial intelligence is shaping medical imaging technology: A Survey of innovations and applications. Bioengineering (Basel) 2023;10:1435.
2. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJ. Artificial intelligence in radiology. Nat Rev Cancer 2018;18:500-10.
3. Drukker K, Chen W, Gichoya J, Gruszauskas N, Kalpathy-Cramer J, Koyejo S, *et al.* Toward fairness in artificial intelligence for medical image analysis: Identification and mitigation of potential biases in the roadmap from data collection to model deployment. J Med Imaging (Bellingham) 2023;10:061104.
4. Babar M, Qureshi B, Koubaa A. Investigating the impact of data heterogeneity on the performance of federated learning algorithm using medical imaging. PLoS One 2024;19:e0302539.
5. Chang Q, Yan Z, Zhou M, Qu H, He X, Zhang H, *et al.* Mining multi-center heterogeneous medical data with distributed synthetic learning. Nat Commun 2023;14:5510.
6. Masoudi S, Harmon SA, Mehralivand S, Walker SM, Raviprakash H, Bagci U, *et al.* Quick guide on radiology image pre-processing for deep learning applications in prostate cancer research. J Med Imaging (Bellingham) 2021;8:010901.

7. Singh NT, Kaur C, Chaudhary A, Goyal S. Preprocessing of Medical Images using Deep Learning: A Comprehensive Review. In: 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS). IEEE; 2023. p. 521-7.
8. Krenzer A, Makowski K, Hekalo A, Fitting D, Troya J, Zoller WG, *et al.* Fast machine learning annotation in the medical domain: A semi-automated video annotation tool for gastroenterologists. Biomed Eng Online 2022;21:33.
9. Galbusera F, Cina A. Image annotation and curation in radiology: An overview for machine learning practitioners. Eur Radiol Exp 2024;8:11.
10. Jawdekar A, Dixit M. A review of image enhancement techniques in medical imaging. In: Agrawal S, Kumar Gupta K, Chan JH, Agrawal J, Gupta M, eds. Machine Intelligence and Smart Systems. Algorithms for Intelligent Systems. Singapore, Springer Nature;2021:25 33.
11. Fu S, Zhang M, Mu C, Shen X. Advancements of medical image enhancement in healthcare applications. J Healthc Eng 2018;2018:7035264.
12. Islam SM, Mondal HS. Image Enhancement Based Medical Image Analysis. In: 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE; 2019. p. 1-5.
13. Razmjooy N, Rajinikanth V, editors. Frontiers of Artificial Intelligence in Medical Imaging. Bristol,UK:IOP Publishing;2022.
14. Bengio Y, Goodfellow I, Aeron C. Deep Learning. In: Deep Learning. Cambridge: MIT Press; 2017.
15. Zhou SK, Greenspan H, Davatzikos C, Duncan JS, van Ginneken B, Madabhushi A, *et al*. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. Proc IEEE Inst Electr Electron Eng 2021;109:820-38.
16. Zhang A, Lipton ZC, Li M, Smola AJ. Dive into Deep Learning. Cambridge: Cambridge University Press; 2023.
17. Litjens G, Kooi T, Bejnordi BE, Setio AA, Ciompi F, Ghafoorian M, *et al*. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60-88.
18. Han S, Pool J, Tran J, Dally WJ. Learning Both Weights and Connections for Efficient Neural Networks. In NIPS'15: Proceedings of the 29th International Conference on Neural Information Processing Systems - Volume 1 Cambridge,MA:MIT Press;2015.
19. Chu KC, Yeh CH, Lin JM, Chen CY, Cheng CY, Yeh YQ, *et al.* Using convolutional neural network denoising to reduce ambiguity in X-ray coherent diffraction imaging. J Synchrotron Radiat 2024;31:1340-5.
20. Ait Nasser A, Akhloufi MA. A review of recent advances in deep learning models for chest disease detection using radiography. Diagnostics (Basel) 2023;13:159.
21. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. Annu Rev Biomed Eng 2017;19:221-48.
22. Yu H, Zhang Y, Chu Y. Reduction of metal artifacts in X Ray CT images using a convolutional neural network. In: Müller B, Wang G, editors. Developments in X Ray Tomography XI. SanDiago; CA:SPIE; 2017. p. 30.
23. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, *et al*. N4ITK: Improved N3 bias correction. IEEE Trans Med Imaging 2010;29:1310-20.
24. Kamnitsas K, Ledig C, Newcombe VF, Simpson JP, Kane AD, Menon DK, *et al.* Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal 2017;36:61-78.
25. Zhang Q, Yang D, Zhu Y, Liu Y, Ye X. An optimized optical-flow-based method for quantitative tracking of ultrasound-guided right diaphragm deformation. BMC Med Imaging 2023;23:108.
26. Yu Y, Acton ST. Speckle reducing anisotropic diffusion. IEEE Trans Image Process 2002;11:1260-70.
27. Wang X, Yi J, Guo J, Song Y, Lyu J, Xu J, Yan W, *et al.* A review of image super resolution approaches based on deep learning and applications in remote sensing. Remote Sens 2022;14:5423.
28. Barros B, Lacerda P, Albuquerque C, Conci A. Pulmonary COVID-19: Learning spatiotemporal features combining CNN and LSTM networks for lung ultrasound video classification. Sensors (Basel) 2021;21:5486.
29. Prinster D, Mahmood A, Saria S, Jeudy J, Lin CT, Yi PH, *et al.* Care to explain? AI explanation types differentially impact chest radiograph diagnostic performance and physician trust in AI. Radiology 2024;313:e233261.
30. Ye Q, Wang Z, Lou Y, Yang Y, Hou J, Liu Z, *et al.* Deep learning approach based on a patch residual for pediatric supracondylar subtle fracture detection. Biomol Biomed 2025. [doi: 10.17305/bb.2024.11341].
31. Wang CH, Chang W, Lee MR, Tay J, Wu CY, Wu MC, *et al.* Deep learning-based diagnosis of pulmonary tuberculosis on chest X-ray in the emergency department: A retrospective study. J Imaging Inform Med 2024;37:589-600.
32. Quek JJ, Nickalls OJ, Wong BS, Tan MO. Deploying artificial intelligence in the detection of adult appendicular and pelvic fractures in the Singapore emergency department after hours: Efficacy, cost savings and non monetary benefits. Singapore Med J 2024. [doi: 10.4103/singaporemedj.SMJ 2023 170].
33. Kavak N, Kavak RP, Güngörer B, Turhan B, Kaymak SD, Duman E, *et al.* Detecting pediatric appendicular fractures using artificial intelligence. Rev Assoc Med Bras (1992) 2024;70:e20240523.
34. López Alcolea J, Fernández Alfonso A, Cano Alonso R, Álvarez Vázquez A, Díaz Moreno A, García Castellanos D, *et al.* Diagnostic performance of artificial intelligence in chest radiographs referred from the emergency department. Diagnostics (Basel) 2024;14:2592.
35. Novak A, Ather S, Gill A, Aylward P, Maskell G, Cowell GW, *et al.* Evaluation of the impact of artificial intelligence-assisted image interpretation on the diagnostic performance of clinicians in identifying pneumothoraces on plain chest X-ray: A multi-case multi-reader study. Emerg Med J 2024;41:602-9.
36. Lee CK, Chen TL, Wu JE, Liao MT, Wang C, Wang W, *et al.* Multimodal deep learning models utilizing chest X-ray and electronic health record data for predictive screening of acute heart failure in emergency department. Comput Methods Programs Biomed 2024;255:108357.
37. Ghatak A, Hillis JM, Mercaldo SF, Newbury-Chaet I, Chin JK, Digumarthy SR, *et al.* The potential clinical utility of an artificial intelligence model for identification of vertebral compression fractures in chest radiographs. J Am Coll Radiol 2025;22:220-9.
38. Kumari J, Kumar E, Kumar D. A structured analysis to study the role of machine learning and deep learning in the healthcare sector with big data analytics. Arch Comput Methods Eng 2023;30:1-29.
39. Laletin V, Ayobi A, Chang PD, Chow DS, Soun JE, Junn JC, *et al.* Diagnostic performance of a deep learning-powered application for aortic dissection triage prioritization and classification. Diagnostics (Basel) 2024;14:1877.
40. Kim J, Kwak CW, Uhmn S, Lee J, Yoo S, Cho MC, *et al*. A novel deep learning-based artificial intelligence system for interpreting urolithiasis in computed tomography. Eur Urol Focus 2024;10:1049-54.
41. Lu CY, Wang YH, Chen HL, Goh YX, Chiu IM, Hou YY, *et al.* Artificial intelligence application in skull bone fracture with segmentation approach. J Imaging Inform Med 2025;38:31-46.
42. Xu Y, Fu Q, Qu M, Chen J, Fan J, Hou S, *et al.* Automated hematoma detection and outcome prediction in patients with traumatic brain injury. CNS Neurosci Ther 2024;30:e70119.
43. Liu HH, Chang CB, Chen YS, Kuo CF, Lin CY, Ma CY, *et al.* Automated detection and differentiation of stanford type A and type B aortic dissections in CTA scans using deep learning. Diagnostics (Basel) 2024;15:12.
44. Zhang C, Peng J, Wang L, Wang Y, Chen W, Sun MW, *et al*. A deep

learning-powered diagnostic model for acute pancreatitis. BMC Med Imaging 2024;24:154.
45. Choi SY, Kim JH, Chung HS, Lim S, Kim EH, Choi A. Impact of a deep learning-based brain CT interpretation algorithm on clinical decision-making for intracranial hemorrhage in the emergency department. Sci Rep 2024;14:22292.
46. Kim JY, Choi HJ, Kim SH, Ju H. Improved differentiation of cavernous malformation and acute intraparenchymal hemorrhage on CT using an AI algorithm. Sci Rep 2024;14:11818.
47. Wang Y, Zhang J, Li M, Miao Z, Wang J, He K, *et al.* SMART: Development and application of a multimodal multi organ trauma screening model for abdominal injuries in emergency settings. Acad Radiol 2024:S1076 2. [doi: 10.1016/j.acra.2024.11.056].
48. Ruitenbeek HC, Oei EH, Schmahl BL, Bos EM, Verdonschot RJ, Visser JJ. Towards clinical implementation of an AI-algorithm for detection of cervical spine fractures on computed tomography. Eur J Radiol 2024;173:111375.
49. Warman P, Warman A, Warman R, Degnan A, Blickman J, Smith D, *et al.* Using an artificial intelligence software improves emergency medicine physician intracranial haemorrhage detection to radiologist levels. Emerg Med J 2024;41:298-303.
50. Kim J, Oh SW, Lee HY, Choi MH, Meyer H, Huwer S, *et al.* Assessment of deep learning-based triage application for acute ischemic stroke on brain MRI in the ER. Acad Radiol 2024;31:4621-8.
51. Lang M, Clifford B, Lo WC, Applewhite BP, Tabari A, Filho AL, *et al*. Clinical evaluation of a 2-minute ultrafast brain MR protocol for evaluation of acute pathology in the emergency and inpatient settings. AJNR Am J Neuroradiol 2024;45:379-85.
52. Kim C, Kwon JM, Lee J, Jo H, Gwon D, Jang JH, *et al.* Deep learning model integrating radiologic and clinical data to predict mortality after ischemic stroke. Heliyon 2024;10:e31000.
53. He B, Dash D, Duanmu Y, Tan TX, Ouyang D, Zou J. AI-enabled assessment of cardiac function and video quality in emergency department point-of-care echocardiograms. J Emerg Med 2024;66:184-91.
54. Park S, Yoon H, Yeon Kang S, Joon Jo I, Heo S, Chang H, *et al.* Artificial intelligence-based evaluation of carotid artery compressibility via point-of-care ultrasound in determining the return of spontaneous circulation during cardiopulmonary resuscitation. Resuscitation 2024;202:110302.
55. Ge C, Jang J, Svrcek P, Fleming V, Kim YH. Exploring deep learning applications using ultrasound single view Cines in acute gallbladder pathologies: Preliminary results. Acad Radiol 2025;32:770-5.
56. Holland L, Hernandez Torres SI, Snider EJ. Using AI segmentation models to improve foreign body detection and triage from ultrasound images. Bioengineering (Basel) 2024;11:128.
57. Biousse V, Najjar RP, Tang Z, Lin MY, Wright DW, Keadey MT, *et al.* Application of a deep learning system to detect papilledema on nonmydriatic ocular fundus photographs in an emergency department. Am J Ophthalmol 2024;261:199-207.
58. Li H, Cao J, You K, Zhang Y, Ye J. Artificial intelligence-assisted management of retinal detachment from ultra-widefield fundus images based on weakly-supervised approach. Front Med (Lausanne) 2024;11:1326004.
59. Shobayo O, Saatchi R, Ramlakhan S. Convolutional neural network to classify infrared thermal images of fractured wrists in pediatrics. Healthcare (Basel) 2024;12:994.
60. Wang Y, Ye Y, Shi S, Mao K, Zheng H, Chen X, *et al.* Prediagnosis recognition of acute ischemic stroke by artificial intelligence from facial images. Aging Cell 2024;23:e14196.
61. Zhao L, Vidwans A, Bearnot CJ, Rayner J, Lin T, Baird J, *et al.* Prediction of anemia in real-time using a smartphone camera processing conjunctival images. PLoS One 2024;19:e0302883.
62. Song A, Lusk JB, Roh KM, Hsu ST, Valikodath NG, Lad EM, *et al.* RobOCTNet: Robotics and deep learning for referable posterior segment pathology detection in an emergency department population. Transl Vis Sci Technol 2024;13:12.
63. Choi YJ, Park MJ, Cho Y, Kim J, Lee E, Son D, *et al.* Screening for RV dysfunction using smartphone ECG analysis app: Validation study with acute pulmonary embolism patients. J Clin Med 2024;13:4792.
64. Maxin AJ, Gulek BG, Lim DH, Kim S, Shaibani R, Winston GM, *et al.* Smartphone pupillometry with machine learning differentiates ischemic from hemorrhagic stroke: A pilot study. J Stroke Cerebrovasc Dis 2025;34:108198.
65. Choi J, Kim J, Spaccarotella C, Esposito G, Oh IY, Cho Y, *et al.* Smartwatch ECG and artificial intelligence in detecting acute coronary syndrome compared to traditional 12-lead ECG. Int J Cardiol Heart Vasc 2025;56:101573.
66. Gharehchopogh FS, Khalifelu ZA. Analysis and Evaluation of Unstructured Data: Text Mining Versus Natural Language Processing. In: 2011 5th International Conference on Application of Information and Communication Technologies (AICT). IEEE; 2011. p. 1-4.
67. Lee SJ, Alzeen M, Ahmed A. Estimation of racial and language disparities in pediatric emergency department triage using statistical modeling and natural language processing. J Am Med Inform Assoc 2024;31:958-67.
68. Farhat H, Alinier G, Tluli R, Chakif M, Rekik FB, Alcantara MC, *et al.* Enhancing patient safety in prehospital environment: Analyzing patient perspectives on non-transport decisions with natural language processing and machine learning. J Patient Saf 2024;20:330-9.
69. Levra AG, Gatti M, Mene R, Shiffer D, Costantino G, Solbiati M, *et al*. A large language model-based clinical decision support system for syncope recognition in the emergency department: A framework for clinical workflow integration. Eur J Intern Med 2025;131:113-20.
70. Huang TY, Chong CF, Lin HY, Chen TY, Chang YC, Lin MC. A pre-trained language model for emergency department intervention prediction using routine physiological data and clinical narratives. Int J Med Inform 2024;191:105564.
71. Zhang X, Wang Y, Jiang Y, Pacella CB, Zhang W. Integrating structured and unstructured data for predicting emergency severity: An association and predictive study using transformer-based natural language processing models. BMC Med Inform Decis Mak 2024;24:372.
72. McMurry AJ, Zipursky AR, Geva A, Olson KL, Jones JR, Ignatov V, *et al.* Moving biosurveillance beyond coded data using AI for symptom detection from physician notes: Retrospective cohort study. J Med Internet Res 2024;26:e53367.
73. Chang YH, Lin YC, Huang FW, Chen DM, Chung YT, Chen WK, *et al.* Using machine learning and natural language processing in triage for prediction of clinical disposition in the emergency department. BMC Emerg Med 2024;24:237.
74. Choi DH, Choi SW, Kim KH, Choi Y, Kim Y. Early identification of suspected serious infection among patients afebrile at initial presentation using neural network models and natural language processing: A development and external validation study in the emergency department. Am J Emerg Med 2024;80:67-76.
75. Patel D, Timsina P, Gorenstein L, Glicksberg BS, Raut G, Cheetirala SN, *et al.* Traditional machine learning, deep learning, and BERT (Large Language Model) approaches for predicting hospitalizations from nurse triage notes: Comparative evaluation of resource management. JMIR AI 2024;3:e52190.
76. Hartman V, Zhang X, Poddar R, McCarty M, Fortenko A, Sholle E, *et al.* Developing and evaluating large language model-generated emergency medicine handoff notes. JAMA Netw Open 2024;7:e2448723.
77. Seo H, Ahn I, Gwon H, Kang HJ, Kim Y, Cho HN, *et al*. Prediction of hospitalization and waiting time within 24 hours of emergency

department patients with unstructured text data. Health Care Manage Sci 2024;27:114 29.

78. Kuo KM, Lin YL, Chang CS, Kuo TJ. An ensemble model for predicting dispositions of emergency department patients. BMC Med Inform Decis Mak 2024;24:105.
79. Hughes JA, Wu Y, Jones L, Douglas C, Brown N, Hazelwood S, *et al.* Analyzing pain patterns in the emergency department: Leveraging clinical text deep learning models for real-world insights. Int J Med Inform 2024;190:105544.
80. Weidman AC, Sedor Schiffhauer Z, Zikmund C, Salcido DD, Guyette FX, Weiss LS, *et al.* Words to live by: Using medic impressions to identify the need for prehospital lifesaving interventions. Acad Emerg Med 2025;00:1-10. [doi: 10.1111/acem.15067].
81. Watson M, Boulitsakis Logothetis S, Green D, Holland M, Chambers P, Al Moubayed N. Performance of machine learning versus the national early warning score for predicting patient deterioration risk: A single-site study of emergency admissions. BMJ Health Care Inform 2024;31:e101088.
82. Chai C, Peng SZ, Zhang R, Li CW, Zhao Y. Advancing emergency department triage prediction with machine learning to optimize triage for abdominal pain surgery patients. Surg Innov 2024;31:583-97.
83. Khademi S, Palmer C, Javed M, Dimaguila GL, Clothier H, Buttery J, *et al.* Near real-time syndromic surveillance of emergency department triage texts using natural language processing: Case study in febrile convulsion detection. JMIR AI 2024;3:e54449.
84. Huang T, Socrates V, Gilson A, Safranek C, Chi L, Wang EA, *et al.* Identifying incarceration status in the electronic health record using large language models in emergency department settings. J Clin Transl Sci 2024;8:e53.
85. Zheng Y, Cai Y, Yan Y, Chen S, Gong K. Novel approach to personalized physician recommendations using semantic features and response metrics: Model evaluation study. JMIR Hum Factors 2024;11:e57670.
86. Boley S, Sidebottom A, Vacquier M, Watson D, Van Eyll B, Friedman S, *et al.* Racial differences in stigmatizing and positive language in emergency medicine notes. J Racial Ethn HealthDisparities 2024. [doi: 10.1007/s40615 024 02080 3].
87. Landau AY, Blanchard A, Kulkarni P, Althobaiti S, Idnay B, Patton DU, *et al.* Harnessing the power of machine learning and electronic health records to support child abuse and neglect identification in emergency department settings. Stud Health Technol Inform 2024;316:1652-6.
88. Abedi V, Misra D, Chaudhary D, Avula V, Schirmer CM, Li J, *et al.* Machine learning-based prediction of stroke in emergency departments. Ther Adv Neurol Disord 2024;17:17562864241239108.
89. Pandey D, Jahanabadi H, D'Arcy J, Doherty S, Vo H, Jones D, *et al.* Early prediction of intensive care unit admission in emergency department patients using machine learning. Aust Crit Care 2024;38:101143.
90. Akhlaghi H, Freeman S, Vari C, McKenna B, Braitberg G, Karro J, *et al.* Machine learning in clinical practice: Evaluation of an artificial intelligence tool after implementation. Emerg Med Australas 2024;36:118-24.
91. Paslı S, Şahin AS, Beşer MF, Topçuoğlu H, Yadigaroğlu M, İmamoğlu M. Assessing the precision of artificial intelligence in ED triage decisions: Insights from a study with ChatGPT. Am J Emerg Med 2024;78:170-5.
92. Colakca C, Ergın M, Ozensoy HS, Sener A, Guru S, Ozhasenekler A. Emergency department triaging using ChatGPT based on emergency severity index principles: A cross-sectional study. Sci Rep 2024;14:22106.
93. Haim GB, Braun A, Eden H, Burshtein L, Barash Y, Irony A, *et al.* AI in the ED: Assessing the efficacy of GPT models versus physicians in medical score calculation. Am J Emerg Med 2024;79:161-6.
94. Liu X, Lai R, Wu C, Yan C, Gan Z, Yang Y, *et al.* Assessing the utility of artificial intelligence throughout the triage outpatients: A prospective randomized controlled clinical study. Front Public Health 2024;12:1391906.
95. Hoppe JM, Auer MK, Strüven A, Massberg S, Stremmel C. ChatGPT with GPT-4 outperforms emergency department physicians in diagnostic accuracy: Retrospective analysis. J Med Internet Res 2024;26:e56110.
96. Haim GB, Saban M, Barash Y, Cirulnik D, Shaham A, Eisenman BZ, *et al*. Evaluating large language model assisted emergency triage: A comparison of acuity assessments by GPT 4 and medical experts. J Clin Nurs 2024;0:1-7. [doi: 10.1111/jocn.17490].
97. Arslan B, Nuhoglu C, Satici MO, Altinbilek E. Evaluating LLM-based generative AI tools in emergency triage: A comparative study of ChatGPT plus, Copilot pro, and triage nurses. Am J Emerg Med 2025;89:174-81.
98. Rosen S, Saban M. Evaluating the reliability of ChatGPT as a tool for imaging test referral: A comparative study with a clinical decision support system. Eur Radiol 2024;34:2826-37.
99. Woo KC, Simon GW, Akindutire O, Aphinyanaphongs Y, Austrian JS, Kim JG, *et al.* Evaluation of GPT-4 ability to identify and generate patient instructions for actionable incidental radiology findings. J Am Med Inform Assoc 2024;31:1983-93.
100. Wang X, Ye S, Feng J, Feng K, Yang H, Li H. Performance of ChatGPT on prehospital acute ischemic stroke and large vessel occlusion (LVO) stroke screening. Digit Health 2024;10:20552076241297127.
101. Amacher SA, Arpagaus A, Sahmer C, Becker C, Gross S, Urben T, *et al.* Prediction of outcomes after cardiac arrest by a generative artificial intelligence model. Resusc Plus 2024;18:100587.
102. Shekhar AC, Kimbrell J, Saharan A, Stebel J, Ashley E, Abbott EE. Use of a large language model (LLM) for ambulance dispatch and triage. Am J Emerg Med 2025;89:27-9.
103. Williams CY, Zack T, Miao BY, Sushil M, Wang M, Kornblith AE, *et al.* Use of a large language model to assess clinical acuity of adults in the emergency department. JAMA Netw Open 2024;7:e248895.
104. Choi DH, Kim Y, Choi SW, Kim KH, Choi Y, Shin SD. Using large language models to extract core injury information from emergency department notes. J Korean Med Sci 2024;39:e291.
105. Glicksberg BS, Timsina P, Patel D, Sawant A, Vaid A, Raut G, *et al*. Evaluating the accuracy of a state-of-the-art large language model for prediction of admissions from the emergency room. J Am Med Inform Assoc 2024;31:1921-8.
106. Bejan C, Reed A, Mikula M, Zhang S, Xu Y, Fabbri D, *et al*. Large language models improve the identification of emergency department visits for symptomatic kidney stones. Sci Rep 2025;15:3503. [doi: 10.1101/2024.08.12.24311870].
107. Tortum F, Kasali K. Exploring the potential of artificial intelligence models for triage in the emergency department. Postgrad Med 2024;136:841-6.
108. Open AI, Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, *et al*. Gpt-4 technical report,2024. Website. Available at:https://arxiv.org/abs/2303.08774 [Last accessed on 2025 Jan 28].
109. Thirunavukarasu AJ, Ting DS, Elangovan K, Gutierrez L, Tan TF, Ting DS. Large language models in medicine. Nat Med 2023;29:1930-40.
110. Iscoe M, Socrates V, Gilson A, Chi L, Li H, Huang T, *et al.* Identifying signs and symptoms of urinary tract infection from emergency department clinical notes using large language models. Acad Emerg Med 2024;31:599-610.
111. Jang JH, Lee SW, Kim DY, Shin SH, Lee SC, Kim DH, *et al.* Use of artificial intelligence-powered ECG to differentiate between cardiac and pulmonary pathologies in patients with acute dyspnoea in the emergency department. Open Heart 2024;11:e002924.

112. Lee SH, Hong WP, Kim J, Cho Y, Lee E. Smartphone AI versus medical experts: A comparative study in prehospital STEMI diagnosis. Yonsei Med J 2024;65:174-80.
113. Park MJ, Choi YJ, Shim M, Cho Y, Park J, Choi J, *et al.* Performance of ECG-derived digital biomarker for screening coronary occlusion in resuscitated out-of-hospital cardiac arrest patients: A comparative study between artificial intelligence and a group of experts. J Clin Med 2024;13:1354.
114. Han J, Ahn K, Cha K, Kim SJ, Jung WJ, Roh YII, *et al.* Prediction of blood pressure using chest compression waveform during cardiopulmonary resuscitation. Resuscitation 2024;202:110331.
115. Kim T, Suh GJ, Kim KS, Kim H, Park H Kwon WY, *et a*l. Development of artificial intelligence driven biosignal sensitive cardiopulmonary resuscitation robot. Resuscitation. 2024;202:110354.
116. Choi DH, Lee H, Joo H, Kong HJ, Lee SB, Kim S, *et al.* Development of prediction model for intensive care unit admission based on heart rate variability: A case-control matched analysis. Diagnostics (Basel) 2024;14:816.
117. Ou Z, Wang H, Zhang B, Liang H, Hu B, Ren L, *et al.* Early identification of stroke through deep learning with multi-modal human speech and movement data. Neural Regen Res 2025;20:234-41.
118. Zhang G, Xie Q, Wang C, Xu J, Liu G, Su C. Intelligent alert system for predicting invasive mechanical ventilation needs via noninvasive parameters: Employing an integrated machine learning method with integration of multicenter databases. Med Biol Eng Comput 2024;62:3445-58.
119. Liu LR, Huang MY, Huang ST, Kung LC, Lee CH, Yao WT, *et al.* An arrhythmia classification approach via deep learning using single-lead ECG without QRS wave detection. Heliyon 2024;10:e27200.
120. Sen A, Navarro L, Avril S, Aguirre M. A data-driven computational methodology towards a pre-hospital acute ischaemic stroke screening tool using haemodynamics waveforms. Comput Methods Programs Biomed 2024;244:107982.
121. Woehrle T, Pfeiffer F, Mandl MM, Sobtzick W, Heitzer J, Krstova A, *et al.* Point-of-care breath sample analysis by semiconductor-based E-nose technology discriminates non-infected subjects from SARS-CoV-2 pneumonia patients: A multi-analyst experiment. MedComm (2020) 2024;5:e726.
122. Herman R, Meyers HP, Smith SW, Bertolone DT, Leone A, Bermpeis K, *et al.* International evaluation of an artificial intelligence-powered electrocardiogram model detecting acute coronary occlusion myocardial infarction. Eur Heart J Digit Health 2024;5:123-33.
123. Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics. Int J Inf Manage 2015;35:137-44.
124. Davoudi A, Chae S, Evans L, Sridharan S, Song J, Bowles KH, *et al.* Fairness gaps in machine learning models for hospitalization and emergency department visit risk prediction in home healthcare patients with heart failure. Int J Med Inform 2024;191:105534.
125. Haraldsson T, Marzano L, Krishna H, Lethval S, Falk N, Bodeby P, *et al.* Exploring hospital overcrowding with an explainable time-to-event machine learning approach. Stud Health Technol Inform 2024;316:678-82.
126. Peláez-Rodríguez C, Torres-López R, Pérez-Aracil J, López-Laguna N, Sánchez-Rodríguez S, Salcedo-Sanz S. An explainable machine learning approach for hospital emergency department visits forecasting using continuous training and multi-model regression. Comput Methods Programs Biomed 2024;245:108033.
127. Aziz W, Nicalaou A, Stylianides C, Panayides A, Kakas A, Kyriacou E, *et al.* Emergency department length of stay classification based on ensemble methods and rule extraction. Stud Health Technol Inform 2024;316:1812-6.
128. Porto BM, Fogliatto FS. Enhanced forecasting of emergency department patient arrivals using feature engineering approach and machine learning. BMC Med Inform Decis Mak 2024;24:377.
129. Kauppi W, Imberg H, Herlitz J, Molin O, Axelsson C, Magnusson C. Advancing a machine learning-based decision support tool for pre-hospital assessment of dyspnoea by emergency medical service clinicians: A retrospective observational study. BMC Emerg Med 2025;25:2.
130. Jiang Y, Zhao Q, Guan J, Wang Y, Chen J, Li Y. Analyzing prehospital delays in recurrent acute ischemic stroke: Insights from interpretable machine learning. Patient Educ Couns 2024;123:108228.
131. Zhu L, Li Y, Zhao Q, Li C, Wu Z, Jiang Y. Assessing the severity of ODT and factors determinants of late arrival in young patients with acute ischemic stroke. Risk Manag Healthc Policy 2024;17:2635-45.
132. López-Izquierdo R, Del Pozo Vegas C, Sanz-García A, Mayo Íscar A, Castro Villamor MA, Silva Alvarado E, *et al.* Clinical phenotypes and short-term outcomes based on prehospital point-of-care testing and on-scene vital signs. NPJ Digit Med 2024;7:197.
133. Kajino K, Daya MR, Onoe A, Nakamura F, Nakajima M, Sakuramoto K, *et al.* Development and validation of a prehospital termination of resuscitation (TOR) rule for out-of hospital cardiac arrest (OHCA) cases using general purpose artificial intelligence (AI). Resuscitation 2024;197:110165.
134. Farhat H, Makhlouf A, Gangaram P, El Aifa K, Howland I, Babay Ep Rekik F, *et al.* Predictive modelling of transport decisions and resources optimisation in pre-hospital setting using machine learning techniques. PLoS One 2024;19:e0301472.
135. Smida T, Price BS, Mizener A, Crowe RP, Bardes JM. Prehospital post-resuscitation vital sign phenotypes are associated with outcomes following out-of-hospital cardiac arrest. Prehosp Emerg Care 2025;29:138-45.
136. Xu YY, Weng SJ, Huang PW, Wang LM, Chen CH, Tsai YT, *et al.* The emergency medical service dispatch recommendation system using simulation based on bed availability. BMC Health Serv Res 2024;24:1513.
137. Peng C, Peng L, Yang F, Yu H, Chen Q, Guo Y, *et al.* The prediction of the survival in patients with severe trauma during prehospital care: Analyses based on NTDB database. Eur J Trauma Emerg Surg 2024;50:1599-609.
138. Nasser L, McLeod SL, Hall JN. Evaluating the reliability of a remote acuity prediction tool in a Canadian academic emergency department. Ann Emerg Med 2024;83:373-9.
139. van Veen M, Steyerberg EW, Ruige M, van Meurs AH, Roukema J, van der Lei J, *et al.* Manchester triage system in paediatric emergency care: Prospective observational study. BMJ 2008;337:a1501.
140. Gilboy N, Tanabe T, Travers DA, Rosenau AM. Emergency Severity Index (ESI): A triage tool for emergency department care. Ver. 4. In: Implementation Handbook. Rockville,MD: AHRQ; 2011.
141. Christian MD. Triage. Crit Care Clin 2019;35:575-89.
142. El Ariss AB, Kijpaisalratana N, Ahmed S, Yuan J, Coleska A, Marshall A, *et al.* Development and validation of a machine learning framework for improved resource allocation in the emergency department. Am J Emerg Med 2024;84:141-8.
143. Chen Q, Qin Y, Jin Z, Zhao X, He J, Wu C, *et al.* Enhancing performance of the national field triage guidelines using machine learning: Development of a prehospital triage model to predict severe trauma. J Med Internet Res 2024;26:e58740.
144. Nsubuga M, Kintu TM, Please H, Stewart K, Navarro SM. Enhancing trauma triage in low-resource settings using machine learning: A performance comparison with the Kampala Trauma Score. BMC Emerg Med 2025;25:14.
145. Viana J, Souza J, Rocha R, Santos A, Freitas A. Identification of avoidable patients at triage in a paediatric emergency department: A decision support system using predictive analytics. BMC Emerg

Med 2024;24:149.
146. Chen AT, Kuzma RS, Friedman AB. Identifying low acuity emergency department visits with a machine learning approach: The low acuity visit algorithms (LAVA). Health Serv Res 2024;59:e14305.
147. Look CS, Teixayavong S, Djärv T, Ho AF, Tan KB, Ong ME. Improved interpretable machine learning emergency department triage tool addressing class imbalance. Digit Health 2024;10:20552076241240910.
148. Yu JY, Kim D, Yoon S, Kim T, Heo S, Chang H, *et al.* Inter hospital external validation of interpretable machine learning based triage score for the emergency department using common data model. Sci Rep 2024;14:6666.
149. Defilippo A, Veltri P, Lió P, Guzzi PH. Leveraging graph neural networks for supporting automatic triage of patients. Sci Rep 2024;14:12548.
150. Wyatt S, Lunde Markussen D, Haizoune M, Vestbø AS, Sima YT, Sandboe MI, *et al.* Leveraging machine learning to identify subgroups of misclassified patients in the emergency department: Multicenter proof-of-concept study. J Med Internet Res 2024;26:e56382.
151. Nanini S, Abid M, Mamouni Y, Wiedemann A, Jouvet P, Bourassa S. Machine and deep learning models for hypoxemia severity triage in CBRNE emergencies. Diagnostics (Basel) 2024;14:2763.
152. Gan T, Liu X, Liu R, Huang J, Liu D, Tu W, *et al.* Machine learning based prediction models for analyzing risk factors in patients with acute abdominal pain: A retrospective study. Front Med (Lausanne) 2024;11:1354925.
153. Grant L, Diagne M, Aroutiunian R, Hopkins D, Bai T, Kondrup F, *et al.* Machine learning outperforms the Canadian Triage and Acuity Scale (CTAS) in predicting need for early critical care. CJEM 2025;27:43-52.
154. Boresta M, Giovannelli T, Roma M. Managing low-acuity patients in an emergency department through simulation-based multiobjective optimization using a neural network metamodel. Health Care Manag Sci 2024;27:415-35.
155. Xu Y, Malik N, Chernbumroong S, Vassallo J, Keene D, Foster M, *et al.* Triage in major incidents: Development and external validation of novel machine learning-derived primary and secondary triage tools. Emerg Med J 2024;41:176-83.
156. Yoon P, Steiner I, Reinhardt G. Analysis of factors influencing length of stay in the emergency department. CJEM 2003;5:155-61.
157. Canellas MM, Kotkowski KA, Pachamanova DA, Perakis G, Reznek MA, Skali Lami O, *et al*. A granular view of emergency department length of stay: Improving predictive power and extracting real-time, actionable insights. Ann Emerg Med 2024;84:386-98.
158. Saggu S, Daneshvar H, Samavi R, Pires P, Sassi RB, Doyle TE, *et al.* Prediction of emergency department revisits among child and youth mental health outpatients using deep learning techniques. BMC Med Inform Decis Mak 2024;24:42.
159. Lehan E, Briand P, O'Brien E, Hafeez AA, Mulder DJ. Synergistic patient factors are driving recent increased pediatric urgent care demand. PLOS Digit Health 2024;3:e0000572.
160. Heyman ET, Ashfaq A, Ekelund U, Ohlsson M, Björk J, Khoshnood AM, *et al*. A novel interpretable deep learning model for diagnosis in emergency department dyspnoea patients based on complete data from an entire health care system. PLoS One 2024;19:e0311081.
161. Flores E, Martínez-Racaj L, Blasco Á, Diaz E, Esteban P, López-Garrigós M, *et al*. A step forward in the diagnosis of urinary tract infections: From machine learning to clinical practice. Comput Struct Biotechnol J 2024;24:533-41.
162. Roshanaei G, Salimi R, Mahjub H, Faradmal J, Yamini A, Tarokhian A. Accurate diagnosis of acute appendicitis in the emergency department: An artificial intelligence-based approach. Intern Emerg Med 2024;19:2347-57.
163. Saboorifar H, Rahimi M, Babaahmadi P, Farokhzadeh A, Behjat M, Tarokhian A. Acute cholecystitis diagnosis in the emergency department: An artificial intelligence-based approach. Langenbecks Arch Surg 2024;409:288.
164. Chang CH, Nguyen PA, Huang CC, Liu CF, Melisa S, Chen CJ, *et al.* Acute myocardial infarction risk prediction in emergency chest pain patients: An external validation study. Int J Med Inform 2025;193:105683.
165. Yilmaz R, Yagin FH, Colak C, Toprak K, Abdel Samee N, Mahmoud NF, *et al.* Analysis of hematological indicators via explainable artificial intelligence in the diagnosis of acute heart failure: A retrospective study. Front Med (Lausanne) 2024;11:1285067.
166. Holmstrom L, Bednarski B, Chugh H, Aziz H, Pham HN, Sargsyan A, *et al.* Artificial intelligence model predicts sudden cardiac arrest manifesting with pulseless electric activity versus ventricular fibrillation. Circ Arrhythm Electrophysiol 2024;17:e012338.
167. Aygun U, Yagin FH, Yagin B, Yasar S, Colak C, Ozkan AS, *et al.* Assessment of sepsis risk at admission to the emergency department: Clinical interpretable prediction model. Diagnostics (Basel) 2024;14:457.
168. Ben-Haim G, Yosef M, Rowand E, Ben-Yosef J, Berman A, Sina S, *et al.* Combination of machine learning algorithms with natural language processing may increase the probability of bacteremia detection in the emergency department: A retrospective, big-data analysis of 94,482 patients. Digit Health 2024;10:20552076241277673.
169. Toprak B, Solleder H, Di Carluccio E, Greenslade JH, Parsonage WA, Schulz K, *et al.* Diagnostic accuracy of a machine learning algorithm using point-of-care high-sensitivity cardiac troponin I for rapid rule-out of myocardial infarction: A retrospective study. Lancet Digit Health 2024;6:e729-38.
170. Song YF, Huang HN, Ma JJ, Xing R, Song YQ, Li L, *et al*. Early prediction of sepsis in emergency department patients using various methods and scoring systems. Nurs Crit Care 2024. [doi: 10.1111/nicc. 13201].
171. Li H, Liu Z, Sun W, Li T, Dong X. Interpretable machine learning for the prediction of death risk in patients with acute diquat poisoning. Sci Rep 2024;14:16101.
172. Brasen CL, Andersen ES, Madsen JB, Hastrup J, Christensen H, Andersen DP, *et al.* Machine learning in diagnostic support in medical emergency departments. Sci Rep 2024;14:17889.
173. Rahadian RE, Okada Y, Shahidah N, Hong D, Ng YY, Chia MY, *et al.* Machine learning prediction of refractory ventricular fibrillation in out-of-hospital cardiac arrest using features available to EMS. Resusc Plus 2024;18:100606.
174. Schipper A, Belgers P, O'Connor R, Jie KE, Dooijes R, Bosma JS, *et al.* Machine-learning based prediction of appendicitis for patients presenting with acute abdominal pain at the emergency department. World J Emerg Surg 2024;19:40.
175. Kijpaisalratana N, Saoraya J, Nhuboonkaew P, Vongkulbhisan K, Musikatavorn K. Real-time machine learning-assisted sepsis alert enhances the timeliness of antibiotic administration and diagnostic accuracy in emergency department patients with sepsis: A cluster-randomized trial. Intern Emerg Med 2024;19:1415-24.
176. Chiu CP, Chou HH, Lin PC, Lee CC, Hsieh SY. Using machine learning to predict bacteremia in urgent care patients on the basis of triage data and laboratory results. Am J Emerg Med 2024;85:80-5.
177. Goodacre S. Using clinical risk models to predict outcomes: What are we predicting and why? Emerg Med J 2023;40:728-30.
178. Rahmatinejad Z, Dehghani T, Hoseini B, Rahmatinejad F, Lotfata A, Reihani H, *et al.* A comparative study of explainable ensemble learning and logistic regression for predicting in-hospital

mortality in the emergency department. Sci Rep 2024;14:3406.
179. Lee S, Lee KS, Park SH, Lee SW, Kim SJ. A machine learning-based decision support system for the prognostication of neurological outcomes in successfully resuscitated out-of-hospital cardiac arrest patients. J Clin Med 2024;13:7600.
180. Richards JE, Yang S, Kozar RA, Scalea TM, Hu P. A machine learning-based Coagulation Risk Index predicts acute traumatic coagulopathy in bleeding trauma patients. J Trauma Acute Care Surg. 2025;98:614-20. doi:10.1097/TA.0000000000004463.
181. Ding H, Feng X, Yang Q, Yang Y, Zhu S, Ji X, *et al*. A risk prediction model for efficient intubation in the emergency department: A 4-year single-center retrospective analysis. J Am Coll Emerg Physicians Open 2024;5:e13190.
182. Ortiz-Barrios M, Petrillo A, Arias-Fonseca S, McClean S, de Felice F, Nugent C, *et al.* An AI-based multiphase framework for improving the mechanical ventilation availability in emergency departments during respiratory disease seasons: A case study. Int J Emerg Med 2024;17:45.
183. Li L, Han X, Zhang Z, Han T, Wu P, Xu Y, *et al.* Construction of prognosis prediction model and visualization system of acute paraquat poisoning based on improved machine learning model. Digit Health 2024;10:20552076241287891.
184. Deng YX, Wang JY, Ko CH, Huang CH, Tsai CL, Fu LC. Deep learning-based Emergency Department In-hospital Cardiac Arrest Score (Deep EDICAS) for early prediction of cardiac arrest and cardiopulmonary resuscitation in the emergency department. BioData Min 2024;17:52.
185. Shashikumar SP, Le JP, Yung N, Ford J, Singh K, Malhotra A, *et al.* Development and validation of a deep learning model for prediction of adult physiological deterioration. Crit Care Explor 2024;6:e1151.
186. Jawad BN, Shaker SM, Altintas I, Eugen-Olsen J, Nehlin JO, Andersen O, *et al.* Development and validation of prognostic machine learning models for short- and long-term mortality among acutely admitted patients based on blood tests. Sci Rep 2024;14:5942.
187. Choi HJ, Lee C, Chun J, Seol R, Lee YM, Son YJ. Development of a predictive model for survival over time in patients with out-of-hospital cardiac arrest using ensemble-based machine learning. Comput Inform Nurs 2024;42:388-95.
188. Park SW, Yeo NY, Kang S, Ha T, Kim TH, Lee D, *et al.* Early prediction of mortality for septic patients visiting emergency room based on explainable machine learning: A real-world multicenter study. J Korean Med Sci 2024;39:e53.
189. Siakopoulou S, Billis A, Logaras E, Stelmach V, Zouka M, Fyntanidou V, *et al.* Experimentation of AI models towards the prediction of medium-risk emergency department cases disposition outcome. Stud Health Technol Inform 2024;316:914-8.
190. Wang CH, Tay J, Wu CY, Wu MC, Su PI, Fang YD, *et al.* External validation and comparison of statistical and machine learning-based models in predicting outcomes following out-of-hospital cardiac arrest: A multicenter retrospective analysis. J Am Heart Assoc 2024;13:e037088.
191. Lee S, Kim DW, Oh NE, Lee H, Park S, Yon DK, *et al.* External validation of an artificial intelligence model using clinical variables, including ICD-10 codes, for predicting in-hospital mortality among trauma patients: A multicenter retrospective cohort study. Sci Rep 2025;15:1100.
192. Lin YT, Deng YX, Tsai CL, Huang CH, Fu LC. Interpretable deep learning system for identifying critical patients through the prediction of triage level, hospitalization, and length of stay: Prospective study. JMIR Med Inform 2024;12:e48862.
193. Nikouline A, Feng J, Rudzicz F, Nathens A, Nolan B. Machine learning in the prediction of massive transfusion in trauma: A retrospective analysis as a proof-of-concept. Eur J Trauma Emerg Surg 2024;50:1073-81.
194. Tsai SC, Lin CH, Chu CJ, Lo HY, Ng CJ, Hsu CC, *et al.* Machine learning models for predicting mortality in patients with cirrhosis and acute upper gastrointestinal bleeding at an emergency department: A retrospective cohort study. Diagnostics (Basel) 2024;14:1919.
195. Hinson JS, Zhao X, Klein E, Badaki-Makun O, Rothman R, Copenhaver M, *et al.* Multisite development and validation of machine learning models to predict severe outcomes and guide decision-making for emergency department patients with influenza. J Am Coll Emerg Physicians Open 2024;5:e13117.
196. Mehrpour O, Saeedi F, Vohra V, Hoyte C. Outcome prediction of methadone poisoning in the United States: Implications of machine learning in the National Poison Data System (NPDS). Drug Chem Toxicol 2024;47:556-63.
197. Chen JY, Hsieh CC, Lee JT, Lin CH, Kao CY. Patient stratification based on the risk of severe illness in emergency departments through collaborative machine learning models. Am J Emerg Med 2024;82:142-52.
198. Gauss T, Moyer JD, Colas C, Pichon M, Delhaye N, Werner M, *et al.* Pilot deployment of a machine-learning enhanced prediction of need for hemorrhage resuscitation after trauma – The ShockMatrix pilot study. BMC Med Inform Decis Mak 2024;24:315.
199. Simon GE, Johnson E, Shortreed SM, Ziebell RA, Rossom RC, Ahmedani BK, *et al.* Predicting suicide death after emergency department visits with mental health or self-harm diagnoses. Gen Hosp Psychiatry 2024;87:13-9.
200. Jawad BN, Altintas I, Eugen-Olsen J, Niazi S, Mansouri A, Rasmussen LJ, *et al.* Prospective and external validation of machine learning models for short-and long-term mortality in acutely admitted patients using blood tests. J Clin Med 2024;13:6437.
201. Chang CH, Chen CJ, Ma YS, Shen YT, Sung MI, Hsu CC, *et al.* Real-time artificial intelligence predicts adverse outcomes in acute pancreatitis in the emergency department: Comparison with clinical decision rule. Acad Emerg Med 2024;31:149-55.
202. Soundararajan K, Adams D, Nathanson B, Mader TJ, Godwin RC, Melvin RL, *et al*. Use of machine learning models to predict neurologically intact survival for advanced age adults following out of hospital cardiac arrest. Acad Emerg Med Official J Soc Acad Emerg Med 2025;32:169-71. [doi:10.1111/acem.15018].
203. Yang S, Hu P, Kalpakis K, Burdette B, Chen H, Parikh G, *et al.* Utilizing ultra-early continuous physiologic data to develop automated measures of clinical severity in a traumatic brain injury population. Sci Rep 2024;14:7618.
204. Shung DL, Chan CE, You K, Nakamura S, Saarinen T, Zheng NS, *et al.* Validation of an electronic health record-based machine learning model compared with clinical risk scores for gastrointestinal bleeding. Gastroenterology 2024;167:1198-212.
205. Wei L, Lv H, Yue C, Yao Y, Gao N, Chai Q, *et al*. A machine learning algorithm-based predictive model for pressure injury risk in emergency patients: A prospective cohort study. Int Emerg Nurs 2024;74:101419.
206. Seger DL, Amato MG, Frits M, Iannaccone C, Mugal A, Chang F, *et al.* A machine learning technology for addressing medication-related risk in older, multimorbid patients. Am J Manag Care 2024;30:e233-9.
207. Ahmed A, Aram KY, Tutun S, Delen D. A study of "left against medical advice" emergency department patients: An optimized explainable artificial intelligence framework. Health Care Manag Sci 2024;27:485-502.
208. Fujiwara G, Okada Y, Suehiro E, Yatsushige H, Hirota S, Hasegawa S, *et al.* Development of machine-learning model to predict anticoagulant use and type in geriatric traumatic brain injury using coagulation parameters. Neurol Med Chir (Tokyo) 2025;65:61-70.
209. Hsu CC, Chu CJ, Ng CJ, Lin CH, Lo HY, Chen SY. Machine learning models for predicting unscheduled return visits of

patients with abdominal pain at emergency department and validation during COVID-19 pandemic: A retrospective cohort study. Medicine (Baltimore) 2024;103:e37220.

210. Sung CW, Ho J, Fan CY, Chen CY, Chen CH, Lin SY, *et al.* Prediction of high-risk emergency department revisits from a machine-learning algorithm: A proof-of-concept study. BMJ Health Care Inform 2024;31:e100859.
211. Chan KS, Zary N. Applications and challenges of implementing artificial intelligence in medical education: Integrative review. JMIR Med Educ 2019;5:e13930.
212. Gordon M, Daniel M, Ajiboye A, Uraiby H, Xu NY, Bartlett R, *et al*. A scoping review of artificial intelligence in medical education: BEME guide no. 84. Med Teach 2024;46:446-70.
213. Sami A, Tanveer F, Sajwani K, Kiran N, Javed MA, Ozsahin DU, *et al*. Medical students' attitudes toward AI in education: Perception, effectiveness, and its credibility. BMC Med Educ 2025;25:82.
214. Aster A, Hütt C, Morton C, Flitton M, Laupichler MC, Raupach T. Development and evaluation of an emergency department serious game for undergraduate medical students. BMC Med Educ 2024;24:1061.
215. Duggan NM, Jin M, Duran Mendicuti MA, Hallisey S, Bernier D, Selame LA, *et al.* Gamified crowdsourcing as a novel approach to lung ultrasound data set labeling: Prospective analysis. J Med Internet Res 2024;26:e51397.
216. Spadafore M, Yilmaz Y, Rally V, Chan TM, Russell M, Thoma B, *et al.* Using natural language processing to evaluate the quality of supervisor narrative comments in competency based medical education. Acad Med 2024;99:534-40.
217. Shamim M, Zaidi S, Rehman A. The revival of essay-type questions in medical education: Harnessing artificial intelligence and machine learning. J Coll Physicians Surg Pak 2024;34:595-9.
218. Iftikhar H, Anjum S, Bhutta ZA, Najam M, Bashir K. Performance of ChatGPT in emergency medicine residency exams in Qatar: A comparative analysis with resident physicians. Qatar Med J 2024;2024:61.
219. Misra SM, Suresh S. Artificial intelligence and objective structured clinical examinations: Using ChatGPT to revolutionize clinical skills assessment in medical education. J Med Educ Curric Dev 2024;11:23821205241263475.
220. Schnapp B, Sehdev M, Schrepel C, Bord S, Pelletier-Bui A, Alvarez A, *et al.* ChatG-PD? Comparing large language model artificial intelligence and faculty rankings of the competitiveness of standardized letters of evaluation. AEM Educ Train 2024;8:e11052.
221. Wang L, Mao Y, Wang L, Sun Y, Song J, Zhang Y. Suitability of GPT-4o as an evaluator of cardiopulmonary resuscitation skills examinations. Resuscitation 2024;204:110404.
222. Huang LW, Chan YW, Tsan YT, Zhang QX, Chan WC, Yang HH. Implementation of a smart teaching and assessment system for high-quality cardiopulmonary resuscitation. Diagnostics (Basel) 2024;14:995.
223. Eskandarani R, Almuhainy A, Alzahrani A. Creating a master training rotation schedule for emergency medicine residents and challenges in using artificial intelligence. Int J Emerg Med 2024;17:84.
224. Johnson D, Chopra S, Bilgic E. Exploring the use of natural language processing to understand emotions of trainees and faculty regarding entrustable professional activity assessments. J Grad Med Educ 2024;16:323-7.
225. Karnan N, Francis J, Vijayvargiya I, Rubino Tan C. Analyzing the effectiveness of AI-generated patient education materials: A comparative study of ChatGPT and Google Gemini. Cureus 2024;16:e74398.
226. Harden RM, Grant J, Buckley G, Hart IR. Best evidence medical education. Adv Health Sci Educ Theory Pract 2000;5:71-90.
227. Webb JJ. Proof of concept: Using ChatGPT to teach emergency physicians how to break bad news. Cureus 2023;15:e38755.
228. Yilmaz Y, Jurado Nunez A, Ariaeinejad A, Lee M, Sherbino J, Chan TM. Harnessing natural language processing to support decisions around workplace-based assessment: Machine learning study of competency-based medical education. JMIR Med Educ 2022;8:e30537.
229. Choi J, Lee Y, Kang GH, Jang YS, Kim W, Choi HY, *et al.* Educational suitability of new channel-type video-laryngoscope with AI-based glottis guidance system for novices wearing personal-protective-equipment. Medicine (Baltimore) 2022;101:e28890.
230. Zhao S, Xiao X, Zhang X, Yan Meng WL, Soghier L, Hahn JK. Automated assessment system for neonatal endotracheal intubation using dilated convolutional neural network. Annu Int Conf IEEE Eng Med Biol Soc 2020;2020:5455-8.

Supplementary Table 1: https://turkjemergmed.com/pages/2025-2-issue-supplementary-files

Supplementary Table 2: https://turkjemergmed.com/pages/2025-2-issue-supplementary-files

Supplementary Table 3: https://turkjemergmed.com/pages/2025-2-issue-supplementary-files

Supplementary Table 4: https://turkjemergmed.com/pages/2025-2-issue-supplementary-files

Supplementary Table 5: https://turkjemergmed.com/pages/2025-2-issue-supplementary-files