Deep learning in Emergency Medicine: Recent contributions and methodological challenges

Francesco Faita

Institute of Clinical Physiology, National Research Council, Pisa, Italy

Abstract

In the last few years, artificial intelligence (AI) technology has grown dramatically impacting several fields of human knowledge and medicine in particular. Among other approaches, deep learning, which is a subset of AI based on specific computational models, such as deep convolutional neural networks and recurrent neural networks, has shown exceptional performance in images and signals processing. Accordingly, emergency medicine will benefit from the adoption of this technology. However, a particular attention should be devoted to the review of these papers in order to exclude overoptimistic results from clinically transferable ones. We presented a group of studies recently published on PubMed and selected by keywords 'deep learning emergency medicine' and 'artificial intelligence emergency medicine' with the aim of highlighting their methodological strengths and weaknesses, as well as their clinical usefulness.

Introduction

Artificial intelligence (AI) has recently emerged as a disruptive technology in biomedicine especially because the availability of large repository of clinical data and the progressive growth of computing capacity.[1] Recently, the number of published papers about this topic has risen exponentially, while the performance of the proposed systems is getting closer (and in many cases exceeding) than those of experts in many medical tasks. AI in medicine is a wide field with applications ranging from natural language processing to deep learning (DL). Accordingly, we restricted our analysis this latter topic because DL solutions using convolutional neural networks has shown exceptional performance in images and signals processing.[2] Accordingly, some AI systems are almost ready to be used in everyday clinical life, especially in the fields of ophthalmology[3] and dermatology,[4] while there are still few applications for emergency medicine departments (ED).[5] However, emergency medicine is an eligible application area for artificial intelligence systems due to its main characteristics of unpredictability of patient influx, criticality of the patient's conditions, high risk of human error and exhaust human capabilities. In particular, three classes of applications that could be implemented using deep learning technology, can be mainly recognized according to the published literature: i) automatic triage systems; ii) algorithms for the prediction of clinical events and for the generation of alerts based on signals monitoring; iii) clinical decision support systems for diagnostic imaging. On the other hand, artificial intelligence is still at an early stage of development and its adoption in clinical practice is not trivial, while patient safety requires a rigorous methodological approach to avoid rhetoric and overly optimistic claims.[6] Therefore, an in-depth critical review and understanding of the *Materials and Methods* section of the scientific papers published in the field of *artificial intelligence and medicine* is crucial to promote a quick and safe adoption of these systems into the clinical practice.

The aim of this review is a brief and limited analysis of papers published in the last two years in the field of emergency medicine based on deep learning technology. In particular, all publications have been classified according to three main categories and analyzed both in terms of the potential impact that they can have in clinical practice, and of the appropriateness of methodological aspects.

Correspondence: Francesco Faita, Institute of Clinical Physiology, National Research Council, via G. Moruzzi 1, 56124 Pisa, Italy.
Tel.: +39.050.3158125 - Fax: +39.050.3152166.
E-mail: f.faita@ifc.cnr.it

Materials and Methods

Deep learning

Deep learning is a particular sub-class of artificial intelligence algorithms that is very popular due to the astonishing results obtained by recently published studies. Fundamentally, it is based on the implementation of neural networks (NN) characterized by a high number of hidden layers with hidden nodes that learn to represent input data according many abstraction methods. The main difference of the DL with respect to the simpler use of NNs consists in the increased number of layers and nodes as well as in a general increase of the network size. In general, this specific feature of the DL technique allows a more accurate representation of relations between complex data.[7] The convolutional neural network (CNN) is the most typical and common example of a DL network used for classification tasks. It consists of a feed-forward

NN designed to mimic the neural processes of the brain. In order to do this, a CNN first subsamples the input data into feature maps through convolution operations performed by the hidden node and layers. Subsequently, all the features are recombined to describe the data at abstraction levels beyond the raw value of the individual sample.[8] When the network concludes that the combination of features describes a specific pattern, this information is propagated to all previous (hidden) nodes (back propagation). Therefore, nodes will associate the maps of features they have created with this specific pattern in a permanent way. These representations are used by the network itself to classify the different patterns that can be identified through the output layers of the NN.

Literature search

PubMed website was used to search for articles with the following keywords: 'deep learning emergency medicine' and 'artificial intelligence emergency medicine'. The output of the research was ordered based on the publication date. Furthermore, the analysis of the results was limited from 1 January 2018 to 30 June 2019.

Abstracts of all articles have been analyzed to appropriately select only papers that report technical description and validation of deep learning systems concerning emergency medicine.

Selected papers have been divided into three main categories: i) triage and risk stratification; ii) clinical events prediction and clinical data monitoring; iii) medical images analysis.

Methodological analysis

In order to assess the reliability and the quality of the papers, the design of the study, as well as the features of artificial intelligence algorithms were evaluated, as described in Table 1.

Most critical points of each study in relation to the abovementioned elements were highlighted in the following *Results* section.

Results

The PubMed research using the key words 'deep learning emergency medicine' provided 60 results, while 'artificial intelligence emergency medicine' highlighted 305 papers. After filtering by date (only papers published in 2018 and 2019) and by content (checking the coherence of the abstract with the keywords), 15 papers were selected and divided into the three main categories, as shown in Table 2. Scientific articles published by the same authors on the same topic, but in different journals, were selected only once.

Triage and risk stratification

Predicting clinical outcomes is a primary goal in ED triage. Accordingly, Raita *et al*.[9] have developed a system that performs this task using routinely available data only. The sample size is large (>135,000 subjects) and it was extracted from a public database that properly characterizes the general adult population of North America. The primary and secondary main outcomes of the study (in-hospital death and hospitalization, respectively) are of great interest for their impact on public health. Splitting of data is optimal with an appropriate balance for training (70%) and tests (30%) populations without overlap. The National Hospital and Ambulatory Medical Care Survey five-level Emergency Severity Index algorithm, which could be considered a high-quality gold standard, was used for reference of patient classification. The DL system was also compared with some other more traditional machine learning approaches; appropriately, the number of systems tested is relatively limited compared to the sample size in order to avoid overfitting problems. However, the system is affected by the black box effect, as the weight of input variables in terms of prediction power was not evaluated.

Subsequently, the same research team published a DL-based triage system trained on a population of about 50,000 children (<18 years).[10] This system has a similar design as the previous work, thus having the same methodological pros and cons, but it showed significant results in terms of clinical performance, being particularly effective in avoiding under-triaged critically ill subjects and over-triaged hospitalized ones.

Triaging patients even in the absence of qualified medical personnel starting from a minimum set of vital signs together with a simple coma scale is the goal of the neural network system proposed by Kim *et al*.[11] Despite a critically important primary outcome (death after ED discharge) and a large sample size (>460,000 subjects), the paper has methodological issues, that actually limit its clinical value. In fact, a 10-fold cross-validation technique, which involves the use of the same data, both for training and for algorithm testing, was adopted for testing algorithm's performances. Furthermore, due to a relatively low number of subjects with outcomes compared to the total

Table 1. List of elements used for quality and reliability evaluation of the selected studies.

Element	Description
Data collection	Retrospective or prospective study Consecutive or convenience sampling Sample size
Data labelling	Quality of the reference standard (gold standard or expert opinion) Model training and validation Availability of training/validation/test subsets Test data not used for training Quality of metric for performance evaluation
Visualization of results	Avoiding of black box effect
Overfitting	High accuracy on limited dataset Low generalization Too many models tested
Spectrum bias	Appropriateness of the dataset in representation of full range of possible clinical data
Data snooping bias	Use of too similar data in training and testing

population (less than 4%), test classes were derived with a stratified approach without randomization, thus implying a sample selection bias. On the other hand, authors presented an interesting sensitivity map, which is useful for understanding the effect of each input variables on the final prediction.

Overcoming limitations of the currently adopted triage scores is the main objective of the paper published by Kwon *et al.*[12] The design of the study is methodologically adequate, as authors recruited an external validation population, together with a recommended splitting of the main study population in training and test subgroups without overlapping. The study is also well designed in terms of outcomes (including hospital mortality and need of hospitalization), while the predictors used for the analysis (common to the main triage scores used worldwide) suggest the adoption of the system in real life situations. Unfortunately, the large sample size (>11 million subjects) is composed by Korean

Table 2. List of selected papers including identification three main categories.

Title	Authors	Journal	Category
Emergency department triage prediction of clinical outcomes using machine learning models	Raita Y, Goto T, Faridi MK, Brown DFM, Camargo Jr CA,, Hasegawa K.	Critical Care	Triage and risk stratification
Machine Learning-Based Prediction of Clinical Outcomes for Children During Emergency Department Triage	Goto T, Camargo Jr CA, Faridi MK, Freishtat RJ, Hasegawa K.	JAMA Network Open	Triage and risk stratification
A data-driven artificial intelligence model for remote triage in the prehospital environment	Kim D, You S, So S, Lee J, Yook S, Jang DP, Kim IY, Park E, Cho K, Cha WC, Shin DW, Cho BH, Park H.	PLoS One	Triage and risk stratification
Validation of deep-learning-based triage and acuity score using a large national dataset	Kwon JM, Lee Y, Lee Y, Lee S, Park H, Park J.	PLoS One	Triage and risk stratification
Predicting hospital admission at emergency department triage using machine learning	Hong WS, Haimovich AD, Taylor RA.	PLoS One	Triage and risk stratification
Mixed convolutional and long short-term memory network for the detection of lethal ventricular arrhythmia	Picon A, Irusta U, Alvarez-Gila A, Aramendi E, Alonso-Atienza F, Figuera C, Ayala U, Garrote E, Wik L, Kramer-Johansen J, Eftestol T.	PLoS One	Clinical events prediction and clinical data monitoring
Developing neural network models for early detection of cardiac arrest in emergency department	Jang DH, Kim J, Jo YH, Lee JH, Hwang JE, Park SM, Lee DK, Park I, Kim D, Chang H.	American Journal of Emergency Medicine	Clinical events prediction and clinical data monitoring
Deep-learning-based out-of-hospital cardiac arrest prognostic system to predict clinical outcomes	Kwon JM, Jeon KH, Kim HM, Kim MJ, Lim S, Kim KH, Song PS, Park J, Choi RK, Oh BH.	Resuscitation	Clinical events prediction and clinical data monitoring
An Algorithm Based on Deep Learning for Predicting In-Hospital Cardiac Arrest	Kwon JM, Lee Y, Lee Y, Lee S, Park J.	Journal of American Heart Association	Clinical events prediction and clinical data monitoring
A deep neural network learning algorithm outperforms a conventional algorithm for emergency department electrocardiogram interpretation	Smith SW, Walsh B, Grauer K, Wang K, Rapin J, Li J, Fennell W, Taboulet P.	Journal of Electrocardiology	Clinical events prediction and clinical data monitoring
Artificial intelligence to predict needs for urgent revascularization from 12-leads electrocardiography in emergency patients	Goto S, Kimura M, Katsumata Y, Goto S, Kamatani T, Ichiihara G, Ko S, Sasaki J, Fukuda K, Sano M.	PLoS One	Clinical events prediction and clinical data monitoring
Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs	Cheng CT, Ho TY, Lee TY, Chang CC, Chou CC, Chen CC, Chung IF, Liao CH.	European Radiology	Medical images analysis
Improving Sensitivity on Identification and Delineation of Intracranial Hemorrhage Lesion Using Cascaded Deep Learning Models	Cho J, Park KS, Karki M, Lee E, Ko S, Kim JK, Lee D, Choe J, Son J, Kim M, Lee S, Lee J, Yoon C, Park S.	Journal of Digital Imaging	Medical images analysis
Deep learning for predicting in-hospital mortality among heart disease patients based on echocardiography	Kwon JM, Kim KEK, Jeon KH, Park JP.	Echocardiography	Medical images analysis
Hybrid 3D/2D Convolutional Neural Network for Hemorrhage Evaluation on Head CT	Chang PD, Kuoy E, Grinband J, Weinberg BD, Thompson M, Homo R, Chen J, Abcede H, Shafie M, Sugrue L, Filippi CG, Su MY, Yu W, Hess C, Chow D.	American Journal of Neuroradiology	Medical images analysis

people, imposing an ethnic bias and excluding the generalization of the study to other countries.

Hong *et al*.[13] realized a retrospective study with the aim of predicting hospitalization at triage time by deep neural networks feed with a large set of input variables. The enrolled population is quite large (>190,000 subjects), heterogeneous (three enrollment sites with variable general characteristics) and well-integrated in the study design (training and test subgroup are not overlapping and correctly sized). However, an extremely high number of network's configurations were tested with risk of overfitting. Accordingly, published results are quite variable with the risk that high benchmark score would be simply obtained by chance.

Clinical events prediction and clinical data monitoring

Ventricular fibrillation is characterized by high occurrence and severity of outcomes. Accordingly, Picon *et al*.[14] developed a system for automatic electrocardiogram (ECG) analysis based on a convolutional deep network. In order to train the network, authors extracted ECG traces both from a public database and from implanted devices (defibrillators during discharge). Despite a good methodology, the proposed system added an only marginal gain in terms of accuracy respect to those offered by other commercial systems already available in the market. In particular, authors claimed a change in specificity and sensitivity from 99.3/98.4 to 99.7/98.9, respectively. As similar systems based on ECG traces analysis, the proposed approach suffers of data snooping bias, with too similar ECG segments (although belonging to different subjects) included in training and test subsets.

Cardiac arrest is also the topic of the paper published by Jang *et al*.,[15] However, their approach does not require ECG traces analysis, but it is based on the processing of simple and easily available Electronic Health Record data. The main objective of the deep neural network is to predict events while patients are still in ED. The study is mainly devoted to the comparison of several network structures, all exceeding the clinically available standard scores. However, performances of the new system are only marginally improved respect to those obtained by deep learning algorithms already present in literature.

With their paper, Kwon *et al*.[16] aimed to evaluate the prognosis of out-of-hospital cardiac arrests. Despite the high impact of the topic, the paper is characterized by a study design that limits its clinical interest. In particular, the population is characterized by a strong ethnic bias and the sample size is relatively small. Furthermore, authors ran an exceeding number of experiments, testing several network configurations with the risk of overfitting. Additionally, the gold standard used to validate the system is not recognized at international level.

Subsequently, the same research group addressed the issue of in-hospital cardiac arrest[17] with the risk of patient death as outcome. In this paper, the study cohort, although small in size, was extracted from two different datasets, with one of them used as independent validation group, thus increasing the reliability of the algorithm. Furthermore, only four simple vital signs are needed as input of the system, increasing its likelihood of clinical adoption. However, blackbox effect is present, as it was not evaluated the weight of individual input parameters on the system's prediction.

The full analysis of the ECG trace is a task usually reserved for experts. Smith *et al*.[18] tested a previously developed deep NN algorithm with the aim of labelling more than 70 different pathophysiological classes. Authors randomly selected ECG traces from a large database mimicking the everyday activity of an ED. Performances of the algorithm were compared with that of a commercial ECG classification system using cardiologist judgement as gold standard reference. The deep learning system overperformed the commercial one in terms of both sensitivity and specificity, thus promoting its forthcoming clinical use.

ECG analysis aimed to predict urgent revascularization at 48 h in ED is the goal of the study by Goto *et al*.[19] Accordingly, the authors collected data from 39,000 consecutive subjects enrolled at a single hospital. Due to the lack of events (need for revascularization), the number of subjects considered for the analysis was limited to around 800, in order to avoid unbalanced data, but weakening the *big data nature* of the study. This reduction is still preferable to the inclusion in the study of an excessive imbalance between the number of subjects for each class of prediction. Unfortunately, the design of the network does not allow to clearly highlight the features of the ECG trace that are useful for prediction (blackbox effect).

Medical images analysis

Biomedical imaging is an extraordinary opportunity for development of DL solutions aimed at segmenting structures in images and performing classification tasks. Accordingly, the deep neural network developed by Cheng *et al*.[20] has a dual purpose: classifying X-ray images with respect to the presence of hip fractures and detect their contours. The application has a good clinical applicability as the correct and early diagnosis positively impacts the healing process and the quality of life of patients. The study is well designed with radiological images reliably labeled according to the diagnosis recorded in the trauma registry and with the inclusion of independent test images for validation of performances. Another interesting aspect of the study is the comparison of network's performance with opinion of experts, some of which (ED physicians and surgeons) are over performed by the network, while others (radiologist and orthopedist) performed better than the automatic system. Finally, the generation of heat maps allows radiologists to understand which features of the image contribute most to the classification, thus avoiding the blackbox effect.

A cascade of multiple deep networks is the solution proposed by Cho *et al*.[21] in order to segment and classify 5 subtypes of intracranial hemorrhage from computed tomography (CT) images. The system would have a remarkable applicability in ED allowing an accurate and early identification of lesions. The number of patients (5700 patients, 3055 with bleeding for more than 135k images overall) enrolled for the validation is more than adequate in order to achieve the goal, although all patients referred to the same hospital. On the other hand, the gold standard adopted for classification is based on the personal opinion of 10 experts and on a complex consensus process for controversial cases. This is a common procedure for radiological examination, but it is sub optimal for DL systems that should be preferably trained and validated against automatic and quantitative measurements.

Unfortunately, the gold standard adopted for classification is not adequate being based on the personal opinion of 10 experts and on a complex consensus procedure for controversial cases.

Biomedical images are often supplied with a detailed textual description. For this reason, Kwon *et al*.[22] have developed a deep network for text mining echocardiographic medical reports aimed at predicting in-hospital mortality from cardiovascular disease. The algorithm was trained on data collected at the cardiovascular university hospital and tested on data collected at the community general hospital. Performances of the system were evaluated against well-validated and largely adopted mortality scores obtaining an overall gain of 15% in accuracy. However, the authors

themselves recognize ethnicity and selection bias as a limitation of the study. Furthermore, they propose the introduction of a further system based on the direct analysis of the ultrasound images (instead of their medical reports) as a next step to make further progress in terms of clinical applicability.

The design of a study is a key factor to promote its successful adoption in clinical practice. Accordingly, Chang *et al.*[23] present an interesting prospective study based on DL image processing. The clinical target is the segmentation and quantification of three types of intracerebral hemorrhage from CT images without contrast. Training and testing of the deep neural network were performed on a large sample of more than 10,000 subjects (for a total of 512,000 CT images) and the developed network was prospectively applied on a further sample of 860 consecutive subjects with a CT for suspected cerebral hemorrhage. Authors strengthen their work with saliency maps and class activation that allow to avoid the black box effect. Finally, a quantitative metric was used to compare network segmentation with the expert's manual approach.

Conclusions

In recent years, AI technology has grown dramatically impacting several fields of human knowledge and medicine in particular.

Emergency medicine will benefit from the adoption of this technology, but care must be taken when reviewing papers in this topic in order to exclude overoptimistic results from clinically transferable ones.

We presented a group of studies published on PubMed since January 2018 and selected by keywords 'deep learning emergency medicine' and 'artificial intelligence emergency medicine' with the aim of highlighting their methodological strengths and weaknesses, as well as their clinical usefulness. Furthermore, we split these research studies in three groups related to potential high impact topics in emergency medicine.

Studies aimed at proposing innovative triage systems are mainly retrospective, despite the wide and continuously growing availability of well-labeled data. On the other hand, most of the proposed systems have been evaluated in comparison to reliable and widely recognized worldwide reference scores. Use of widely-available simple input data allows to avoid black box effect by clearly defining the importance of each individual predictor with respect to the final decision of the algorithm. However, several studies tested the performance of (too) many models generating overfitting issues, with the negative consequence of selecting best models that perform (extremely) well only by chance. Finally, due to the great heterogeneity of the National Health Systems worldwide and the data collection procedures during triage, it may be necessary to adopt a standardized evaluation procedure of the DL networks proposed for use in EDs.

The DL systems for the prediction of clinical events and for the generation of alerts based on signal monitoring are in an advanced development stage and nearly ready for technology transfer, just requiring an industrialization phase leading to a rapid inclusion in clinical practice. This is mainly due to the wide availability of well-labeled reliable data that support independent validation of algorithms. Despite this fact, this class of systems also has features to improve; in particular, the design of prospective studies would be required, as well as the integration of tools providing information about the contribution and importance of input data on network's output. The complex task of analyzing biomedical images is another target of DL systems that can represent a valid solution for emergency medicine unmet needs. In particular, systems have been proposed for the segmentation and classification of pathologies using all available imaging technologies, from the common X-ray, to the more complex CT and magnetic resonance or the widely available ultrasound. The major strengths of these systems are both in terms of accuracy of measurements and diagnostic power, even in the presence of non-specialized personnel. Furthermore, these systems are becoming time and money saving. On the contrary, one of the main barriers to their adoption in the clinical practice is the lack of valid image repository, that are publicly available and supplied with a reliable gold standard for system's validation.

As a final remark, 15 papers have been reviewed and only 4 resulted to be characterized by strong methodologies that completely satisfied all quality requirements (Table 1). This is not surprising because the popularity and the novelty of the topic, with several research groups having a strong background in theoretical AI algorithms but a lower knowledge in modelling clinical problems. On the other hand, reviewers of medical journal are often not completely equipped with critical tool for the evaluation of systems powered by AI technology.

In conclusion, although with some difficulties to face, the correct question is not whether the AI will be a valid resource for Emergency Medicine but when.

References

1. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25:44-56.
2. Anwar SM, Majid M, Qayyum A, et al. Medical image analysis using convolutional neural networks: a review. J Med Syst 2018;42:22.
3. De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med 2018;24:1342-50.
4. Esteva A, Kuprel B, Novoa R, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542:115-8.
5. Shafaf N, Malek H. Applications of machine learning approaches in emergency medicine; a review article. Arch Acad Emerg Med 2019;7:e34.
6. England JR, Cheng PM. Artificial intelligence for medical image analysis: a guide for authors and reviewers. AJR Am J Roentgenol 2019;212:513-9.
7. Lee JG, Jun S, Cho YW, et al. Deep learning in medical imaging: general overview. Korean J Radiol 2017;18:570-84.
8. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436-44.
9. Raita Y, Goto T, Faridi MK, et al. Emergency department triage prediction of clinical outcomes using machine learning models. Crit Care 2019;23:64.
10. Goto T, Camargo Jr CA, Faridi MK, et al. Machine learning-based prediction of clinical outcomes for children during emergency department triage. JAMA Netw Open 2019;2:e186937.
11. Kim D, You S, So S, et al. A data-driven artificial intelligence model for remote triage in the prehospital environment. PLoS One 2018;13:e0206006.
12. Kwon JM, Lee Y, Lee Y, et al. Validation of deep-learning-based triage and acuity score using a large national dataset. PLoS One 2018;13:e0205836.

13. Hong WS, Haimovich AD, Taylor RA. Predicting hospital admission at emergency department triage using machine learning. PLoS One 2018;13:e0201016.
14. Picon A, Irusta U, Alvarez-Gila A, et al. Mixed convolutional and long short-term memory network for the detection of lethal ventricular arrhythmia. PLoS One 2019;14:e0216756.
15. Jang DH, Kim J, Jo YH, et al. Developing neural network models for early detection of cardiac arrest in emergency department. Am J Emerg Med 2020;38:43-9.
16. Kwon JM, Jeon KH, Kim HM, et al. Deep-learning-based out-of-hospital cardiac arrest prognostic system to predict clinical outcomes. Resuscitation 2019;139:84-91.
17. Kwon JM, Lee Y, Lee Y, et al. An algorithm based on deep learning for predicting in-hospital cardiac arrest. J Am Heart Assoc 2018;7:e008678.
18. Smith SW, Walsh B, Grauer K, et al. A deep neural network learning algorithm outperforms a conventional algorithm for emergency department electrocardiogram interpretation. J Electrocardiol 2019;52:88-95.
19. Goto S, Kimura M, Katsumata Y, et al. Artificial intelligence to predict needs for urgent revascularization from 12-leads electrocardiography in emergency patients. PLoS One 2019;14:e0210103.
20. Cheng CT, Ho TY, Lee TY, et al. Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs. Eur Radiol 2019;29:5469-77.
21. Cho J, Park KS, Karki M, et al. Improving sensitivity on identification and delineation of intracranial hemorrhage lesion using cascaded deep learning models. J Digit Imaging 2019;32:450-61.
22. Kwon JM, Kim KH, Jeon KH, Park J. Deep learning for predicting in-hospital mortality among heart disease patients based on echocardiography. Echocardiography 2019;36:213-8.
23. Chang PD, Kuoy E, Grinband J, et al. Hybrid 3D/2D convolutional neural network for hemorrhage evaluation on head CT. AJNR Am J Neuroradiol 2018;39:1609-16.