



**HAL**  
open science

# Deep learning in medical image analysis: A third eye for doctors

A. Fourcade, R.H. Khonsari

► **To cite this version:**

A. Fourcade, R.H. Khonsari. Deep learning in medical image analysis: A third eye for doctors. *Journal of Stomatology, Oral and Maxillofacial Surgery*, 2019, 120, pp.279 - 288. <10.1016/j.jormas.2019.06.002>. <hal-03488104>

**HAL Id: hal-03488104**

**<https://hal.science/hal-03488104v1>**

Submitted on 20 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

## **Deep learning in medical image analysis: a third eye for doctors**

**Fourcade** Arthur (1), **Khonsari** Roman Hossein (2)

1. Service de chirurgie plastique, maxillofaciale et stomatologie, Centre Hospitalier de Gonesse

2. Assistance Publique – Hôpitaux de Paris, Service de chirurgie maxillofaciale et plastique, Hôpital Necker – Enfants Malades, Centre de Référence Maladies Rares MAFACE, Filière Maladies Rares TeteCou, Université Paris Descartes, Université de Paris, Paris, France

### **Corresponding author**

Roman H. Khonsari, Service de chirurgie plastique, maxillofaciale et stomatologie, Centre Hospitalier de Gonesse, 2 boulevard du 19 mars 1962, 96500 Gonesse. Tel : +33134532075 ; Fax : +33134532758 ; email : arthur.fourcade@ch-gonesse.fr

### **Keywords**

Deep learning; artificial intelligence; neural network; image analysis; systematic review; computer vision

## **Abstract**

### **AIM & SCOPE**

Artificial intelligence (AI) in medicine is a fast-growing field. The rise of deep learning algorithms, such as convolutional neural networks (CNNs), offers fascinating perspectives for the automation of medical image analysis. In this systematic review article, we screened the current literature and investigated the following question: 'Can deep learning algorithms for image recognition improve visual diagnosis in medicine?'

### **MATERIALS & METHODS**

We provide a systematic review of the articles using CNNs for medical image analysis, published in the medical literature before May 2019. Articles were screened based on the following items: type of image analysis approach (detection or classification), algorithm architecture, dataset used, training phase, test, comparison method (with specialists or other), results (accuracy, sensibility and specificity) and conclusion.

### **RESULTS**

We identified 352 articles in the PubMed database and excluded 327 items for which performance was not assessed (review articles) or for which tasks other than detection or classification, such as segmentation, were assessed. The 25 papers included were published from 2013 to 2019 and were related to a vast array of medical specialties. Authors were mostly from North America and Asia. Large amounts of qualitative medical images were necessary to train the CNNs, often resulting from international collaboration. The most common CNNs such as AlexNet and GoogleNet, designed for the analysis of natural images, proved their applicability to medical images.

### **CONCLUSION**

CNNs are not replacement solutions for medical doctors, but will contribute to optimize routine tasks and thus have a potential positive impact on our practice. Specialties with a strong visual component such as radiology and pathology will be deeply transformed. Medical practitioners, including surgeons, have a key role to play in the development and implementation of such devices.

## **Introduction**

### **1. Artificial intelligence and image analysis**

Artificial intelligence (AI) in medicine is a fast-growing field, generating hopes and raising perplexing issues. AI can be defined as the ability for a computer to mimic the cognitive abilities of a human being. AI corresponds to a large array of techniques. Among them, machine learning is one of the most relevant approaches in the medical field. Three converging technical progresses have facilitated the medical applications of machine learning: (1) the global rise of ‘big data’, that is the constitution and the analysis of very large databases, (2) the spectacular increase of the computing power of processors and (3) the design of new deep learning algorithms.

These technological advances allowed algorithms to process images and sounds, and be integrated into the digital tools of everyday life and into automated professional workflows.

The medical field faces a current massive growth in volume, complexity and heterogeneity of raw data. Effective and medically-focused big data analysis is a major issue in public health. AI offers three promising perspectives in this field: (1) risks prediction via correlations analyses, (2) genomic analysis and phenotype-genotype association studies, (3) automation of medical image analysis. In this report, we focused on the third topic and investigated the following question: ‘Can image recognition deep learning algorithms improve medical visual diagnosis?’ [1].

### **2. Historical background**

The fundamentals of AI were formalized in the 1950s [2]. Initially, two contrasting conceptions of AI were developed: (1) cognitivism, that is the development of rule-based programs referred to as expert systems [3,4] and (2) connexionism, corresponding to the development of naive programs educated secondarily by data. Connexionism currently dominates the world of AI.

Algorithms known as artificial neurons, organized into artificial networks, have led to the design of programs that can be ‘educated’ [5]. The first neural networks were designed in the 1960s. They consisted in sets of artificial neurons organized into superimposed layers. After going through an input layer for presentation purposes, data was analyzed via intermediate layers and ended into an output layer that produced a result. The learning abilities of neural networks were exploited only in the 1990s, with the initiation of ‘deep learning’, which corresponded to the development of very large multi-layered neural networks [6]. In addition to the development of big data analysis and to the increase in computation power, deep learning was boosted in the years 2010 due to the development of a certain type of neural network known as Convolutional Neural Networks (CNN). CNNs had specifically high performances in the field of pattern recognition. Most of the theoretical background of CNNs derives from the achievements of a French computer scientist in the 1980s, Yann LeCun, who created the LeNET network [7]. LeNet was an automated handwriting recognition algorithm, intended to be used by banks for reading checks. In 2010, major international AI teams took part into the ‘ImageNET Challenge’, a competition during which they had the task to classify millions of natural images into thousands of different categories [8]. The current navigation devices embedded into autonomous cars are derived from the networks designed by LeCun nearly forty years ago (Figure 1).

CNNs are inspired by the structure of the primary visual cortex [9]; image recognition proceeds from the automatic extraction of the visual features of images. Automatic extraction is a breakthrough when compared with traditional manual extraction techniques, which were time consuming and less precise [10]. CNNs are based on a specific multilayered architecture that can be trained by data (Figure 2).

Each individual piece of input data corresponds to an image associated with a label (for instance a medical diagnosis for a chest X-ray). The process of training a CNN with labelled images is known as supervised learning. During training, the network analyzes the image, considered as a matrix composed of pixels of varying intensities, by transiting it through the different layers. Some

layers have the ability to modify the input data using features extraction, and are thus able to produce abstract images determined by the level of extracted features; others layers reduce the data into vectors, that are finally interpreted as probabilities by the last layers: ‘Is the image on the chest X-ray a lung cancer?’, ‘Is this stain on the skin a melanoma?’ (Figure 3).

Initially, the naive network, before proper training, labels the input pictures with a certain amount of error. Errors are then back-propagated [11] within the network, which will be able to correct the characteristics of its neurons to match the proper label of the input image. The CNN thus extracts the features of the images uploaded during the training process and keeps the memory of these features into its neural layers. After an efficient training process involving a large input dataset, the network will be able to classify new un-labelled data and provide results as label categories.

## Material and Methods

Here we provide a systematic review of the publications using CNN technology for medical image analysis, available in the National Library of Medicine database (PubMed). The search equation was the following: (convolutional OR deep learning) AND (classification OR detection) AND (image OR photography), filtered for 'Human studies' and 'Title/Abstract' as search fields.

The selected articles were screened according to a standard grid containing the following items: (1) **aim of the study**: detection or classification, (2) **methods**: network architecture, dataset, training, validation, test, comparison method (with specialists or other), (3) **results**: accuracy, sensibility and specificity and (4) **conclusion**. More precisely, within the methods section, the following data regarding network architecture were collected: (a) **pre-training** (yes/no), (b) **network parameters**, (c) **layers** (convolution and pooling, fully connected and softmax) and (d) **software**. Datasets were characterized using the following parameters: (a) **variety / veracity**: origin and certification, (b) **volume**: number, size, pre-training, augmentation, set distribution, (c) **speed** (Graphic Processor Unit (GPU) type) (Figure 4).

## Results

We identified 352 articles in the PubMed database based on our search equation between 01/01/2013 and 05/20/2019. We excluded 327 articles where CNNs performance was not assessed (review articles) or where tasks other than detection or classification, such as segmentation, were assessed (Figure 5). The 25 papers included were ordered according to medical specialties: dermatology [12,13], ophthalmology [14-18], cardiothoracic imaging [19-24], senology [25-27], oral and maxillofacial surgery [28-32], hepato-gastro-enterology [33-36]. Demographic datas about the included papers and informations on the journals in which the works were published are provided in Tables 1 and 2. These results are summarized in Figures 6 and 7, which showed a stable increase of the volume of publications on CNN performances in the last few years: 1 paper in 2013 [27], 8 papers in 2016 [13,15,16,19,22,25,29,33], 6 papers in 2017 [12,14,20,21,26,28], 5 papers in 2018 [17,23,24,30,34] and already 5 papers in 2019 [18,31,32,35,36]. Our results furthermore indicate that most of the papers were produced by North American and Asian teams: 12 papers were from the USA and Canada [12-14,16,20,21,23,24,28,30,31,34] and 13 papers were written by Asian teams [13,16-18,22,26,29-33,35,36]. Interestingly, international collaboration between Asian / South American teams and European / North American teams allowed compiling very large databases in 4 cases [13,16,26,31]. The assessment of the characteristics of the networks (table 3) showed that most teams used previously developed codes such as GoogleNet [12,16,17,20,21,32,36] and AlexNet [19,21,29,33,36]. Most of the 25 networks were pre-trained [12,16-19,21,23-25,30-36], and most teams had included proofs of veracity for the labeling of the images, based on standard methods depending on the dataset used in the study, such as for instance histological confirmation [12,25,27] or assessment by 1-4 specialists (table 4) [14-16,18-22,24,28-30,32,34]. The quantity of included images (table 5) was variable and was not very informative given the variety of the datasets: from 170 pictures of skins lesions used for melanoma detection [13] to 139886 funduscopy images used for diabetic retinopathy diagnosis [16]. Only 5/25 papers

compared the performances of the networks to medical professionals [12,15,16,32,34]; nevertheless, 13/25 studies compared the performances of the networks to traditional detection or classification techniques [13,14,17,22,25-29,32-35]. The results of the quantitative assessment of performances of the networks in all studies were satisfactory; 18 figures were provided by the authors: the lowest precision score was 0.75 [28] and the lowest sensibility was 0.70 [27]. All other values of precision, sensibility, specificity and AUC (area under curve) were between 0.8 and 0.9 [13,19,20,22,25,26,29] or over 0.9 [12,14-18,21,23,24,30-36]. Only 9/25 studies provided visualization methods in order to better understand the inner mechanisms of the network (table 6) [12,14,15,17,18,21,23,27,28].

## Discussion

Most of the studies included into this review have been published within the last three years, and authors are from North America and Asia. These demographics underline the need for European Countries to launch specific plans in order to promote medical IA. Several French initiatives are promising and may contribute to build an EU medical AI community, such as the PaRis Artificial Intelligence Research InstitutE (PRAIRIE), and the development of specific training programs at the University of Paris (the largest medical school in Europe), with a dedicated international Master's Degree, or at CentraleSupélec in association with Inria Saclay (two high-level engineering schools and research centers).

The choice of the network architecture is conditioned by specific tasks. Nevertheless, most authors have used previously developed networks already efficient on natural images, such as *AlexNET* [37] and *GoogLeNET* [38], especially in their pre-trained versions [39]. These networks easily run on softwares as *Caffe* [40], *Theano* [41] or *Tensorflow* [42]. All of these softwares are freely available online and are based on the widespread programming language *Python*.

The data used in the 25 studies included macroscopic and microscopic images of different types: clinical photographs [12-18,30-36], pathology slides [27], X-rays [20,21,23-26,28] and CT images [19,22,29]. Four criteria are often cited in the literature as crucial elements for the design of reliable networks and are referred to as the four 'V'.

(1) **Volume** and **Variety**. The training phase requires a significant number of images. Beyond the quantity, heterogeneous databases increase robustness. Augmentation techniques (rotating - cutting - resizing the images) are often used to increase Volume and Variety without decreasing the third V, **Veracity**.

(2) **Veracity** is conditioned by two quality criteria.

(a) **Image quality** – resolution, limited variability due to angle of view, zoom or brightness – often increased by pre-treatment [43].

(b) **Label quality**, depending on the expertise of the practitioners who created the database. Database review by a scientific committee is thus recommended before training.

(3) **Velocity**, related to power of the processor. GPU, made of several parallel cores, are well fitted for CNNs, specifically during the training phase when a lot of images are simultaneously processed.

CNN training is a key step, for which specific technical skills are required, in order to avoid overfitting [44] on limited data, leading to issues for when using the network to analyze wider datasets. Training thus requires evaluation and monitoring.

Transparency is a fundamental parameter in medical AI. CNNs are opaque structures often referred to as ‘black boxes’. Some systems offer partial visualization techniques (heat maps, probability maps) in order to provide some view of the inner functioning of the CNN. The understanding of how these networks ‘work’ is a relevant and major challenge in medical AI. Nevertheless, using computer vision devices that have ‘non-human’ image analysis abilities is by itself potentially fruitful as hidden correlations between images can potentially be uncovered based on parameters that are not perceived by the human brain, even by the brain of a trained expert.

AI has clearly shown its efficiency in several clinical visual tasks, but comparative clinical studies showing the integration of this technique into clinical workflows are still missing for most applications. Nevertheless, both the robustness of the current results and the potentially simple interfaces that can be designed using trained CNNs provide bases for straightforward, time-saving, reliable and practical applications. CNNs could then be conceived as colleagues providing expert second opinions on tricky clinical questions. In addition, CNNs are intrinsically not subjected to confusing factors such as tiredness, individual beliefs or hierarchical issues, and thus minimize inter- and intra-individual variabilities when achieving a specific task.

However, adverse effects of the use of CNNs consist in deskilling, phenomenon described during the automation of any type of task. Another non-intended consequence of the clinical use of CNNs could be an overreliance to the machine by the practitioner. This could lead to the omission of crucial factors, such as clinical data, that the machine does not integrate into its decision-making process. Indeed, CNNs answer specific questions and sole visual criteria remain insufficient to perform a diagnosis in most cases. Some authors have suggested combining CNNs with algorithms screening other categories of data (for instance clinical, biological, molecular). In brief, networks provide a certain decision with a given probability of accurateness, which is then to integrate into a wider array of diagnostic arguments.

A key point of medical application of CNNs is that the process directly involves the patients themselves. Melanoma screening methods can be used as self-medicine devices, and more generally, CNNs can be seen as solutions improving management and prognosis by promoting earlier diagnosis and earlier treatment. This point raises the confidence patients will place into these devices: excessive or minimal confidence in CNNs may interfere within the relationship between patients and doctors.

The medical use of CNNs will only be successful via collaboration between several actors, each having a specific and fundamental role in the process.

- (1) **Patients.** They are the source of the data and their consent is required for its use.
- (2) **Practitioners.** They define relevant medical questions for which CNNs could be of use and collect the data, which they interpret and label. They are then crucial for the validation of the CNNs in a clinical environment.
- (3) **Engineers.** Specific training programs for medical AI are required, based on collaborations between academic hospitals, medical schools and engineering schools.
- (4) **Entrepreneurs and investors.** By partnering with medical doctors, investors create the financial conditions required to hire competent engineers and benefit from powerful datacenters and computation power.

(5) **Lawmakers.** Data protection institutions, such as the Commission Nationale de l'Informatique et des Libertés (CNIL) and the Institut National des Données de Santé (INDS) in France supervise the use of clinical data and the application of these medical devices. Collaborations with the national medical councils (Conseil National de l'Ordre des Médecins in France) will result in the formulation of ethical and legal questions, mostly regarding medical.

With the help of the digitization of the medical data, all the actors involved in the design and practical application of CNNs will interact in a fruitful manner, thanks to the optimization of data collection, interpretation, storage, sharing and use. Multi-centric platforms will contribute to augment data volume and veracity.

## **Conclusion**

CNNs are not replacement solutions for medical doctors, but will contribute to optimize routine tasks and thus have a potential positive impact on our practice. Specialties with a strong visual component such as radiology and pathology will be deeply transformed by CNNs but all the fields of medical and surgical practice will be affected by this technology. The role of practitioners is key for the development and implementation of such devices. Medical doctors currently have a historical chance to take part into a scientific revolution by understanding deep learning, taking part in the conception and evaluation of new devices but also by contribution to conceive a framework for the regulation of this new type of medical activity.

## References

1. Fourcade A, Khonsari RH. Apprentissage profond : un troisième œil pour les médecins. Université Paris-Est Créteil UPEC; 2017.
2. McCarthy J, Minsky M, Rochester N, Shannon C. A proposal for the Dartmouth Summer Research Project on Artificial Intelligence. 1955.
3. Miller R, Pople HJ, Myers J. Internist-1, an experimental computer-based diagnostic consultant for general internal medicine. *N Engl J Med.* 1982;307(8):468-76.
4. Shortliffe E, Davis R, Axline S, Buchanan B, Green C, Cohen S. Computer-based consultations in clinical therapeutics: Explanation and rule acquisition capabilities of the MYCIN system. *Comput Biomed Res.* 1975;8(4):303-20.
5. Rosenblatt F. The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychol Rev.* 1958;65(6):386-408.
6. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521(7553):436-44.
7. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* 1989;1(4):541-51.
8. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. In: *Conference on Computer Vision and Pattern Recognition.* 2009.
9. Hubel D, Wiesel T. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Physiol.* 1962;160:106-54.
10. Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. *Annu Rev Biomed Eng.* 2017;19(1):221-48.

11. Rumelhart D, Hinton G, Williams R. Learning representations by back-propagating errors. *Nature*. 1986;323:533-6.
12. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115-8.
13. Nasr-Esfahani E, Samavi S, Karimi N, Soroushmehr SMR, Jafari MH, Ward K, et al. Melanoma detection by analysis of clinical images using convolutional neural network. In: *IEEE Engineering in Medicine and Biology Society*. 2016. p. 1373-6.
14. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*. 2017;124(7):962-9.
15. Grinsven MJJP Van, Ginneken B Van, Hoyng CB, Theelen T, Clara IS. Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images. *IEEE Trans Med Imaging*. 2016;35(5):1273-84.
16. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402-10.
17. Ahn JM, Kim S, Ahn KS, Cho SH, Lee KB, Kim US. A deep learning model for the detection of both advanced and early glaucoma using fundus photography. *PLOS ONE*. 2018;13(11):e0207982.
18. Phan S, Satoh S, Yoda Y, Kashiwagi K, Oshika T. *Jpn J Ophthalmol* (2019) 63: 276.
19. Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans Med Imaging*. 2016;35(5):1207-16.

20. Cicero M, Bilbily A, Colak E, Dowdell T, Gray B, Perampaladas K, et al. Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs. *Invest Radiol.* 2017;52(5):281-7.
21. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology.* 2017;284(2):574-82.
22. Li W, Cao P, Zhao D, Wang J. Pulmonary nodule classification with deep convolutional neural networks on computed tomography images. *Comput Math Methods Med.* 2016;6215085.
23. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLOS MED.* 2018;15(11): e1002683
24. Taylor AG, Mielke C, Mongan J. Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks: A retrospective study. *PLOS MED.* 2018;15(11):e1002697.
25. Arevalo J, Gonzalez F, Ramos-Pollan R, Oliveira J, Angel M, Guevara Lopez M. Representation learning for mammography mass lesion classification with convolutional neural networks. *Comput Methods Programs Biomed.* 2015;127:248-57.
26. Jadoon MM, Zhang Q, Haq IU, Butt S, Jadoon A. Three-class mammogram classification based on descriptive CNN features. *Biomed Res Int.* 2017;3640901.
27. Ciresan C, Giusti A, Gambardella L, Schmidhuber J. Mitosis detection in breast cancer histology images with deep neural networks. *Med Image Comput Comput Assist Interv.* 2013;16(2):411-8.
28. Arik SÖ, Ibragimov B, Xing L. Fully automated quantitative cephalometry using convolutional neural networks. *J Med Imaging.* 2017;4(1):1-11.

29. Miki Y, Muramatsu C, Hayashi T, Zhou X, Hara T, Katsumata A, et al. Classification of teeth in cone-beam CT using deep convolutional neural network. *Comput Biol Med*. 2017;1(80):24-9.
30. Uthoff RD, Song B, Sunny S, Patrick S, Suresh A, et al. Point-of-care, smartphone-based, dual-modality, dual-view, oral cancer screening device with neural network classification for low-resource communities. *PLOS ONE*. 2018;13(12): e0207493.
31. Gurovich Y, Hanani Y, Bar O, Nadav G, Fleischer N, Gelbman D, et al. Identifying facial phenotypes of genetic disorders using deep learning. *Nature Medicine*. 2019;25(1): 60–64.
32. Jeyaraj PR, Samuel Nadar ER. *J Cancer Res Clin Oncol* (2019) 145: 829.
33. Jia X, Meng Q. A deep convolutional neural network for bleeding detection in wireless capsule endoscopy images. In: *Conf Proc IEEE Eng Med Biol Soc*. 2016. p. 639-42.
34. Urban G, Tripathi T, Alkayali T, Mittal M, Jalali F, Karnes W, et al. Deep Learning Localizes and Identifies Polyps in Real Time With 96% Accuracy in Screening Colonoscopy. *Gastroenterology*. 2018;155:1069–1078.
35. Horie Y, Yoshio T, Aoyama K, Yoshimizu S, Horiuchi Y, et al. Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. *Gastrointestinal Endoscopy*. 2019;89(1):25-32.
36. Alaskar H, Hussain A, Al-Aseem N, Liatsis P, Al-Jumeily D. Application of Convolutional Neural Networks for Automated Ulcer Detection in Wireless Capsule Endoscopy Images. *Sensors*. 2019;19(6):1265.
37. Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks. In: *Neural Information Processing Systems*. 2012. p. 1097-105.

38. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: Computer Vision and Pattern Recognition. 2015. p. 2818-26.
39. Shin H, Roth HR, Gao M, Lu L, Member S, Xu Z, et al. Deep convolutional neural networks for Computer-Aided Detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging. 2016;35(5):1285-98.
40. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, et al. Caffe: convolutional architecture for fast feature embedding. In: ACM International Conference on Multi-media. 2014. p. 675-8.
41. Bastien F, Lamblin P, Pascanu R, Bergstra J, Goodfellow I, Bergeron A, et al. Theano: new features and speed improvements. Deep Learn Unsupervised Featur Learn. 2012;1211.5590.
42. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. Tensorflow: large-scale machine learning on heterogeneous distributed systems. arXiv. 2016;1603.04467.
43. Kohli M, Prevedello LM, Filice RW, Geis JR. Implementing machine learning in radiology practice and research. AJR Am J Roentgenol. 2017;208(4):754-60.
44. Caruana R, Lawrence S, Giles L. Overfitting in neural nets: backpropagation, conjugate gradient, and early stopping. In: International Conference on Neural Information Processing Systems. 2001. p. 402-8.

## **Figure legends**

*Figure 1.* Concepts and theories in deep learning.

*Figure 2.* Structure of a typical artificial neural network: the multilayered Perceptron.

*Figure 3.* Structure of a Convolutional Neural Network (CNN).

*Figure 4.* Canvas for data extraction from the included articles.

*Figure 5.* Inclusion process for the systematic review.

*Figure 6.* Geographical origin of the included publications.

*Figure 7.* Number of scientific publications per year related to medical imaging interpretation and deep learning.

## **Table legends**

**Table 1.** Systematic review of deep learning applications in medical image analysis. Articles assessing detection and classification abilities of neural networks. Year of publication - Journal - Impact Factor.

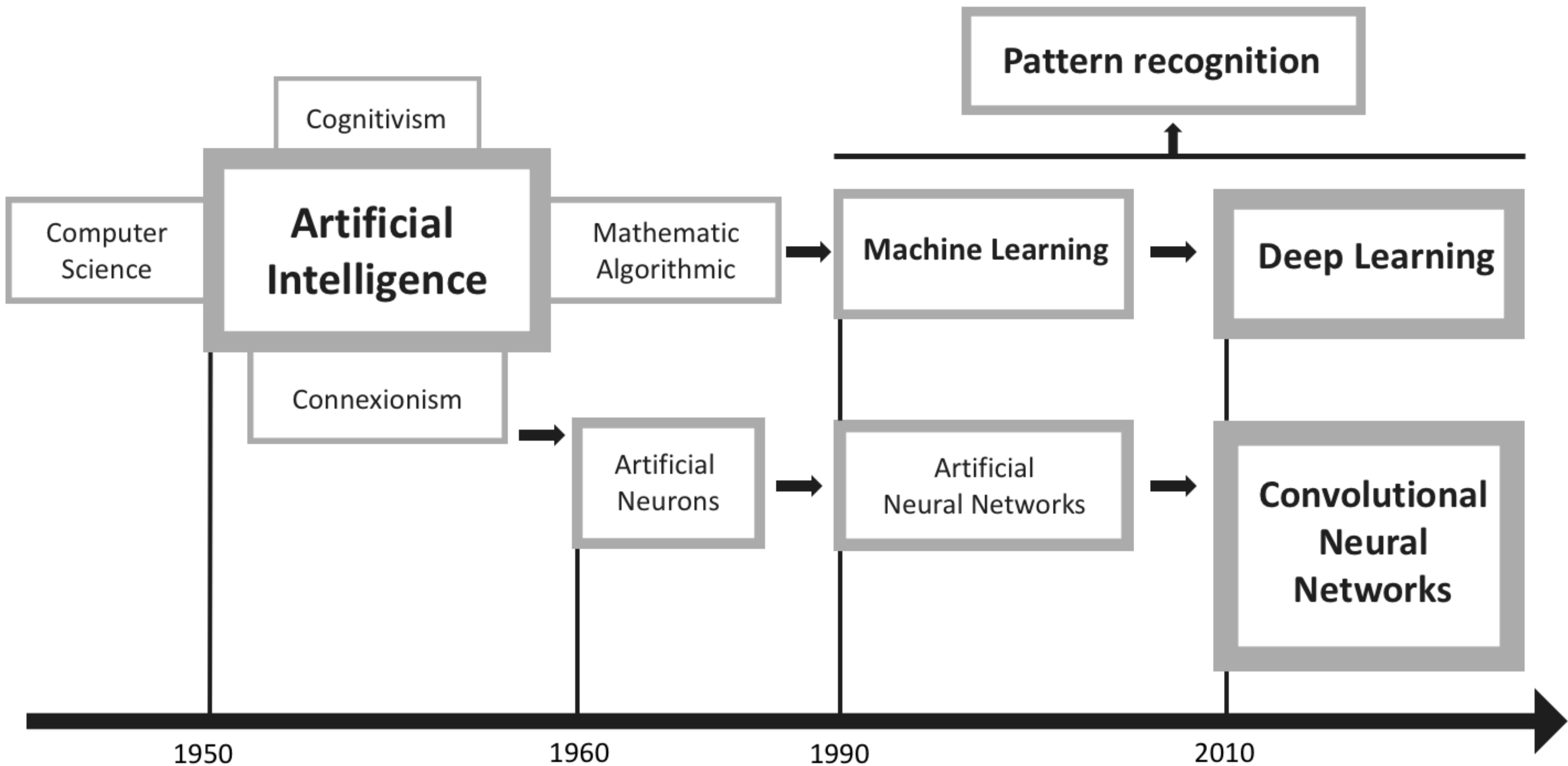
**Table 2.** Systematic review of deep learning applications in medical image analysis. Articles assessing detection and classification abilities of neural networks. Origin - Team - Objective.

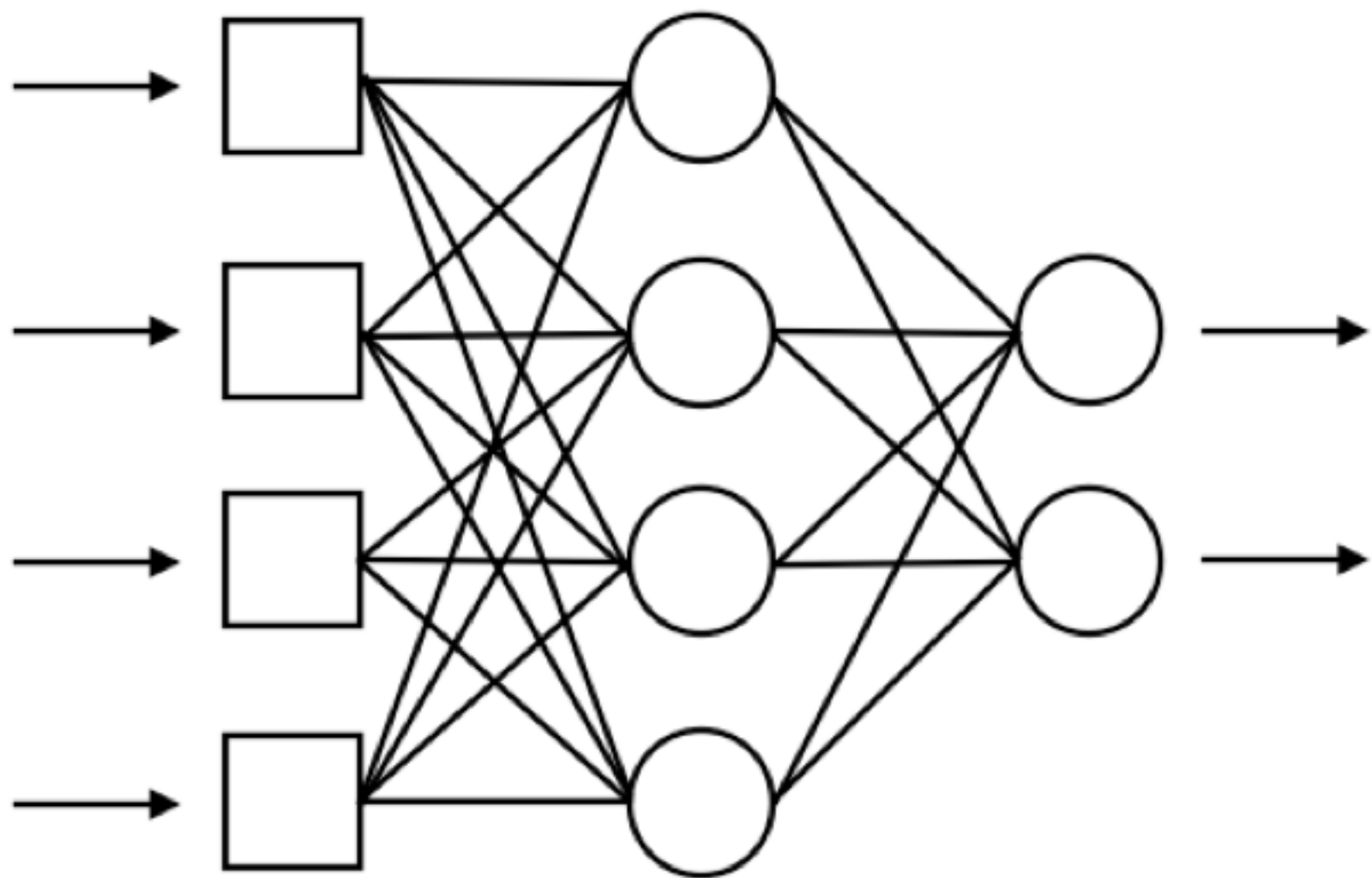
**Table 3.** Systematic review of deep learning applications in medical image analysis. Articles assessing detection and classification abilities of neural networks. Network - Training step.

**Table 4.** Systematic review of deep learning applications in medical image analysis. Articles assessing detection and classification abilities of neural networks. Dataset quality

**Table 5.** Systematic review of deep learning applications in medical image analysis. Articles assessing detection and classification abilities of neural networks. Dataset size.

**Table 6.** Systematic review of deep learning applications in medical image analysis. Articles assessing detection and classification abilities of neural networks. Results and comparisons.

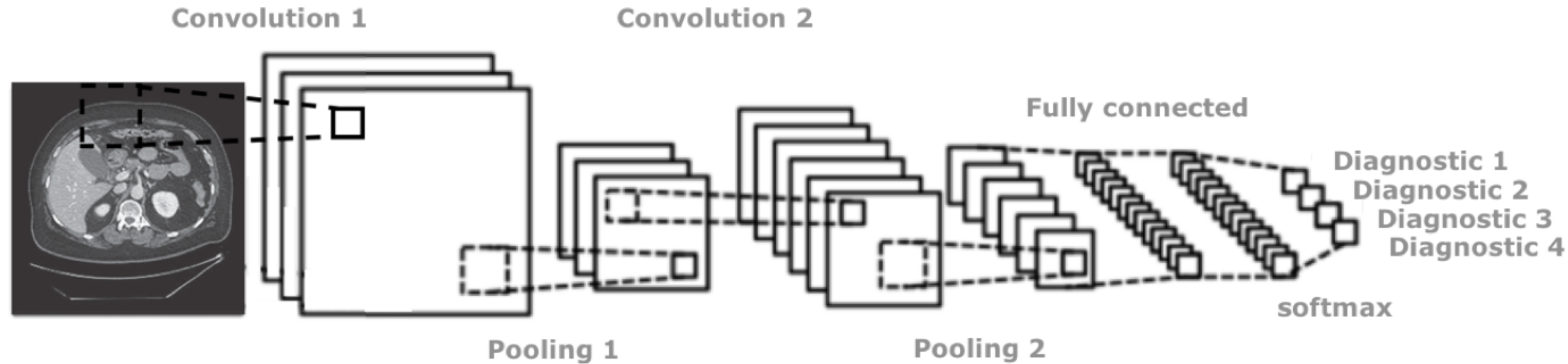




Input  
layer

Hidden  
layer

Output  
layer



## Objective

Detection and/or classification

## Methods

- Network architecture :
  - o Pre-trained or from scratch
  - o Parameters
  - o Layers :
    - Convolution and pooling
    - Fully connected and softmax
  - o Software
- Data set :
  - o Variety / Veracity
    - Origin
    - Certification
  - o Volume
    - Number
    - Size
    - Pre-training
    - Augmentation
    - Sets distribution
  - o Speed (*GPU* used)
- Training
- Validation
- Test
- Comparison : specialists and/or other techniques

## Results

- Accuracy
- Sensibility and specificity - PPV and NPV - AUC

## Conclusion

352 articles identified

- 1) *Convolutional* OR *Deep learning*
- 2) *Classification\** OR *Detection\**
- 3) *Image\** OR *Photography\**

1 AND 2 AND 3

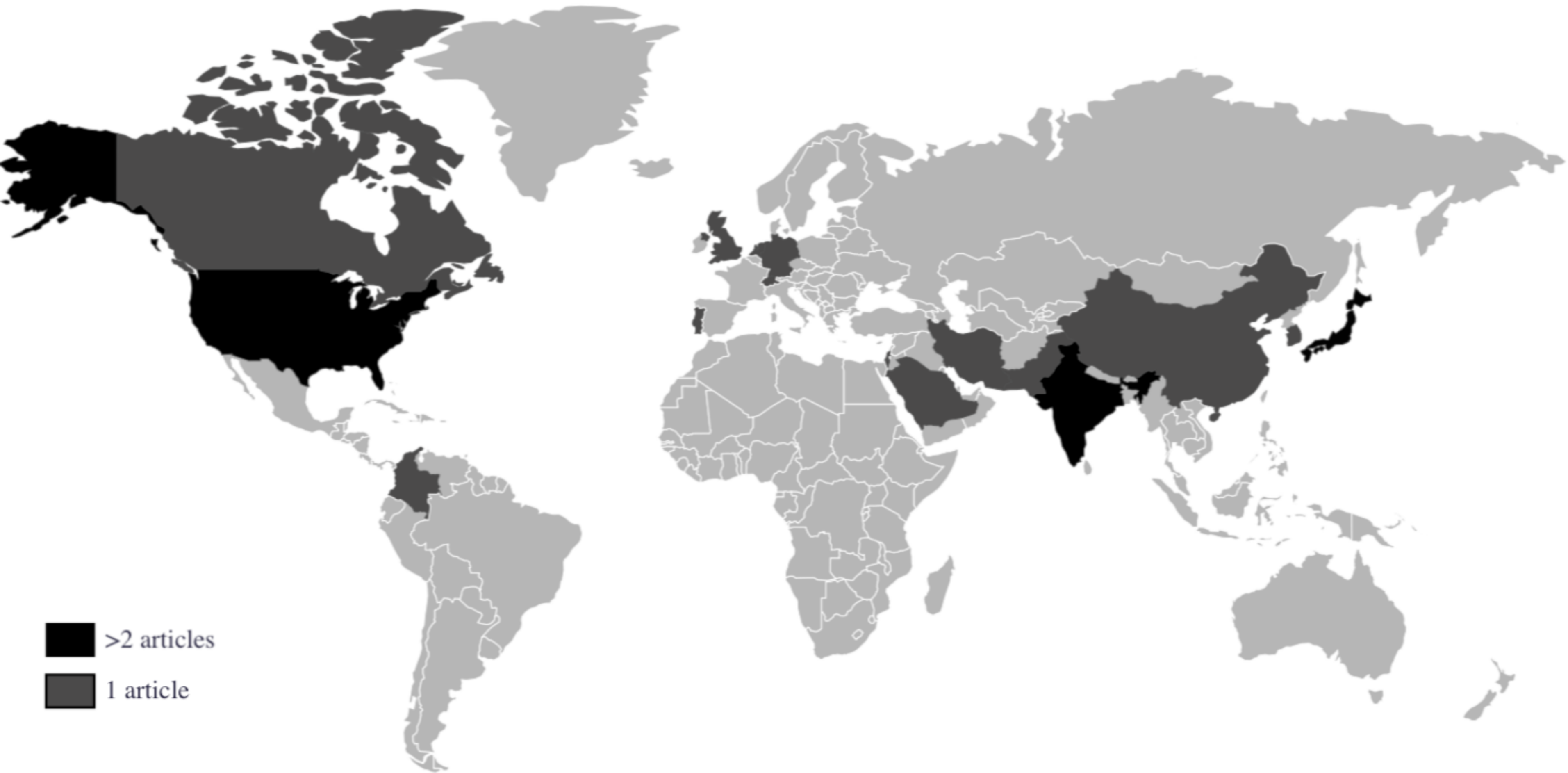
327 articles excluded

- CNN use for another task that classification or detection in medical images (ex segmentation)
- No evaluation of the CNN performance

25 articles selected

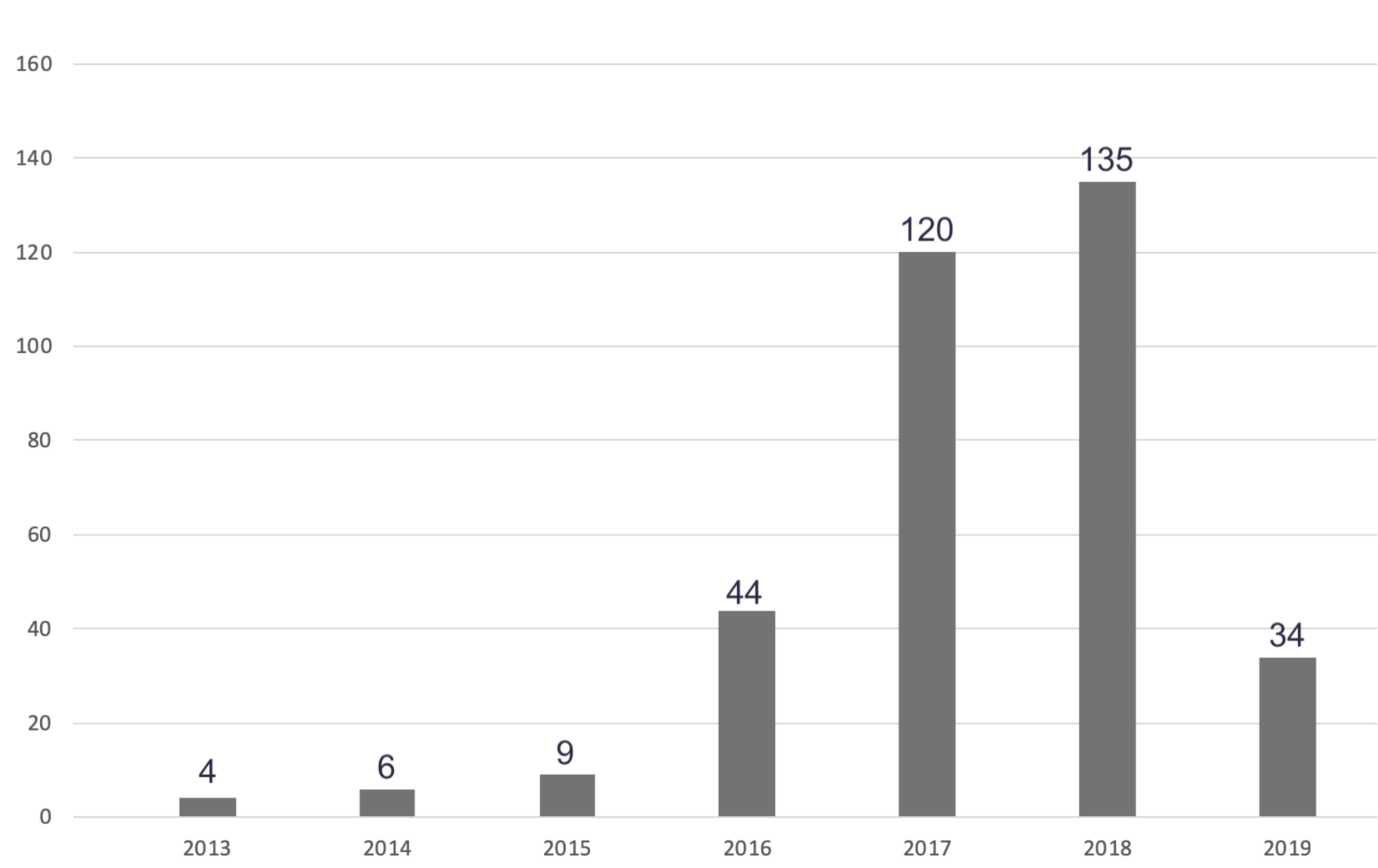
Evaluation of the CNN performance in tasks of :

- Lesions detection in medical image
- Medical image classification



■ >2 articles

■ 1 article



**Table 1.** Systematic review of deep learning applications in medical image analysis. Articles assessing detection and classification abilities of neural networks. Year of publication - Journal - Impact Factor.

<b>Authors</b>	<b>Year</b>	<b>Journal</b>	<b>Impact Factor</b>
<b>DERMATOLOGY</b>			
<i>Esteva &amp; al</i> (12)	2017	Nature Medecine	41,577
<i>Nasr-esfahani &amp; al</i> (13)	2016	IEEE Engineering in Medicine and Biology Society	3.05
<b>OPHTALMOLOGY</b>			
<i>Gargeya &amp; al</i> (14)	2017	Journal of American Academy of Ophthalmology	8.2
<i>Grinsven &amp; al</i> (15)	2016	IEEE Transactions on Medical Imaging	3.942
<i>Gulshan &amp; al</i> (16)	2016	Journal of American Medical Association	44.4
<i>Ahn &amp; al</i> (17)	2018	PLOS Medecine	11.675
<i>Phan &amp; al</i> (18)	2019	Japanese Journal of Ophtalmology	1.775
<b>THORACIC</b>			
<i>Anthimopoulos &amp; al</i> (19)	2016	IEEE Transactions on Medical Imaging	3.942
<i>Cicero &amp; al</i> (20)	2017	Investigative Radiology	5.195
<i>Lakhani &amp; al</i> (21)	2017	Radiology	7.296
<i>Li &amp; al</i> (22)	2016	Computational and Mathematical Methods in Medicine	0.937
<i>Zech &amp; al</i> (23)	2018	PLOS medecine	11.675
<i>Taylor &amp; al</i> (24)	2018	PLOS medecine	11.675
<b>SENOLOGY</b>			
<i>Arevalo &amp; al</i> (25)	2016	Computer Methods and Programs in Biomedicine	1.862
<i>Jadoon &amp; al</i> (26)	2017	Biomed Research International	2.476
<i>Ciresan &amp; al</i> (27)	2013	Medical Image Computing and Computer-Assisted Intervention	-
<b>O-M-F SURGERY</b>			

<i>Arik &amp; al</i> (28)	2017	Journal of Medical Imaging	1.109
<i>Miki &amp; al</i> (29)	2016	Computers in Biology and Medicine	1.836
<i>Uthoff &amp; al</i> (30)	2018	PLOS One	2.766
<i>Gurovich &amp; al</i> (31)	2019	Nature Medecine	41.577
<i>Jeyaraj &amp; al</i> (32)	2019	Journal of Cancer Research and Clinical Oncology	3.081
<b>GASTRO-ENTEROLOGY</b>			
<i>Jia &amp; al</i> (33)	2016	Engineering in Medicine and Biology Society	0.76
<i>Urban &amp; al</i> (34)	2018	Gastroenterology	20.877
<i>Horie &amp; al</i> (35)	2019	Gastrointestinal Endoscopy	5.369
<i>Alaskar &amp; al</i> (36)	2019	Sensors	2.475

*Table 1. Systematic review of deep learning applications in medical image analysis. Articles assessing detection and classification abilities of neural networks. Origin - Team - Objective.*

<b>Authors</b>	<b>Origin</b>	<b>Team</b>	<b>Objective</b>
<b>DERMATOLOGY</b>			
<i>Esteva &amp; al</i> (12)	California, USA	Mixed <sup>1</sup>	Skin lesion classification
<i>Nasr-esfahani &amp; al</i> (13)	Isfahan, Iran / Michigan, USA	Mixed	Melanoma detection
<b>OPHTALMOLOGY</b>			
<i>Gargeya &amp; al</i> (14)	California, USA	Mixed	Diabetic retinopathy detection
<i>Grinsven &amp; al</i> (15)	Nijmegen, Netherlands	Engineers	Retinal hemorrhages detection
<i>Gulshan &amp; al</i> (16)	California, USA / Texas, USA / Madurai, India	Mixed	Diabetic retinopathy detection
<i>Ahn &amp; al</i> (17)	Seoul, Korea	Mixed	Glaucoma detection
<i>Phan &amp; al</i> (18)	Tokyo, Japan	Mixed	Glaucoma detection
<b>THORACIC</b>			
<i>Anthimopoulos &amp; al</i> (19)	Bern, Switzerland	Engineers	Interstitial lung disease classification
<i>Cicero &amp; al</i> (20)	Toronto, Canada	Mixed	Thoracic disease classification
<i>Lakhani &amp; al</i> (21)	Pennsylvania, USA	Mixed	Tuberculosis detection
<i>Li &amp; al</i> (22)	Shenyang, China	Mixed	Pulmonary node detection
<i>Zech &amp; al</i> (23)	California, USA	Mixed	Pneumonia detection
<i>Taylor &amp; al</i> (24)	California USA	Mixed	Pneumothorax detection
<b>SENOLOGY</b>			
<i>Arevalo &amp; al</i> (25)	Bogota, Colombia / Aveiro, Portugal	Mixed	Lesions classification on mammograms
<i>Jadoon &amp; al</i> (26)	London, UK / Islamabad, Pakistan	Engineers	Lesions classification on mammograms
<i>Ciresan &amp; al</i> (27)	Lugano, Switzerland	Engineers	Mitosis detection on histological slices of breast tumors
<b>O-M-F SURGERY</b>			
<i>Arik &amp; al</i> (28)	California, USA	Mixed	Anatomical point-of-interest detection and classification

<sup>1</sup> Team of engineers and medical doctors.

<i>Miki &amp; al</i> (29)	Gifu, Japan	Mixed	Teeth classification
<i>Uthoff &amp; al</i> (30)	Arizona, USA / Bangalore, India	Mixed	Oral cancer detection
<i>Gurovich &amp; al</i> (31)	Israel / Germany / USA	Mixed	Facial phenotyping of genetic disorders
<i>Jeyaraj &amp; al</i> (32)	Sivakasi, India	Engineers	Oral cancer detection
<b>GASTRO-ENTEROLOGY</b>			
<i>Jia &amp; al</i> (33)	Hong Kong, Hong Kong	Engineers	Gastrointestinal bleeding detection
<i>Urban &amp; al</i> (34)	California, USA	Mixed	Polyps detection
<i>Horie &amp; al</i> (35)	Japan	Mixed	Esophageal cancer detection
<i>Alaskar &amp; al</i> (36)	Alkharj, Saudi Arabia / Liverpool, UK	Engineers	Esophageal and gastric ulcer detection

*Table 1. Systematic review of deep learning applications in medical image analysis. Articles assessing detection and classification abilities of neural networks. Network - Training step.*

<b>Authors</b>	<b>Networks</b>	<b>Pre-training</b>	<b>Transfer-learning</b>	<b>Software</b>	<b>Hardware</b>
<b>DERMATOLOGY</b>					
<i>Esteva &amp; al</i> (12)	<i>GoogleNet Inception v3</i>	Yes ( <i>Imagenet</i> )	Yes ( <i>fine tuning</i> )	<i>Google TensorFlow</i>	-
<i>Nasr-esfahani &amp; al</i> (13)	Personal	No	No	-	2 GPU
<b>OPHTHALMOLOGY</b>					
<i>Gargeya &amp; al</i> (14)	5	No	No	-	2 CPU or <i>iPhone</i>
<i>Grinsven &amp; al</i> (15)	<i>OxfordNet</i>	No	No	-	1 GPU
<i>Gulshan &amp; al</i> (16)	<i>GoogleNet Inception v3</i>	Yes ( <i>Imagenet</i> )	Yes ( <i>fine tuning</i> )	<i>Google TensorFlow</i>	-
<i>Ahn &amp; al</i> (17)	Personal and <i>GoogleNet Inception v3</i>	Yes	Yes	<i>Google TensorFlow</i>	-
<i>Phan &amp; al</i> (18)	<i>VGG19, ResNet152 and DenseNet201</i>	Yes	Yes	-	-
<b>THORACIC</b>					
<i>Anthimopoulos &amp; al</i> (19)	Personal, <i>AlexNet</i> and <i>VGG-Net</i>	<i>AlexNet-PT</i>	<i>AlexNet-PT</i>	<i>Theano &amp; Caffe</i>	1 GPU
<i>Cicero &amp; al</i> (20)	<i>GoogleNet Inception v3</i>	No	No	<i>Caffe</i>	3 GPU
<i>Lakhani &amp; al</i> (21)	<i>GoogleNet Inception v3 and AlexNet</i>	Yes ( <i>Imagenet</i> ) & no	Yes ( <i>fine tuning</i> ) & no	<i>Caffe</i>	1 GPU
<i>Li &amp; al</i> (22)	Personal	No	No	-	2 CPU
<i>Zech &amp; al</i> (23)	<i>ResNet</i> and <i>DenseNet VGG16/19,</i>	Yes	Yes	<i>Pytorch</i>	-
<i>Taylor &amp; al</i> (24)	<i>Xception, Inception, and ResNet</i>	Yes	Yes	<i>Google TensorFlow</i>	8 GPU
<b>SENOLOGY</b>					
<i>Arevalo &amp; al</i> (25)	Personal	Both	No	<i>Theano</i>	1 GPU

<i>Jadoon &amp; al</i> (26)	<i>CNN-CT</i> and <i>CNN-WT</i>	No	No	-	-
<i>Ciresan &amp; al</i> (27)	Personal	No	No	-	1 GPU
<b>O-M-F SURGERY</b>					
<i>Arik &amp; al</i> (28)	Personal	No	No	<i>MATLAB</i>	-
<i>Miki &amp; al</i> (29)	<i>AlexNet</i>	No	No	<i>Caffe</i>	1 GPU
<i>Uthoff &amp; al</i> (30)	<i>VGG</i>	Yes	Yes	-	-
<i>Gurovich &amp; al</i> (31)	-	Yes	Yes	-	-
<i>Jeyaraj &amp; al</i> (32)	<i>GoogleNet</i> <i>Inception v3</i>	Yes	Yes	-	1 GPU
<b>GASTRO-ENTEROLOGY</b>					
<i>Jia &amp; al</i> (33)	<i>AlexNet</i>	No	No	<i>Caffe</i>	1 GPU
<i>Urban &amp; al</i> (34)	<i>VGG</i> and <i>ResNET</i>	Yes	Yes	-	1 GPU
<i>Horie &amp; al</i> (35)	Personal	Yes	Yes	-	-
<i>Alaskar &amp; al</i> (36)	<i>AlexNet</i> and <i>GoogleNet</i>	Yes	Yes	-	-

*Table 1. Systematic review of deep learning applications in medical image analysis. Articles assessing detection and classification abilities of neural networks. Dataset quality.*

Authors	Data type	Dataset	Veracity	Size	Pre-processing
<b>DERMATOLOGY</b>					
<i>Esteva &amp; al</i> (12)	Skin lesion photographs	21	Histological proof	299x299	No
<i>Nasr-esfahani &amp; al</i> (13)	Skin lesion photographs	1	-	188x188	Yes
<b>OPHTALMOLOGY</b>					
<i>Gargeya &amp; al</i> (14)	Retinal fundus photographs	3	Labelling	512x512x3	Yes
<i>Grinsven &amp; al</i> (15)	Retinal fundus photographs	2	Proofreading by 3 ophthalmologists	512x512	Yes
<i>Gulshan &amp; al</i> (16)	Retinal fundus photographs	6	Proofreading by 54 ophthalmologists	299x299	No
<i>Ahn &amp; al</i> (17)	Retinal fundus photographs	1	-	224x224	Yes
<i>Phan &amp; al</i> (18)	Retinal fundus photographs	2	Proofreading by glaucoma expert	256x256 512x512	Yes
<b>THORACIC</b>					
<i>Anthimopoulos &amp; al</i> (19)	Computed tomography slices	2	Proofreading by radiologists	32x32	ROI
<i>Cicero &amp; al</i> (20)	Chest X-ray	1	Proofreading by 2 radiologists	256x256	No
<i>Lakhani &amp; al</i> (21)	Chest X-ray	4	Proofreading by 1 radiologist	256x256	No
<i>Li &amp; al</i> (22)	Computed tomography slices	1	Proofreading by 4 radiologists	32x22	ROI
<i>Zech &amp; al</i> (23)	Chest X-ray	3	-	224x224	Yes
<i>Taylor &amp; al</i> (24)	Chest X-ray	1	Proofreading by 6 radiologists	512x512	Yes
<b>SENOLOGY</b>					
<i>Arevalo &amp; al</i> (25)	Mammogram images	1	Histological proof	150x150	Yes
<i>Jadoon &amp; al</i> (26)	Mammogram images	4	-	128x128	Yes
<i>Ciresan &amp; al</i> (27)	Histological slice photographs	1	Proofreading by pathologists	2084x2084	Mitosis centered images
<b>O-M-F SURGERY</b>					

<i>Arik &amp; al</i> (28)	Lateral cephalogram	3	POI placed by 2 specialists	81x81	POI centered images
<i>Miki &amp; al</i> (29)	Axial slice CBCT	2	Manually defined ROI	227x227	Yes
<i>Uthoff &amp; al</i> (30)	Oral lesion photographs	1	Proofreading by specialist	-	Yes
<i>Gurovich &amp; al</i> (31)	Facial photographs	1	-	-	Yes
<i>Jeyaraj &amp; al</i> (32)	Oral lesion photographs	1	Proofreading by specialist	250x250	Yes
<b>GASTRO-ENTEROLOGY</b>					
<i>Jia &amp; al</i> (33)	Wireless capsule Endoscopy images	1	-	240x240x3	No
<i>Urban &amp; al</i> (34)	Colonoscopy images	1	Proofreading by 3 experts colonoscopists	224x224	Yes
<i>Horie &amp; al</i> (35)	Endoscopic images	1	-	-	Yes
<i>Alaskar &amp; al</i> (36)	Wireless capsule Endoscopy images		-	224x224 and 227x227	No

*Table 1. Systematic review of deep learning applications in medical image analysis. Articles assessing detection and classification abilities of neural networks. Dataset size.*

<b>Authors</b>	<b>Initial dataset</b>	<b>Augmentation techniques</b>	<b>Training &amp; validation dataset</b>	<b>Test dataset</b>
<b>DERMATOLOGY</b>				
<i>Esteva &amp; al</i> (12)	129405	Yes	127463	1942
<i>Nasr-esfahani &amp; al</i> (13)	170	Yes (x36)	4896	1224
<b>OPHTALMOLOGY</b>				
<i>Gargeya &amp; al</i> (14)	77348	Yes	75137	1748 + 463
<i>Grinsven &amp; al</i> (15)	7879	Yes	5287	2592
<i>Gulshan &amp; al</i> (16)	139886	No	128175	9963 + 1748
<i>Ahn &amp; al</i> (17)	1542	Yes	1078	464
<i>Phan &amp; al</i> (18)	3312	Yes	75%	25%
<b>THORACIC</b>				
<i>Anthimopoulos &amp; al</i> (19)	14696 images from 120 CT scans	Yes	13646	1050
<i>Cicero &amp; al</i> (20)	35038	Yes	32586	2443
<i>Lakhani &amp; al</i> (21)	1007	Both	857	150
<i>Li &amp; al</i> (22)	62492 images from 1010 CT scans	No	54680	7811
<i>Zech &amp; al</i> (23)	158323	-	80%	20 %
<i>Taylor &amp; al</i> (24)	13292 (3107 pneumothorax)	Yes	85%	15%
<b>SENOLOGY</b>				
<i>Arevalo &amp; al</i> (25)	736	Yes (x8)	442	294
<i>Jadoon &amp; al</i> (26)	2796	Yes (x7)	-	-
<i>Ciresan &amp; al</i> (27)	300 images of mitosis from 50 histological slices	Yes	35 (66000 mitosis & 151 million non-mitosis)	15
<b>O-M-F SURGERY</b>				

<i>Arik &amp; al</i> (28)	19 POI images from 400 X-ray	Yes (x25)	150	250
<i>Miki &amp; al</i> (29)	35259 slices from 52 CBCT	Yes	40 CBCT	12 CBCT
<i>Uthoff &amp; al</i> (30)	170	Yes (x8)	-	-
<i>Gurovich &amp; al</i> (31)	17000	No		502
<i>Jeyaraj &amp; al</i> (32)	500	-	-	-
<b>GASTRO- ENTEROLOGY</b>				
<i>Jia &amp; al</i> (33)	10000	Yes	8200	1800
<i>Urban &amp; al</i> (34)	2000 coloscopy (8641 images)	-	2000	20
<i>Horie &amp; al</i> (35)	9546	-	8428	1118
<i>Alaskar &amp; al</i> (36)	1875	-	421	105

**Table 1.** Systematic review of deep learning applications in medical image analysis. Articles assessing detection and classification abilities of neural networks. Results and comparisons. AUC: area under curve; Sp: specificity; Sn: sensitivity.

Authors	Comparison to specialists	Comparison to traditional techniques	Associated techniques	Results	Visualisation techniques
<b>DERMATOLOGY</b>					
<i>Esteva &amp; al</i> (12)	21 dermatologists	No	No	AUC 0.96 & 0.94	Yes
<i>Nasr-esfahani &amp; al</i> (13)	No	Yes	No	Precision 0.81	No
<b>OPHTHALMOLOGY</b>					
<i>Gargeya &amp; al</i> (14)	No	Yes	No	AUC 0.95	Yes (heat-map)
<i>Grinsven &amp; al</i> (15)	2 ophtalmologists	No	No	AUC 0.97	Yes
<i>Gulshan &amp; al</i> (16)	15 ophtalmologists	No	No	AUC 0.99	No
<i>Ahn &amp; al</i> (17)	No	Yes	No	AUC 0.93	Yes
<i>Phan &amp; al</i> (18)	No	No	No	AUC >0.9	Yes (heat-map)
<b>THORACIC</b>					
<i>Anthimopoulos &amp; al</i> (19)	No	No	No	Precision 0.856	No
<i>Cicero &amp; al</i> (20)	No	No	No	AUC between 0.85 & 0.96	No
<i>Lakhani &amp; al</i> (21)	No	No	2 CNN & Radiologists + CNN	AUC 0.98	Yes (heat-map)
<i>Li &amp; al</i> (22)	No	Yes	No	Precision 0.864	No
<i>Zech &amp; al</i> (23)	No	No	Yes	AUC 0,931	Yes (heat-map)
<i>Taylor &amp; al</i> (24)	No	No	No	AUC 0.94	No
<b>SENOLOGY</b>					
<i>Arevalo &amp; al</i> (25)	No	Yes	Yes	AUC 0.860	No
<i>Jadoon &amp; al</i> (26)	No	Yes	No	AUC 0.855	No

<i>Ciresan &amp; al</i> (27)	No	Yes	No	Sn 0.70 - Sp 0.88	Yes
<b>O-M-F SURGERY</b>					
<i>Arik &amp; al</i> (28)	No	Yes	No	Precision 0.75	Yes
<i>Miki &amp; al</i> (29)	No	Yes	No	Precision 0.88	No
<i>Uthoff &amp; al</i> (30)	No	No	No	AUC 0.908	No
<i>Gurovich &amp; al</i> (31)	No	No	No	Accuracy 91%	No
<i>Jeyaraj &amp; al</i> (32)	Yes	Yes	Yes	Accuracy 94,5%	No
<b>GASTRO- ENTEROLOGY</b>					
<i>Jia &amp; al</i> (33)	No	Yes (2)	Yes	Precision 0.999	No
<i>Urban &amp; al</i> (34)	Yes	Yes	No	Accuracy 96,4% AUC 0,991	No
<i>Horie &amp; al</i> (35)	-	No	No	Sn 98%	No
<i>Alaskar &amp; al</i> (36)	No	Yes	No	Accuracy 100%	No