A survey of deep learning models in medical therapeutic areas

Alberto Nogales,[1] Álvaro García-Tejedor,[1] Diana Monge,[2] Juan Serrano Vara[1] and Cristina Antón[2]

[1] CEIEC, Research Institute, Universidad Francisco de Vitoria, Ctra. M-515 Pozuelo-Majadahonda km 1800, 28223 Pozuelo de Alarcón, Spain
alberto.nogales@ceiec.es, a.gtejedor@ceiec.es, juanjose.serrano@ceiec.es

[2] Faculty of Medicine, Research Institute, Universidad Francisco de Vitoria, Ctra. M-515 Pozuelo-Majadahonda km 1800, 28223 Pozuelo de Alarcón, Spain
d.monge@ufv.es, c.anton@ufv.es

Abstract. Artificial intelligence is a broad field that comprises a wide range of techniques, where deep learning is presently the one with the most impact. Moreover, the medical field is an area where data both complex and massive and the importance of the decisions made by doctors make it one of the fields in which deep learning techniques can have the greatest impact. A systematic review following the Cochrane recommendations with a multidisciplinary team comprised of physicians, research methodologists and computer scientists has been conducted. This survey aims to identify the main therapeutic areas and the deep learning models used for diagnosis and treatment tasks. The most relevant databases included were MedLine, Embase, Cochrane Central, Astrophysics Data System, Europe PubMed Central, Web of Science and Science Direct. An inclusion and exclusion criteria were defined and applied in the first and second peer review screening. A set of quality criteria was developed to select the papers obtained after the second screening. Finally, 126 studies from the initial 3493 papers were selected and 64 were described. Results show that the number of publications on deep learning in medicine is increasing every year. Also, convolutional neural networks are the most widely used models and the most developed area is oncology where they are used mainly for image analysis.

Keywords: Survey, Artificial Intelligence, Deep Learning, Medicine.

1 Introduction

The incorporation of information and communications technologies has led to an exponential increase in data generation in all areas of society. Only the use of sensors has generated an estimated 500 zettabytes of data in 2019 [1]. The field of healthcare has not remained outside this increase in information that is widely available both within and outside of public health institutions (social media, mobile devices, e-health apps, etc.). Healthcare-related data can have very different types and, hence, provide extremely diverse information: sociodemographic, clinical, genetic, related to treatments and their results, economic, administrative and about the preferences of patients and medical professionals [2].

Suitable integration and analysis of this enormous amount of data can help to create a medicine that is more efficient, personalized, participative, preventive, predictive and population-based. However, owing to a large number of variables and data, this analysis and its corresponding evaluation are impossible to conduct with conventional statistical tools. To do so, methodologies, techniques and tools that use artificial intelligence must be incorporated. This lets hidden patterns be determined and revealed, transforming them into knowledge to predict the future behavior of relevant variables and to identify others that were not previously taken into account to help make decisions at healthcare organizations and resolve highly-complex and real medical problems [3].

Artificial intelligence (AI) is the branch of computer science that analyzes and deciphers the mechanisms that generate intelligent behaviors in human beings, to then reproduce these behaviors in machines, not necessarily with the same mechanisms [4]. As a discipline, AI encompasses a large number of techniques, with different theoretical foundations and scopes of application. However, it is the field of machine learning (ML) that currently provides the most promising results. ML is a scientific discipline in the field of artificial intelligence that studies and develops algorithms to analyze data that let a system learn or in other words generalize behaviors by detecting patterns in the information supplied by way of examples and experience [5]. ML systems can make autonomous decisions based on predicting situations that may occur, although to do so they require large quantities of data, precisely the situation we find for the field of medicine [6].

The term ML encompasses several theoretical and practical approaches to the problem of making a computer system capable of extracting information from the data it analyses. One of these approaches are artificial neural networks (ANN). They are computational systems comprised of a set of simple processing elements (neurons) that are interconnected (network), whose behavior is determined by the topology and weights of the connections [7]. A more formal definition is the one given in [8], a computational system that consists of a large number of simple items, highly interconnected, which processes information by responding dynamically to external stimuli. ANNs learn from data in several ways: supervised, unsupervised or reinforced. But in all cases, they require a large amount of input data to learn and a careful training process to avoid overfitting that occurs when the model obtains good accuracies at training, but it fails to predict data not seen before [9]. Overfitting chances increase with the number of layers that compose the network, although the most interesting properties of neural networks are revealed in the deepest architectures. This impasse remained until deep neural networks, implementing deep learning algorithms, were proposed. Deep learning (DL) models are multiple-layer, hierarchical ANNs able to learn representations of data with increasing levels of abstraction starting from the input data [10]. These methods have dramatically improved state-of-the-art for speech recognition, image recognition, object detection and many other domains. Figure 1 shows the hierarchy in artificial intelligence and the different disciplines mentioned above.

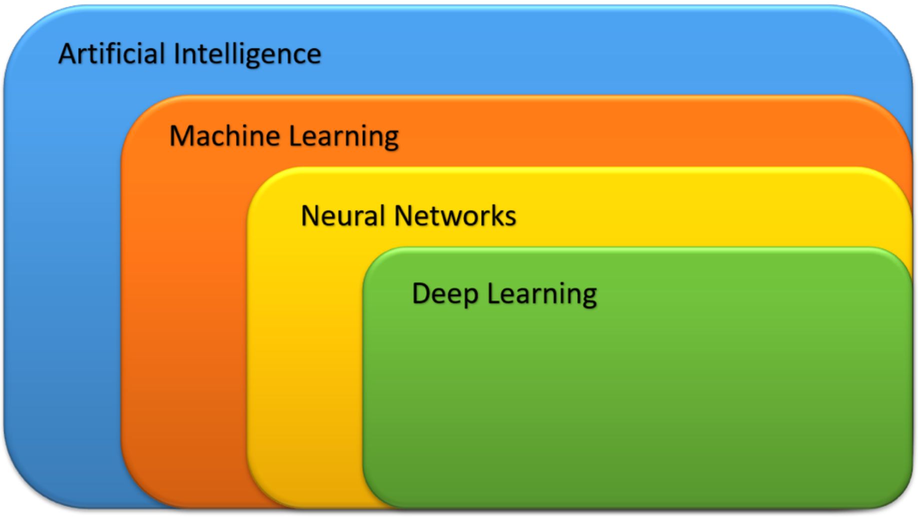

Fig. 1. Hierarchy between artificial intelligence disciplines.

In the medical domain, the areas where DL techniques have been most used are related to image diagnosis [11] and the analysis and classification of biomedical and clinical data, [12] and [13]. However, DL models have also been used to develop tools that help to segment the population, according to risk levels and adapt healthcare to each defined profile, letting patients' needs be anticipated. It has also been used for other purposes, such as developing public health, environmental and labor plans, including educational programs that can help prevent diseases; making predictions via disease probability and prognosis studies; evaluating quality management services and programs; optimizing teleservices and strengthening self-care and permitting decision making based on real data, [14] and [15].

This paper sets out a systematic review of the articles published in the medical field in which DL techniques have been applied. To do this, a methodology was first defined to semi-automatically obtain the relevant articles, eliminating those that were not pertinent to the scope of this study or whose impact on the scientific community was less. This methodology was based on a search for the best-known scientific sources, as well as applying important inclusive and exclusive quality criteria from the fields of medicine and computer science. After filtering the initial material, the contributions of the 126 selected articles were statistically analysed and 64 were described. The analysis revealed in which medical fields more studies have been carried out and which DL models are the most used. Although there are other reviews in medicine and Deep Learning like [16], [17] and [18] the aim of the present one should be a source of reference for physicians to know which use cases have been solved in their field. The potential for computer scientists could be finding under-exploited niches. For that purpose a deep statistical and graphical analysis is provided alongside a set of citations.

This article is organised in the following sections: Section 2 summarises the DL models used today and introduces the data types from the field of medicine. Section 3 details the methodology used to obtain the articles selected. Section 4 contains the in-depth study of the articles that were selected at the end, setting out their theoretical foundations, contributions and applications. Finally, Section 5 presents the conclusions obtained in this research study.

2 State of art

Starting in 2011, after the creation of AlexNet [19], the number of studies about deep learning published in medical bibliographic databases has progressively increased. Figure 2 shows the number of annual deep learning publications on PubMed, which practically doubled every year since 2015, except for 2017. And if we bear in mind the number of total publications in the 2000-2020 period, we can see that two-thirds of these are from 2019 and 2020.

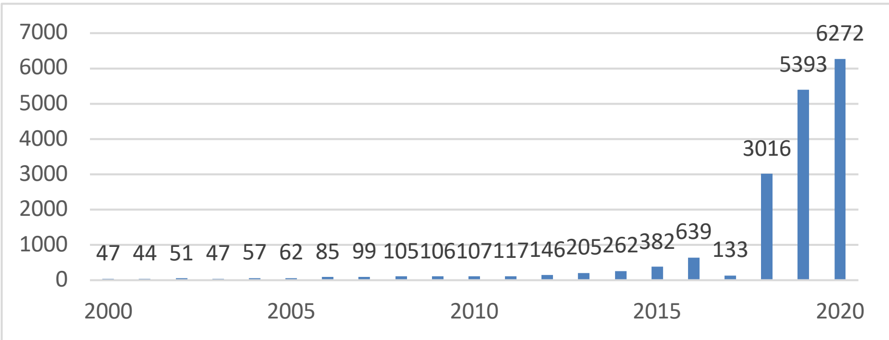

Fig. 2. Distribution by the publication year of the deep learning articles indexed in PubMed from 2000 to 2020 (n=2817).

DL techniques are based on multiple models and architectures, although the effectiveness of all of them is directly related to the nature and quality of the data with which they will be trained. This section describes the data types that are commonly used in medicine, as well as the architectures and models that best adapt to them.

2.1 Data types in medicine

In the medical field, the data types found may be structured, images, texts, time series, Electronic Health Records (EHRs) and graphs.

Structured data is defined [20] as: "any set of data values conforming to a common schema or type", basically data arranged in tables, such as databases or CSV or Excel files. They follow a row and column structure, the latter with a header. Columns define the characteristics of the individuals and rows, the values taken by the individuals for the characteristics in question. Images are obtained from medical tests like x-rays, scans and retinal fundus images. Texts include all written information used to monitor patients, such as their medical records and reports. Time series are electrocardiograms (ECGs) or electroencephalography (EEGs). Here, the information is a set of repeated observations of a single unit or individual at regular intervals over a large number of observations [21]. EHRs are a specific data structure in the medical field, which includes full patient information in diverse formats, including images or text. Finally, graphs can be a special way of modelling medical information, for example, the connections (edges) between different brain zones (nodes). In conclusion, depending on the nature of the data, one DL model or another will be most effective, as detailed in the following classification.

2.2 Deep learning models

The main classification of DL models is based on which learning method is implemented and how training data is used. Under this criterion, there are three different learning methods: supervised, unsupervised and semi-supervised.

In supervised learning, the neural network learns from labeled training data, so the network knows a priori the expected outputs for the input dataset. Three different models that follow this learning type are defined below.

Multilayer perceptron (MLP): This is the simplest DL model. It consists of a feed-forward supervised neural network with an input layer, an output layer and an arbitrary number of hidden layers. They perform well with simple datasets like structured ones and they are normally used to predict the probability that a given event occurs or the value of a particular parameter.

Convolutional neural networks (CNNs): These networks are one of the most widely used deep learning architectures today. Most CNNs are used to classify images and videos. Figure 3 shows the typical structure of a CNN used for image classification.

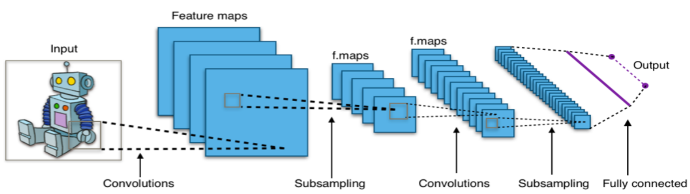

Fig. 3 Structure by layers and functioning of a CNN.[1]

Due to their structure and operating method, they can identify specific characteristics (for example, a tumor) in a delocalized way, meaning independently of its position on the image. The different capacities of these networks can be controlled by varying their depth and breadth. They also make strong and mostly correct assumptions about the nature of images (namely, stationarity of statistics and locality of pixel dependencies) [22]. As CNN is the most developed architecture in Deep Learning, we can find modification as 3D-CNNs or Graph CNNs [23], [24]. Therefore, these models are being used as an aid to medical diagnosis in fields like radiology for tasks like lesions classification, image segmentation or detection of the abnormalities in the medical tests.

Recurrent neural networks (RNNs): They are defined in [25] as a network that can process a sequence of arbitrary length by recursively applying a transition function to

[1] https://en.wikipedia.org/wiki/Convolutional_neural_network

its internal hidden state vector h_t obtained from the input sequence. The use of RNNs has become widespread, primarily due to their great utility for processing data whose type is a time series. The main feature of RNNs is that the output of all or some of their neurons is in turn connected to the inputs of neurons in the same or a previous layer, letting the network gain knowledge of the previous state (memory), meaning they become equipped with a sort of time meaning. Figure 4 shows an example of RNN where the neurons are interconnected. As the main capability that differentiates these models from other is saving previous states, they have mainly been used in medical tests whose information can only be understood by analyzing temporal values like biomedical signals. So, applications of RNN can be found in the area of cardiology or neurology where tests like electrocardiogram or electroencephalogram are used [26], [27].

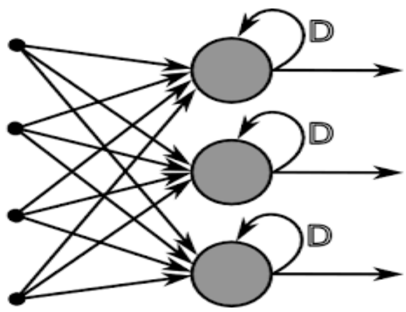

Fig. 4 Example of neural connections within an RNN.[2]

There are several types of recurring networks, the most widely used being Long Short-Term Memory networks (LSTM). LSTM networks arose due to the problem referred to as long-term dependencies. According to [28], LSTMs can learn to bridge time intervals over 1000 steps even when there are noisy, incompressible input sequences, without loss of short time lag capabilities.

The second learning type (unsupervised learning) uses non-labelled input data, that is, there is no a priori knowledge and the results to be obtained from processing the input data are unknown [29]. These neural networks can learn to organize information without providing an error calculation to evaluate the possible solution.

Deep autoencoders (AUD) are included within this group. This model, defined in [30], is a special type of feedforward neural network where both the input and the output are the same and is composed of two chained blocks. The first one, the decoder, reduces the size of the input data until the features that univocally characterised the input data are condensed into a small piece of data (the code). The second one, the decoder, upsamples that piece of data until the input data is reconstructed. Figure 5 shows the main feature of the autoencoder: the input and output layers are both the same size the output should replicate the input, while the hidden layers are smaller sized, as the input patterns

[2] https://missinglink.ai/guides/neural-network-concepts/recurrent-neural-network-glossary-uses-types-basic-structure/

are progressively coded and decoded throughout the process.Their capability to extract the fundamental features of the input has caused them to be used mainly to reduce data dimensionality, but also to reduce noise in input data (such as images). They are often used for data (image and signal) reconstruction, denoising or augmentation [31], [32], [33]. These tasks can be considered to belong to the computer science field mostly but are useful in medicine. Applications in the medical field include segmentation, detection and classification in images that are difficult to manage due to its size or that need to be improved in terms of resolution, [34], [35], [36].

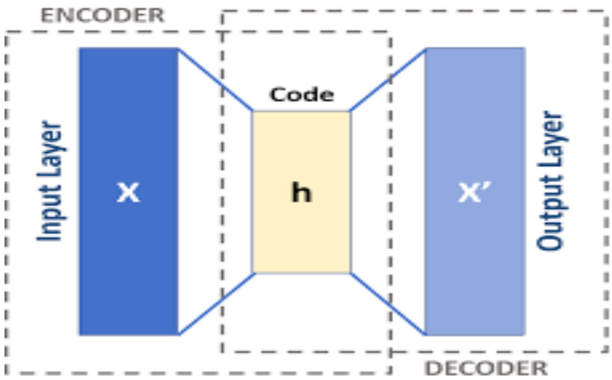

Fig. 5 Example of autoencoder architecture.[3]

In addition to the two learning types described above, there are also architectures implemented through mixed learning types (supervised and unsupervised) called semi-supervised. Generative adversarial networks (GAN) would fit into this class.

GAN is an architecture composed of two neural networks, a generator and a discriminator or classifier, that compete between them in an adversarial training process [37]. The set, as a whole, can learn to imitate any data distribution. The generative network will be in charge of generating instances that belong to the data distribution (a specific data structure, such as images) realistic enough to deceive the second, whose job is to discern between real and generated data structures. The discriminator estimates the probability of this generated data to belong to the data distribution (authentic) or not (fake). As the discriminator classifies the generated data as fake, the generator learns to generate instances closer to the data distribution. By following this process, both models improve the way they perform. In cases of scarce data, GAN can be used to generate synthetic instances of different classes. They are also applied in data reconstruction like signal denoising or image reconstruction. For example, cleaning up artifacts in electroencephalographic tests [38]. They have also been used in dataset manipulations like image superresolution (ontaining more detailed radiographs) and segmentation (resonance images where different elements are tagged) or creation of new synthetic instances (in cases where the training dataset is not enough) [39], [40], [41].

[3] https://en.wikipedia.org/wiki/Autoencoder

3 Materials and methods

3.1 Criteria for selecting articles

To fulfill the objectives that are set out above, the multidisciplinary team designated for the project defined the following criteria.

Papers that describe the development of deep learning models in medicine were included, excluding those focused on the fields of biotechnology, biology and studies conducted with animals.

As no validated tool exists to evaluate the quality of studies describing the development of artificial intelligence models in medicine, we drew up the list of requirements that the papers had to meet, with the help of experts on this topic. The requirements are:

1. The implementation of the model is published in a peer-reviewed journal included in the Q1 impact index (considering the time the papers were accessed) of the Journal Citation Reports (JCR).
2. The paper includes a detailed description of the development of the model so that it can be replicated.
3. The initial dataset must be distributed at a percentage close to 80-20% between training and validation data. This is a well-known good practice in ML related to the Pareto principle [42].
4. Information is included in the model's error or accuracy and evaluation against a baseline.
5. The sample (dataset) must be representative of the study population, both qualitatively and quantitatively.
6. If dataset replication is included, the process must be adequately explained.

3.2 Search strategy for identifying the studies

To define the search strategy, we used the Medical Subject Headings (MeSH) terms, a terminological vocabulary for science articles. In our case, MeSH headings were deep learning and medicine. A complete list of terms under these headings can be found in Appendix A. We also added open terms from the medical and computational sciences fields that were not mutually exclusive. The terms used in medicine were: clinical decision making, image analysis, image processing, medicine, health care and health. The terms in computational sciences were: machine learning, artificial intelligence, bioinformatics, feature learning, feature representation, supervised learning, unsupervised learning, neural networks, deep neural networks, convolutional neural networks, convolution, deep autoencoders, autoencoder, deep belief networks, generative adversarial networks, recurrent neural networks and LSTM. The search and title and abstract extraction period of the papers were at maximum until 15th September of 2020.

3.3 Sources used to extract the studies

The databases we consulted were: Scopus, EMBASE, MEDLINE, CINAHL, PsyArticles and the Astrophysics Data System. They were accessed using the following search engines: Science Direct, PubMed, Europe PubMed Central, Web of Science (WOS) and EBSCO Discovery.

3.4 Data extraction, classification of studies and analysis

To unify all the final files from the search (XML, CSV, etc.) in the aforementioned databases, we wrote several scripts using different Python libraries: Pandas,[4] which allow for easy handling of data structures, NumPy[5] for vector and matrix structures and API ElementTree XML,[6] whose purpose is to manage XML files. The final result was an Excel file that contained all the papers with their titles, abstracts, publication years and the journal in which each one was published.

The selection of studies was done with twofold screenings: the first by title and abstract by peer review, with a third referee if there is no agreement on whether or not it meets the criterion of "a study developing a deep learning model in medicine". The full texts of the selected papers were then obtained and, after reading them, the studies that did not meet the inclusion criteria were discarded. This second screening was done in two phases: first by applying the filtering criterion by the JCR Q1 quartile and then the rest of the criteria.

Data extraction and journal classification were done by creating an Excel file with the journal name, impact factor, quartile, category, H index and the total number of citations. This information was obtained from the WOS and JCR.

The studies finally selected were also classified according to different criteria both in the field of computer science and medicine. In the first case, the factors taken into account were the nature of the data worked with (structured data, images, time series, etc.) and the deep learning model applied (CNN, RNN, MLP, etc.). Criteria for the medical field were the therapeutic area of study (neurology, cardiology, oncology, ophthalmology, etc.), the medical segment in which the results could be applied (diagnosis, classification, surgery, monitoring of treatment or predicting prognoses for diseases), the tests and technologies analysed, whether or not results from the model were verified with external databases not included in the initial dataset with which it was developed and validated, and if the applications resulting from the study were described for common clinical practice.

3.5 Statistical analysis

Finally, and taking into account the classification of the selected papers, a statistical description of the works was obtained. For this, several scripts were developed in Python, fed by the relevant Excel files where the information was saved, and whose output was processed with Matplotlib[7] graph gallery to create the different charts.

The graphs reflect the following criteria: publication year of the articles in the final selection, countries of origin of the centers to which the authors belong; nature of the

---

[4] https://pandas.pydata.org/

[5] https://numpy.org/

[6] https://docs.python.org/2/library/xml.etree.elementtree.html

data used in the selected publications; nature of the data used in the study; use of replicator or booster (obtain several data sets from original data set by resampling on sample space) concerning the initial dataset; the DL models implemented in these publications; comparison between the model used and the data type; therapeutic area of the field of medicine in which the research will be applied; the purpose of the model in medicine; whether or not there was validation with external databases; and description of how the development would be applied to clinical practice.

4 Results

4.1 PRISMA flow diagram

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram [43] below in Figure 6 summarises the results obtained in the search and subsequent screening phases until obtaining the final selection of the articles reviewed.

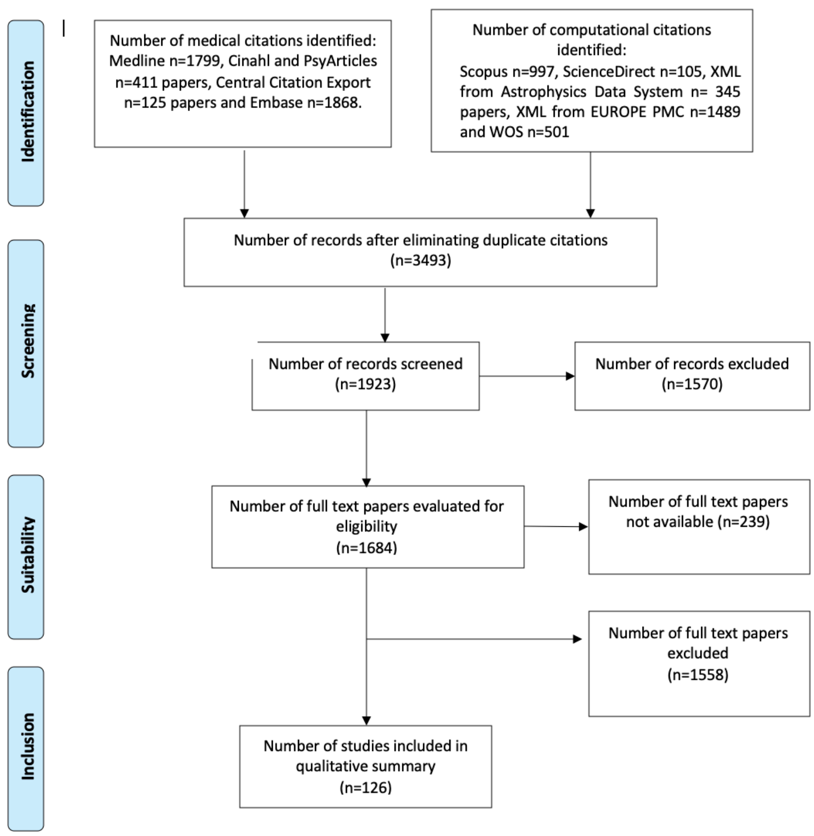

Fig. 6 PRISMA diagram of the bibliographic review conducted. [43]

A total of 7640 papers were obtained from the consultations made (Scopus 997, ScienceDirect 105, XML from Astrophysics Data System 345, XML from EUROPE PMC 1489, WOS 501, Medline 1799, Cinahl and PsyArticles 411, Central Citation Export 125, Embase 1868) where, after eliminating duplicates, 3493 papers remained that we used to start the first screening by title and abstract. In this phase 1570 papers were ruled out, due to not meeting the criterion of “being a deep learning study developed on medicine,” and 1923 were passed on to the second screening. Of the 1923 papers in the second screening, we could not obtain the full texts of 239, so 1684 were analysed, of which 126 studies met the inclusion criteria for review.

Of the articles discarded in the second screening, 647 were excluded because the journal in which they were published had an impact index lower than Q1 at the moment the search was done.

And of the 911 remaining, 516 were ruled out due to not meeting these criteria: division of data for training and validation, description of the model to be replicated, and comparison with other baseline models. A further 177 were discarded due to using a non-representative sample of the study population, 32 for not specifying the expansion model for the initial dataset, 186 papers due to being scientific specific areas, 121 biotechnology and 65 medical engineering.

At the end of our selection process, the number of papers considered most relevant was 126. Table 1 summarises the main causes for the exclusion of papers in the second screening.

Table 1. Reasons for exclusion of articles in the second screening.

Reason for Exclusion	Number of papers
JCR Q1 quartile	647
Replicable model	516
Sample (dataset) representative of the study population, both qualitatively and quantitatively	177
If dataset replication is included, the process must be adequately explained	32
Centered on biotechnology or medical engineering field	186

The distribution by the publication year of the 7640 papers obtained corroborated the rising trend, especially since 2014, as can be seen in Figure 7. After that year, the number of publications from one year to another is observed to double except in the last two years. This could be a consequence of an increase in the evaluation criteria due to the large number of people working on Deep Learning. It is noteworthy mentioning that in the graph, the number of 2020 publications is only through September.

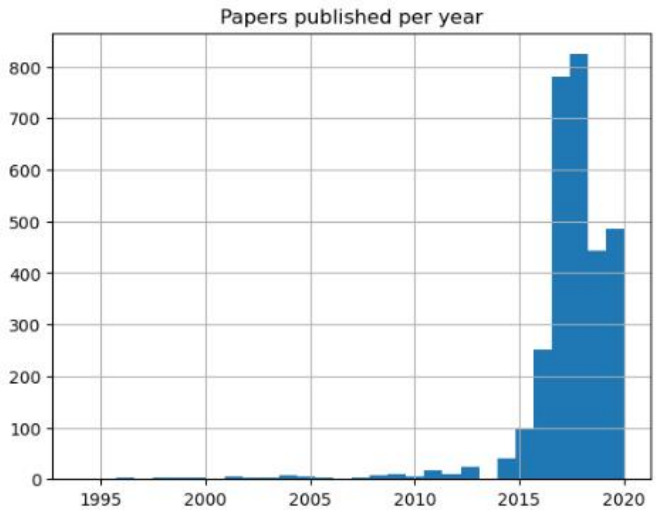

Fig. 7. Distribution by the publication year of the papers obtained without duplicates (n=3493)

4.2 Paper summarising

Due to a space problem, Table 2 only compiles 64 papers (the rest have been referenced in Appendix B) that we considered the most relevant from the 126 that have been chosen after the process. It has a reference to the paper, the therapeutic area where the model has been applied, which is the main aim of the research, what kind of Deep Learning model has been used, which type of data, how the dataset was formed and the results of the models in terms of accuracy or lost.

Table 2. Summary of the main papers reviewed.

rapeutic area	Objective	Model design	Type of data	Sample	Re
iovascular	Predict heart failure	RNN with GTUs	EHRs/Time series	3884 Patients with heart failure and 28,903 controls	AU
iovascular	Segment left ventricle images with greater precision	Deep belief networks	Ultrasound image of the heart 2D	400 images with five different heart diseases and 80 normal echocardiogram images	Ha 0.8
matology	Diagnose possible soft tissue injuries	DeepResolve, a 3D-CNN model.	Nuclear MRIs 3D	124 double echo steady state from 17 patients	MS
logy	Study of tumor tissue samples. Localize areas of necrosis and lymphocyte infiltration	Two CNNs	Pathology cancer images (hematoxylin and eosin)	5,202 images tumor-infiltrating lymphocytes	AU
halmology	Retina age-related macular degeneration diagnostic	CNN	Retinal 3D images obtained by Optical Coherence Tomography	269 patients with AMD, 115 control patients	AU
halmology	Diagnose retinal lesions	CNN	2D Ocular fundus images	243 retina images	Pre 0.8 and
ology-Psychiatry	Automatic interpretation system in Parkinson's disease	CNN	123I-fluoropropyl carbomethoxyiodophenyl nortropane single-photon emission computed tomography (FP-CIT SPECT) 2D images	431 patient cases	Ac
tious Disease	Create a screening system for Malaria	CNN	Giemsa-stained thin blood smear slides cell images	27,558 cell images 150 infected and 50 healthy patients	Ac
ology-Psychiatry	Decide Acute Ischemic Stroke patients' treatment through volume lesions prediction	CNN	Diffusion-weighted imaging maps using MRI	222 patients. 187 treated with rtPA (recombinant tissue-type plasminogen activator)	AU

thetist	Adapt anesthetics treatment dose for different patient profiles	LSTM	Data registry	231 patients with basic information and vital signs data	Cor coe
ology	CAD system to classify tomographies and evaluate the malignity degree in gastro-intestinal stromal tumors (GISTs)	Hybrid system between convolutional networks and radiomics	Abdominal CT 3D images	231 computed abdominal tomographies	AU
ology-Psychiatry	Schizophrenia detection	Deep discriminant autoencoder network	Magnetic resonance images	474 patients with schizophrenia and 607 healthy subjects	Acc
iovascular	Diagnose, stratification and treatment planning for patients with aortic valve pathologies	Marginal space deep learning	Transesophageal ultrasound volume and 3D geometry of the aortic valve images	3,795 volumes from the aortic valves from 150 patients	Pos and erro
mology	CAD system to diagnose interstitial lung disease	CNN	CT image patches 2D	14,696 images from 120 patients with proven diagnose	Acc
roenterology	Staging liver fibrosis through MR	CNN	Gadoxetic acid–enhanced 2D MRI	144,180 images from 634 patients	AU and
halmology	Diabetic retinopathy detection and stage classification	Bayesian CNN	Ocular fundus images 2D	Over 85,000 images	AU
ology	Detect malign solid lesions and prevent overtreatment in false positives	CNN	Mammography images	45,000 images	AU
iovascular	Monitoring cerebral arterial perfusion via spin labeling	CNN	Arterial spin labeling (ASL) perfusion images	140 subjects	AU
ology-Psychiatry	Identify different autism spectrum disorders	Denoising AE	Resting state functional magnetic resonance imaging (rs-fMRI), T1 structural cerebral images and phenotypic information	505 individuals with autism and 520 matched typical controls	Acc
ology-Psychiatry	CAD for early Alzheimer disease stages	Multimodal DBM	3D MRI and PET	93 Alzheimer Disease, 204 MCI Mild Cognitive	Acc 0.7

				Impairment converters and normal control subjects	
halmology	Detect retinal hemorrhages	CNN	Color ocular fundus images	6,679 random sampling images from Kaggle's Diabetic Retinopathy Detection	AU
ology	Mammography diagnosis of early malignant breast cancer with microcalcifications	Stacked AE	Mammography	667 benign and 333 malignant	Acc
ology	CAD to discriminate benign cysts from malignant masses	CNN	Digital Mammography images and the biopsy result of the lesions	1,000 malignant masses and 600 cysts images and their biopsy	AU
halmology	System to detect and evaluate glaucoma	CNN: ResNet and U-Net	Ocular fundus images	168 images with glaucoma and 428 control	AU spe
ology	Dermoscopy CAD system for acral lentiginous melanoma diagnosis	CNN	Dermoscopy images	350 images of melanomas and 374 benign nevus	Acc
iovascular	Breast arterial calcification on mammograms classifier to evaluate the risk of coronary disease	CNN	Mammography images	840 images of mammograms from 210 different patients	Mis
roenterology	Detection and localization system of gastrointestinal anomalies via endoscopy	CNN	Frames from endoscopy videos	205 normal and 360 abnormal images	AU
natology	Recognize nails nychomycosis lesions	Region-based-CNN	Patient demographics and clinical images	49,567 images	AU AU of (tase
ology-Psychiatry	Predict the survival of patients with amyotrophic lateral sclerosis	CNN	Clinical characteristics and MRI 3D	135 patients with short-, medium- and long-term survival	Acc

halmology	Differentiate Age-Related Macular Degeneration lesions in optical coherence tomography	Modification of VGG16 CNN	Optical coherence tomography images	52,690 AMD patients' images and 48,312 control	AU and diff
iovascular	Obstructive coronary disease automatic prediction system	CNN	Stress 99mTc-sestamibi or tetrofosmin myocardial perfusion images	1,638 patients	Ser and
halmology	Predict the evolution of diabetic retinopathy with fundus images	CNN	Ocular fundus images	90,000 images with their diagnoses	AU
logy	CAD system to classify breast ultrasound lesions and lung CT nodules	Stacked denoising AE	Lung computed axial tomography 2D images and reast ultrasound lesions	520 breast sonograms from 520 patients (275 benign and 245 malignant lesions) and lung CT image data from 1,010 patients (700 malignant and 700 benign nodules)	AU
logy	CAD to prevent errors in diagnosing prostate cancer	CNN	MRI 2D	444 images from 195 patients with prostate cancer	AU
logy	Computer automated estimation of breast percentage density	CNN	Digital mammograms	661 from 444 patients	AU
iovascular	Determinate limits between the endocardium and epicardium of the left ventricle	RNN with automatic segmentation techniques	MRI 2D	MICCAI 2009 left ventricle segmentation challenge database	Acc cas
ous	Classify medical diagnostic images according to the modality they were produced and classify illustrations according to their production attributes	CNN and a synergic signal system	12 categories of medical diagnostic images, such as CT images, MRI images and PET images, and 18 categories of illustrations	6,776 images for training and 4,166 for tests	Acc

ology	Classification of breast cancer histology microscopy images	CNN with a Support Vector Machine (SVM)	Microscopy image patches	249 images belonging to 20 histologic categories	Acc clas acc nor sifi
ology	CAD for breast cancer histopathological diagnosis	CNN	Microscopy histopathological images	7,909 images of eight subclasses of breast cancers	Acc
ology-Psychiatry	Analyze cerebral cognitive functions	3D CNN, resting state networks	Functional MRI	68 subjects perform 7 activities, and a state of rest	Acc
matology	CAD for diagnosis of knee osteoarthritis	Deep Siamese CNN	Radiography images	7,821 subjects with 6 monitoring phases	Acc
ology	Segment areas of dense fibroglandular tissue in the breast	CNN	Mammography images	Mammograms from 604 women	Acc
roenterology	Screening system for undiagnosed hepatic magnetic resonance images	CNN	Liver MRIs	522 liver MRI cases with and without contrast for known or suspected liver cirrhosis or focal liver lesion	Rec tive gre
ology	Discriminate lung cancer lesions in adenocarcinoma, squamous and small cell carcinoma	CNN	CT image 2D	63,890 patients with cancer and 171,345 healthy	Log a sc
ology	CAD system to detect and differentiate breast lesions with ultrasound	CNNs inspired in AlexNet, U-Net and LeNet	Ultrasound imaging	306 malignant and 136 benign tumors images	Bes 0.8
tic Surgery	Detect the two-dimensional position of different medical instruments in endoscopy and microscopy surgery	Convolutional detection-regression network	Single-instrument Retinal Microsurgery Instrument Tracking dataset, multi-instrument Endo-Visceral surgery and multi-instrument in vivo images	940 frames of the training data (4,479 frames) and 910 frames for the test data (4,495 frames)	Acc

ology	Probability of cancer on mammograms	CNN	Digital mammograms images	29,107 left mediolateral oblique, right mediolateral oblique, left cranial-caudal and right cranial-caudal mammograms images	AU
ology	Cervix cancer screening	Multiscale CNN	Microscope images	200 female subjects aged from 22 to 64	Me tion
ous	Speed up CT images collection and rebuild the data	DenseNet and a deconvolution model	CT 2D images	3,059 images from several parts of human body	RM
matology	Radiographies CAD for hip osteoarthritis diagnosis	CNN	Radiography images	420 radiography images (219 control group, 201 ostearthritis)	Ac
ology	CAD to diagnose lung cancer in low-dosage computed tomography	Eyetracking sparse attentional model and convolutional neural network	CT images 3D	6,960 lung nodule regions, 3480 of which were positive samples and rest were negative samples (non-nodule)	Ac
ous	Processing text from CT reports in order to classify their respective images	CNN	CT images 2D and text (reports)	9,000 training and 1,000 testing images	Ac 0.5 thre
ology-Psychiatry	Device that lets people with amyotrophic lateral sclerosis write	CNN	P300 signals from electroencephalography	38,750 P300 and 66,450 non-P300 samples	Ac
halmology	CAD to diagnose rhegmatogenous retinal detachment	CNN	Ocular fundus images	411 patients with the disease and 420 controls	F-1 0.5 on

ology	Whole-slide histopathology images to outline the malignant regions	CNN	Whole-slide prostate histopathology images	2,663 images from 32 whole-slide prostate histopathology images	Dic
ology	Binary classification of posteroanterior chest x-ray	CNN	Computed tomography (CT)	Three datasets: 224,316, 112,120 and 15,783	92%
ology	Automatically evaluate the quality of multicenter structural brain MRI images	CNN	MRI images	1064 brain images of autism patients and healthy controls. MRI data from 110 multiple sclerosis patient	AU
halmology	Image quality in the context of diabetic retinopathy	CNN	Fundus images	7000 colour fundus images	Acc
hinolaryngology	Automated Plysomnography scoring	CNN+LSTM	Electroencephalography, electrooculography, and electromyography data	42,560 hours of PSG data from 5213 patients	F1-
ocrinology	Automatic diagnosis and severity-classification model for acromegaly	CNN	Facial photographs	2148 photographs at different severity levels	90.
halmology	Diagnosis of Age-related Macular Degeneration	CNN	AREDS (age related eye disease study) image	130,000 fundus images	94. 98.
halmology	Predict age and sex from retinal fundus images	CNN	Fundus images	219,302 from normal participants without hypertension, diabetes mellitus (DM), and any smoking history	AU

ology	Abnormality detection in chest radiographs	CNN	Radiographs	112,120 frontal view chest radiographs from 30,805 patients and 17,202 frontal view chest radiographs with a binary class label for normal vs abnormal	AU 0.9 and
ology	Classify white blood cells	CNN	Leukocyte images	5,000 images from local hospital	94. 95. 94.
iology	Predict Stroke Patient Mortality	MLP	11 variables	15,099 stroke patients with primaryInternational Classification of Diseases diagnostic codes	AU
eral	Predict patients' hospital mortality	RNN	Public electronic health record database. Fifteen physiological measurements.	32604 unique ICU admissions	Ser

D age-related Macular Degeneration, CAD Computer Aided Diagnosis, CNN Convolutional Neural Network, MRI Magnetic Resonance Images, P
Computed Tomography, OCT Optical Coherence Tomography, D dimensions, AUC Area Under the Curve, MSE Mean Squared Error, RMSE Root M
Coefficient

4.3 Statistics and analysis of the studies included

At the end of the screening process, we had obtained 126 papers. At this point, we verified the rising trend of journals with deep learning papers for medicine. Figure 8 is a bar chart showing the distribution of the 126 papers by year of publication, where one can observe the increasing trend in the number of publications in recent years. From 2016 to 2018, this number more than tripled. This fits with the historical process because although the term Deep Learning was coined by Hinton with his seminal work [10], in 2006, the big milestone is considered to be AlexNet for image recognition in 2012, [19]. The smaller number of papers that were published in the last two years corroborates what has been shown in Fig. 7.

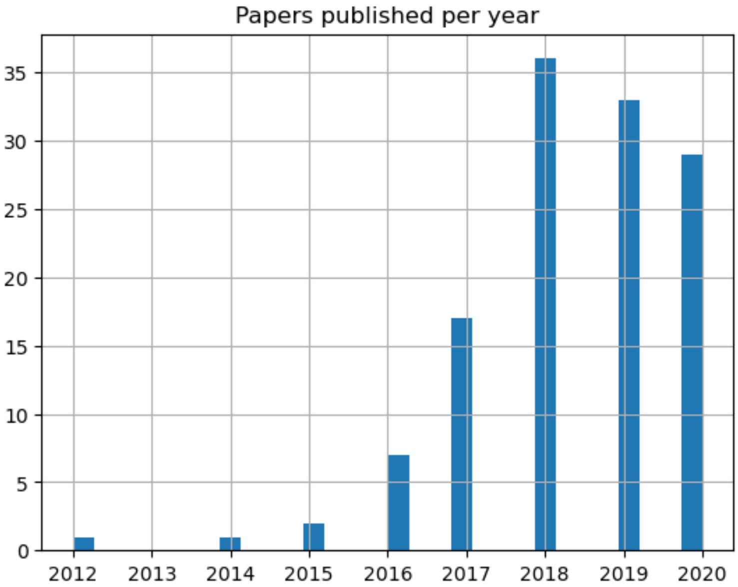

Fig. 8. Distribution by year of the articles selected (n=126).

Another interesting piece of information is the number of papers published by each country. To do so, we collected the affiliation of each first author and compiled the bar chart in Figure 9. As can be seen, the most prominent country is the United States, followed by China in a distant second. The second group of countries includes Korea and the Netherlands. The rest of the countries are only provided from one to three papers. This information fully coincides with that provided by Nature in absolute numbers[8] in terms of research output.

[8] https://www.natureindex.com/annual-tables/2019/country/all

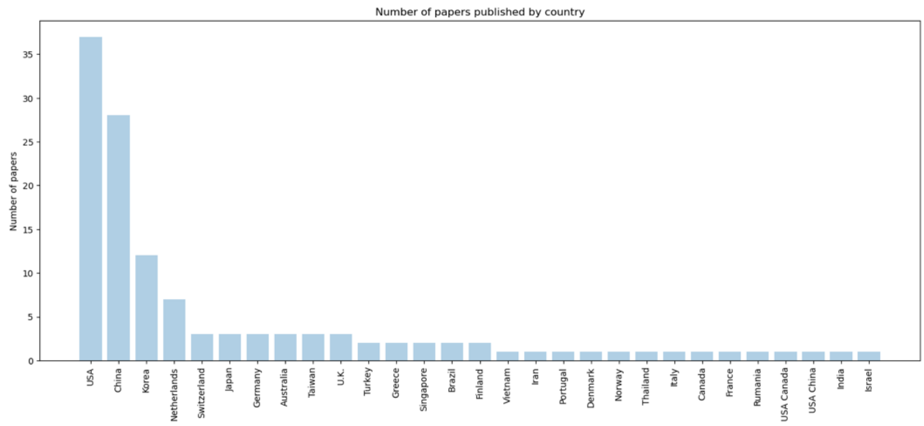

Fig. 9. Distribution by country of the articles selected (n=126).

One of the largest determinants when deciding which deep learning models to use is the nature of the data they will be working with. Figure 9 is a pie chart depicting the distribution of the data types showing that the majority of models —90.5% (114 papers)— work with images and only a small percentage —4.8% (6)— work with time series for example [44, 96, 102] or with structured data —3.2% (4)— like [56, 101, 108]. Only one paper, [95], works with two data types, in which radiographs were used along with their medical descriptions and only another that uses graphs. This information is supported by data published by the National Institute of Health (NIH) where funding in cancer is in the top positions. Considering that most tests in cancer diagnosis are related to medical imaging, there should be a wide range of this type of data. For example, imaging test in the US has greatly increased in recent years, [110].

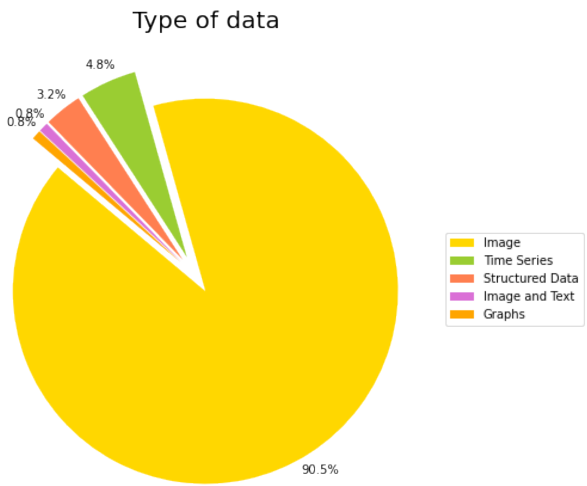

Fig. 9. Distribution of the models by data type (n=126).

It is worth mentioning that 44.8% (60) models used boosters to expand the sample size that they started with initially [45-53, 57, 62, 79, 81, 87-92, 94-98, 102, 107] Figure 10.

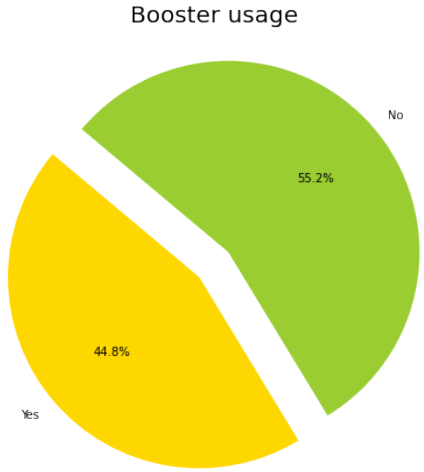

Fig. 10. Distribution of models by whether or not they used boosters to expand the sample size (n=126).

The data type used is directly related to the type of model developed. As can be seen in the pie chart in Figure 11, the most used network, at a large margin, is CNN, in 75.4% of the cases (95 papers). This makes sense, as the CNNs normally use images for working. Then there is a set of papers that uses Autoencoders, 15.6% (20). The rest of the architectures are primarily used at the same percentage: MLP and RNN, 4.1% (5 times each) and GAN, 0,8% (1). [55, 62, 65, 76] uses Autoencoders in the area of neurology for the particular cases of schizophrenia and autism or lung and breast cancer. In the case of MLP a particular marginal space Deep Learning model for diagnosis, stratification and treatment planning for patients that have an aortic valve implanted. Related to RNN [44, 53, 109], vital signs are used alongside some patient's information. Finally, GAN is only used once. Thus, one can speculate with the hypothesis that, in a near future, CNN models will be part of the diagnosis system. Also, a front is opened in the study of Autoencoders mostly used for image segmentation. In the case of RNNs, the difficulty of obtaining this type of medical data is an obstacle to its evolution.

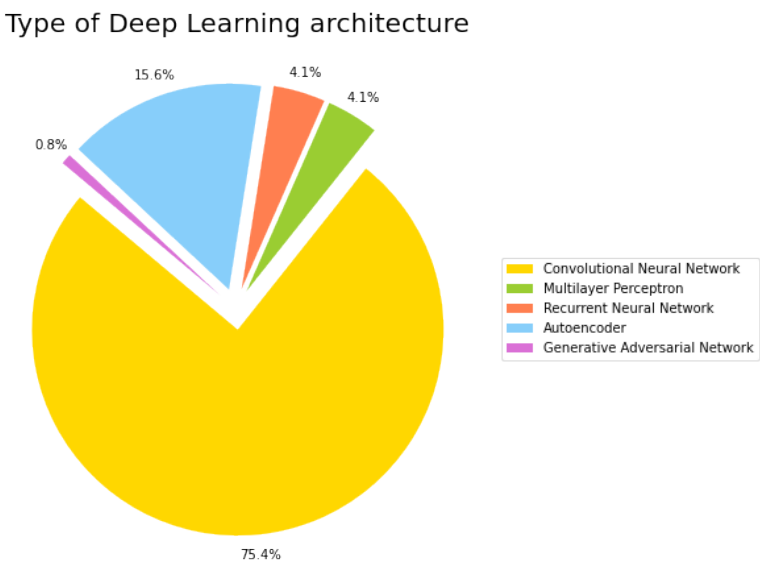

Fig. 11. Distribution of the models by deep learning architecture type (n=126).

To support the conclusion obtained in the last step, Figure 12 shows a bubble diagram where the X-axis represents the model types and the Y-axis the data types. The largest bubble, 69.84% (88 papers), represents the CNN models developed from images. Then, there are 11.89% (15 papers) using Autoencoders with images. The use of CNN with time series, MLP with images, MLP with structured data, RNN with structured data and RNN with time series have two cases each. There is only one paper where CNN is used with text and images [95] or graphs, MLP with time series and images with RNN or GAN. Other cases did not arise in this survey. These results support what has been concluded in the previous paragraph. It is also remarkable that the use of different data sources as the text seems extra information to guide training stages in Deep Learning models. It should also be highlighted that Natural Language Processing is nowadays a hot topic in Deep Learning with the greatest improvements.

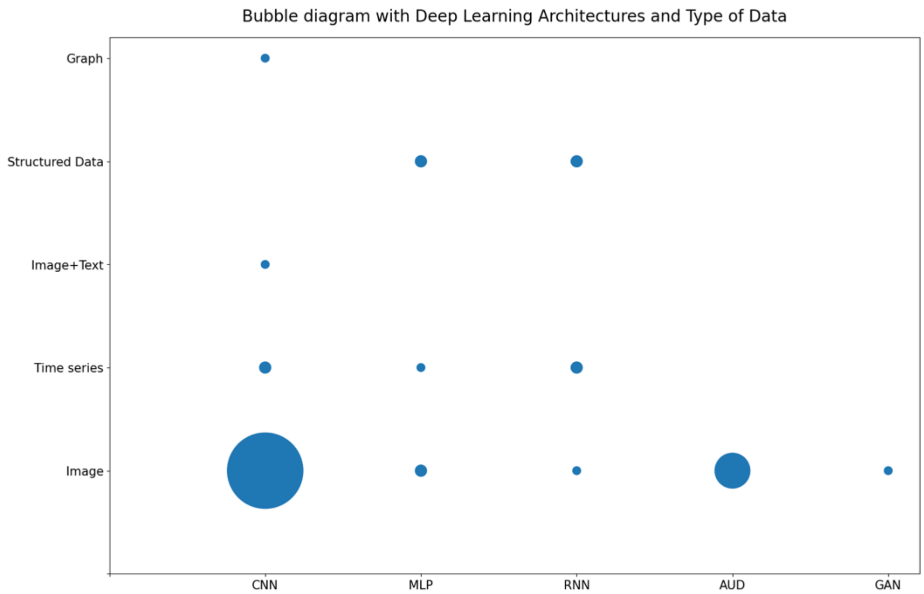

Fig. 12. Representation of the relationship between the data type and the model architecture used (n=126).

The therapeutic areas in which most papers are published are oncology (32.5%, 41), followed by cardiology (11.9%, 15) and neuropsychiatry with 11.1% (14), Figure 13. Standing out in oncology is the Computer-aided diagnosis (CAD) to help physicians to classify models of disease (histology) and facilitate image diagnosis of the tumors which includes mammography, Computer Axial Tomography (CT) and magnetic resonance. The development in Deep Learning for medical imaging can be seen in the wide range of areas where it has been applied: ophthalmology, pneumology or dermatology. There are also curious researches in diseases like malaria, a very common infectious disease in developing or third world countries.

As we can see in Table 2, breast cancer screening and diagnosis support is one of the main objectives [60, 65, 66, 77, 81, 82, 85, 88, 90], followed by the development of CAD in lung cancer [76, 87, 94].

In cardiology, the majority of the papers are about support to diagnosis using images from different tools like ultrasound [45, 56], magnetic resonance [79] and myocardial or cerebral arterial perfusion [61, 74]. [44] uses electronic health records to predict possible future heart failure onset via a time series.

In neuropsychiatry, the aim of many studies is the diagnosis [50, 55, 62, 63, 83, 100] but also, we can find studies to predict disease evolution [72, 108], to allow patients to write through their eyes movement [96].

Ophthalmology is another therapeutic area where deep learning has developed many models with ocular fundus images to detect retinal diseases like age-related macular degeneration [48, 63], hemorrhages [64], microaneurysms [49], diabetic retinopathy [59, 77, 101] and rhegmatogenous retinal detachment [97] and glaucoma [50].

Other therapeutic areas found in this review are radio image 8.7%, ophthalmology 7.9% or traumatology 4.8% for more details see Fig. 13 and Table 2.

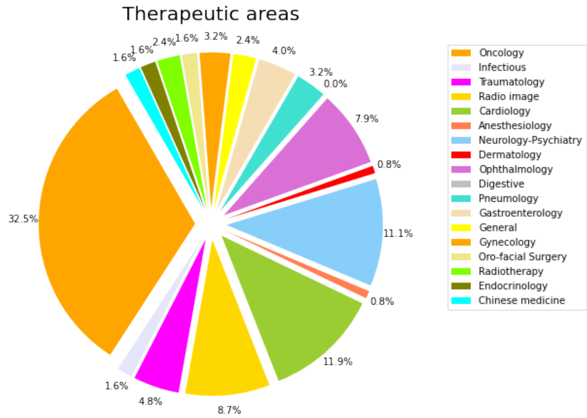

Fig. 13. Distribution by areas of applicability to medicine (n=126).

As seen, diagnosis-based models were the objective of 82.5% (104) of the studies, while 3.2% (4) centred on monitoring treatment, and the rest, 14.3% (18) have miscellaneous topics as their objectives, such as disease classification, robotic surgery and prediction of prognosis, Figure 14. As medicine is a field where the wrong decisions could have irreversible consequences, most of the work with DL methods are applied to diagnosis. The controversy is about letting machines decide, so their role nowadays is aiding physicians in taking better and faster decisions and this can be done mainly at the diagnostic stage.

Most of the models exposed in this survey have an accuracy of 80% or more. So, it can be concluded that the performance of the different models in different areas and use cases are quite good. Only in particular cases in the field of anesthesia using patients' biosignal, detecting autism with MRI or keen osteoarthritis from radiographies have an accuracy under 80%. Most of the papers used Accuracy and AUC as a metric to measure the performance of the models.

Metrics can be grouped taking into account some charactertiscs. Accuracy is the simplest one and uses the correct predictions, unlike error metrics as MSE or Log loss error. In the particular case of medicine, it is very useful to use the false positives and negatives: Sensitivity, Specificity and F-score or F-measure. This is related to what is highlighted in [111] about the economic impact and risk in diagnosing a healthy patient as sick. Related to this, there are graphical representations as AUC and Precision-Recall curves. When using images, Hammoude distance, Position error, Mean corner distance error and Dice coefficient are used. Finally, concordance correlation coefficient which is an agreement between two variables.

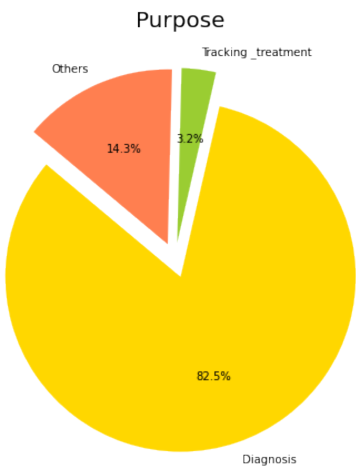

Fig. 14. Distribution according to the purpose of the model (n=126).

Only 25.4% (31) of the 126 papers were validated with databases different than the initial dataset, and only two studies (1.6%) detailed its application to clinical practice, Figure 15. These studies validated their results in different databases than the initial dataset were: [51], [57], [74], [82], [86], [92] , [94], [98-107], [94]. And only two [82, 101] describes the application of the model in the current clinical practice.

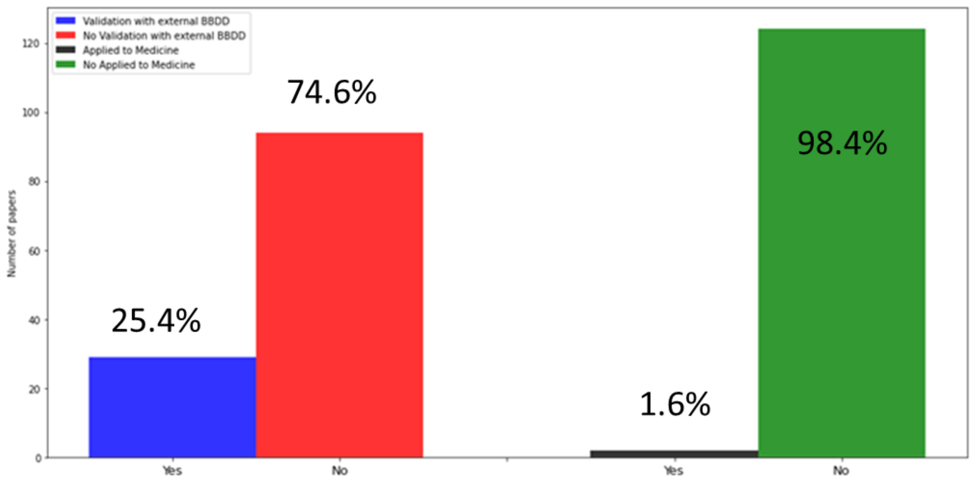

Fig. 15. Representation of the percentage of models validated with databases different than the initial dataset (n=126), and distribution of the articles depending on whether their application to common clinical practice is described (n=126).

5 Discussion and future works

The purpose of our study was to conduct a comprehensive review of the state-of-the-art in the application of deep learning techniques in the field of medicine. A methodology was defined for selecting a series of papers that could be considered representative. This methodology started with the search in different sources of scientific knowledge, obtaining 4505 initial papers. This number was progressively refined by eliminating duplicates and articles not in this field, as well as other exclusion criteria defined by computer scientists and physicians. At the end of the process, 126 papers were selected and briefly summarised and analysed from a qualitative and quantitative perspective. The most straightforward conclusion that can be drawn is that deep learning techniques are widespread in the oncology discipline. Given that here the most used data for diagnosis are images, and that convolutional models are directly related to the treatment of these images, it is logical that most deep learning applications found during the review use this type of architecture. The next relevant areas are cardiology, ophthalmology, and neuropsychiatry, where images also play a prominent role in diagnosis.

One of the main limitations of the study was the need to discard papers published in JCR impact quartiles below Q1. This was because the large volume of references to be included did not permit a correct description of all studies. This is why the objectives of this research team that conducted the study include writing a second paper that would complete and allow for the determination of whether or not the quality of the studies published differs depending on the quartile in which they are published. On the other hand, because this is a review in which various disciplines converge (computer science and all medical specialties) and despite the careful methodological process, there may be published studies for which we did not have access. We also found no information in the publications about the models used by companies such as Google, Intel, Microsoft, Philips and Siemens, probably due to the confidentiality of the data and the patents of the models.

However, it is worth noting that two types of neural applications are significantly absent or underrepresented in the results obtained from this study. The prediction and diagnosis of a patient's medical evolution, mortality risk, or the emergence of diseases through the analysis of discrete/continuous signals (historical vital signs, EEG/ECG data, etc.) have not been widely used in successful scenarios. Preventive medicine focused on the early detection of potentially dangerous situations will use these analysis techniques to produce real-time alarms associated with previously analy~~z~~sed patterns during normal-life situations. NLP (Natural Language Processing) using NMT (Neural Machine Translation) models is also poorly represented in the medical domain, compared with the relevance that processing of human communications is having within artificial intelligence and applied linguistics areas: speech-to-text conversion, translation, summari~~z~~sation, disambiguation, understanding and generation of Natural Language. It is foreseeable that in the coming years, applications related to human language, whether written or spoken, will coloni~~z~~se the medical domain. A large amount of this type of data still unprocessed (medical records among them) and the possibility of using them in combination with other data (numbers and images) will favor the development of multimodal neural applications and will facilitate medical tasks not directly related to the diagnosis.

We have also compared our work with other reviews of deep learning in medicine published over the past five years. These documents were obtained from MEDLINE and, after the screening, 72 of them were considered. Their conclusions roughly correspond to the areas and applications highlighted in this systematic review. The largest difference found is that none of the publications follow the methodological expectations of the Cochrane reviews. They commonly lacked a definition of inclusion criteria that add the characteristics that must be detailed in the papers that describe the implementation of deep learning models in medicine. From the point of view of computer sciences, it is worth mentioning that data types were not considered, which however was done in the present paper.

None of the articles included in our review was conducted in Spain, which may be because current clinical data protection laws make it difficult to implement DL models, as well as the lack of a common structure in electronic medical records between different healthcare centers. We should also mention how the data from medical records are recorded and structured, because the majority of the reports are written in open text, with no encoded data that would permit a suitable extraction of variables, or enough detail to be able to develop deep learning models that could predict the risk or progression of diseases following patients' characteristics, combining this with sociodemographic population data.

To conclude, a high number of studies published in the Q1 did not meet the defined quality criteria. Further, the process to replicate the sample was not always detailed, and we found it quite surprising that the sizes of the initial datasets could be so small~~,~~ when consider that the basis of AI is big data. The lack of information in the papers about the validation of the models developed with external databases and the absence of descriptions of how the results could be used in routine clinical practice should be emphasized [112]. It may be necessary to reach a consensus on quality criteria for the studies and papers about deep learning in medicine.

Acknowledgments

We would like to thank Jaime Pérez Palomera, Borja García Lamas, Ignacio Moll and Pedro Chazarra from CEIEC and Marina Diaz Fernández from the Universidad de Francisco de Vitoria Library for their support during the course of this research.

References

1. Djedouboum, A. C., Ari, A., Adamou, A., Gueroui, A. M., Mohamadou, A., & Aliouat, Z. (2018). Big Data Collection in Large-Scale Wireless Sensor Networks. *Sensors*, *18*(12), 4474.
2. Sánchez-Mendiola. M., Martinez-Franco, A.I., García-Nocett, D.F. and Cervantes-Pérez, F. (2018). *Informática Biomédica: Big Data y Analítica en las Ciencias de la Salud*. Third Edition. Elsevier.
3. Normandeau, K. (2013). Beyond volume, variety and velocity is the issue of big data veracity. *Inside big data*.
4. Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited.
5. Samuel, A. L. (1988). Some studies in machine learning using the game of checkers. II recent progress. In Computer Games I (pp. 366-400). Springer, New York, NY.
6. Domingos, P. M. (2012). A few useful things to know about machine learning. *Commun. acm*, *55*(10), 78-87.
7. Müller, B., Reinhardt, J., & Strickland, M. T. (2012). *Neural networks: an introduction*. Springer Science & Business Media.
8. Hecht-Nielsen, R. (1988). Neurocomputing: picking the human brain. *IEEE spectrum*, *25*(3), 36-41.
9. Salman, S., & Liu, X. (2019). Overfitting mechanism and avoidance in deep neural networks. *arXiv preprint arXiv:1901.06566*.
10. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436.
11. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, *42*, 60-88.
12. Pham, T. T. (2018). Applying Machine Learning for Automated Classification of Biomedical Data in Subject-Independent Settings. Springer.
13. Zhou, S. K., Greenspan, H., & Shen, D. (Eds.). (2017). *Deep learning for medical image analysis*. Academic Press.
14. Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., & Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology*, *2*(4), 230-243.
15. Miller, D. D., & Brown, E. W. (2018). Artificial intelligence in medical practice: the question to the answer? *The American journal of medicine*, *131*(2), 129-133.
16. Jang, H. J., & Cho, K. O. (2019). Applications of deep learning for the analysis of medical data. *Archives of pharmacal research*, 1-13.
17. Bakator, M., & Radosav, D. (2018). Deep learning and medical diagnosis: A review of literature. *Multimodal Technologies and Interaction*, *2*(3), 47.
18. Lundervold, A. S., & Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik*, *29*(2), 102-127.
19. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).
20. Arasu, A., & Garcia-Molina, H. (2003). Extracting structured data from web pages. In Proceedings of the 2003 ACM SIGMOD international conference on Management of data (pp. 337-348). ACM.
21. Velicer, W. F., & Molenaar, P. C. (2012). Time series analysis for psychological research. *Handbook of Psychology, Second Edition*, *2*.

22. LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, *1*(4), 541-551.
23. Ji, S., Xu, W., Yang, M., & Yu, K. (2012). 3D convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, *35*(1), 221-231.
24. Kipf, T. N. & Welling, M. (2016). Semi-Supervised Classification with Graph Convolutional Networks. *5th International Conference on Learning Representations*.
25. Elman, J. (1990). Finding Structure in Time. Cognitive Science, 14, 179—211.
26. Saadatnejad, S., Oveisi, M., & Hashemi, M. (2019). LSTM-based ECG classification for continuous monitoring on personal wearable devices. *IEEE journal of biomedical and health informatics*, *24*(2), 515-523.
27. León, J., Escobar, J. J., Ortiz, A., Ortega, J., González, J., Martín-Smith, P., & Damas, M. (2020). Deep learning for EEG-based Motor Imagery classification: Accuracy-cost trade-off. *Plos one*, *15*(6), e0234178.
28. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.
29. Sathya, R., & Abraham, A. (2013). Comparison of supervised and unsupervised learning algorithms for pattern classification. *International Journal of Advanced Research in Artificial Intelligence*, *2*(2), 34-38.
30. Ballard, D. H. (1987, July). Modular Learning in Neural Networks. In *AAAI* (pp. 279-284).
31. Saravanan, S., & Juliet, S. (2020). Deep Medical Image Reconstruction with Autoencoders using Deep Boltzmann Machine Training. *EAI Endorsed Transactions on Pervasive Health and Technology*, 166360.
32. Gondara, L. (2016, December). Medical image denoising using convolutional denoising autoencoders. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)* (pp. 241-246). IEEE.
33. Pesteie, M., Abolmaesumi, P., & Rohling, R. N. (2019). Adaptive augmentation of medical data using independently conditional variational auto-encoders. *IEEE Transactions on Medical Imaging*, *38*(12), 2807-2820.
34. Evan, M. Y., Iglesias, J. E., Dalca, A. V., & Sabuncu, M. R. (2020, January). An Auto-Encoder Strategy for Adaptive Image Segmentation. In *Medical Imaging with Deep Learning*.
35. Uzunova, H., Schultz, S., Handels, H., & Ehrhardt, J. (2019). Unsupervised pathology detection in medical images using conditional variational autoencoders. *International journal of computer assisted radiology and surgery*, *14*(3), 451-461.
36. Chen, M., Shi, X., Zhang, Y., Wu, D., & Guizani, M. (2017). Deep features learning for medical image analysis with convolutional autoencoder neural network. *IEEE Transactions on Big Data*.
37. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A.C., & Bengio, Y. (2014). Generative Adversarial Nets. *NIPS*.
38. Luo, T. J., Fan, Y., Chen, L., Guo, G., & Zhou, C. (2020). EEG Signal Reconstruction Using a Generative Adversarial Network With Wasserstein Distance and Temporal-Spatial-Frequency Loss. *Frontiers in Neuroinformatics*, *14*.
39. Gupta, R., Sharma, A., & Kumar, A. (2020). Super-Resolution using GANs for Medical Imaging. *Procedia Computer Science*, *173*, 28-35.
40. Zhang, C., Song, Y., Liu, S., Lill, S., Wang, C., Tang, Z., ... & Cai, W. (2018, December). MS-GAN: GAN-based semantic segmentation of multiple sclerosis lesions in brain

magnetic resonance imaging. In *2018 Digital Image Computing: Techniques and Applications (DICTA)* (pp. 1-8). IEEE.

41. Sandfort, V., Yan, K., Pickhardt, P. J., & Summers, R. M. (2019). Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Scientific reports*, *9*(1), 1-9.
42. Suthaharan, S. (2016). Machine learning models and algorithms for big data classification. *Integr. Ser. Inf. Syst*, *36*, 1-12.
43. Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*, *151*(4), 264-269.
44. Choi, E., Schuetz, A., Stewart, W. F. & Sun, J. (2017). Using recurrent neural network models for early detection of heart failure onset. *JAMIA*, 24, 361-370.
45. Carneiro, G., Nascimento, J. C. & Freitas, A. (2012). The Segmentation of the Left Ventricle of the Heart From Ultrasound Data Using Deep Learning Architectures and Derivative-Based Search Methods. *IEEE Trans. Image Processing*, 21, 968-982.
46. Chaudhari, A.S., Fang, Z., Kogan, F., Wood, J., Stevens, K.J., Gibbons, E.K., Lee, J.H., Gold, G., & Hargreaves, B.A. (2018). Super-resolution musculoskeletal MRI using deep learning. *Magnetic resonance in medicine, 80 5*, 2139-2154.
47. Saltz, J.H., Gupta, R., Hou, L., Kurç, T.M., Singh, P., Nguyen, V.M., Samaras, D., Shroyer, K.R., Zhao, T., Batiste, R., Arnam, J.S., Shmulevich, I., Rao, A.U., Lazar, A.J., Sharma, A., & Thorsson, V. (2018). Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images. *Cell reports, 23 1*, 181-193.e7.
48. Apostolopoulos, S., Ciller, C., Zanet, S.D., Wolf, S., & Sznitman, R. (2016). RetiNet: Automatic AMD identification in OCT volumetric data. *CoRR, abs/1610.03628*.
49. Lam, C., Yu, C.Y., Huang, L., & Rubin, D.Y. (2018). Retinal Lesion Detection With Deep Learning Using Image Patches. *Investigative ophthalmology & visual science*.
50. Choi, H., Ha, S., Im, H.J., Paek, S.H., & Lee, D.S. (2017). Refining diagnosis of Parkinson's disease with deep learning-based interpretation of dopamine transporter imaging. *NeuroImage: Clinical*.
51. Rajaraman, S., Antani, S.K., Poostchi, M., Silamut, K., Hossain, M.A., Maude, R.J., Jaeger, S., & Thoma, G.R. (2018). Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. *PeerJ*.
52. Nielsen, A.B., Hansen, M.B., Tietze, A., & Mouridsen, K. (2018). Prediction of Tissue Outcome and Assessment of Treatment Effect in Acute Ischemic Stroke Using Deep Learning. *Stroke, 49 6*, 1394-1401.
53. Lee, H., Ryu, H., Chung, E., & Jung, C. (2017). Prediction of Bispectral Index during Target-controlled Infusion of Propofol and Remifentanil: A Deep Learning Approach. *Anesthesiology, 128 3*, 492-501.
54. Ning, Z., Luo, J., Li, Y.J., Han, S., Feng, Q., Xu, Y., Chen, W., Chen, T., & Zhang, Y. (2018). Pattern Classification for Gastrointestinal Stromal Tumors by Integration of Radiomics and Deep Convolutional Features. *IEEE Journal of Biomedical and Health Informatics, 23*, 1181-1191.
55. Zeng, L., Wang, H., Hu, P., Yang, B., Pu, W., Shen, H., Chen, X., Liu, Z., Yin, H., Tan, Q., Wang, K., & Hu, D. (2018). Multi-Site Diagnostic Classification of Schizophrenia Using Discriminant Deep Learning with Functional Connectivity MRI. *EBioMedicine*.
56. Ghesu, F.C., Georgescu, B., Zheng, Y., Hornegger, J., & Comaniciu, D. (2015). Marginal Space Deep Learning: Efficient Architecture for Detection in Volumetric Image Data. *MICCAI*.

57. Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., & Mougiakakou, S.G. (2016). Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network. *IEEE Transactions on Medical Imaging, 35*, 1207-1216.
58. Yasaka, K., Akai, H., Kunimatsu, A., Abe, O., & Kiryu, S. (2017). Liver Fibrosis: Deep Convolutional Neural Network for Staging by Using Gadoxetic Acid-enhanced Hepatobiliary Phase MR Images. *Radiology, 287 1*, 146-155.
59. Leibig, C., Allken, V., Ayhan, M.S., Berens, P., & Wahl, S. (2017). Leveraging uncertainty information from deep neural networks for disease detection. *Scientific Reports*.
60. Kooi, T., Litjens, G.J., Ginneken, B.V., Gubern-Mérida, A., Sánchez, C.I., Mann, R.M., Heeten, A.D., & Karssemeijer, N. (2017). Large scale deep learning for computer aided detection of mammographic lesions. *Medical image analysis, 35*, 303-312.
61. Kim, E., Kim, H., Han, K., Kang, B.J., Sohn, Y., Woo, O.H., & Lee, C.W. (2018). Applying Data-driven Imaging Biomarker in Mammography for Breast Cancer Screening: Preliminary Study. *Scientific Reports*.
62. Heinsfeld, A.S., Franco, A.R., Craddock, R.C., Buchweitz, A., & Meneguzzi, F. (2018). Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *NeuroImage: Clinical*.
63. Suk, H., Lee, S., & Shen, D. (2014). Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage, 101*, 569-582.
64. Grinsven, M.J., Ginneken, B.V., Hoyng, C.B., Theelen, T., & Sánchez, C.I. (2016). Fast Convolutional Neural Network Training Using Selective Data Sampling: Application to Haemorrhage Detection in Color Fundus Images. *IEEE Transactions on Medical Imaging, 35*, 1273-1284.
65. Wang, J., Yang, X., Cai, H., Tan, W., Jin, C., & Li, L. (2016). Discrimination of Breast Cancer with Microcalcifications on Mammography by Deep Learning. *Scientific reports*.
66. Kooi, T., Ginneken, B.V., Karssemeijer, N., & Heeten, A.D. (2017). Discriminating solitary cysts from soft tissue lesions in mammography using a pretrained deep convolutional neural network. *Medical physics, 44 3*, 1017-1027.
67. Fu, H., Cheng, J., Xu, Y., Zhang, C., Wong, D.W., Liu, J., & Cao, X. (2018). Disc-Aware Ensemble Network for Glaucoma Screening From Fundus Image. *IEEE Transactions on Medical Imaging, 37*, 2493-2501.
68. Yu, C., Yang, S., Kim, W., Jung, J., Chung, K., Lee, S.W., & Oh, B. (2018). Acral melanoma detection using a convolutional neural network for dermoscopy images. *PloS one*.
69. Wang, J., Ding, H., Bidgoli, F.A., Zhou, B., Iribarren, C., Molloi, S.Y., & Baldi, P. (2017). Detecting Cardiovascular Disease from Mammograms With Deep Learning. *IEEE Transactions on Medical Imaging, 36*, 1172-1181.
70. Iakovidis, D.K., Georgakopoulos, S.V., Vasilakakis, M., Koulaouzidis, A., & Plagianakos, V.P. (2018). Detecting and Locating Gastrointestinal Anomalies Using Deep Learning and Iterative Cluster Unification. *IEEE Transactions on Medical Imaging, 37*, 2196-2210.
71. Han, S.S., Park, G.H., Lim, W., Kim, M.S., Na, J.I., Park, I., & Chang, S.E. (2018). Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: Automatic construction of onychomycosis datasets by region-based convolutional deep neural network. *PloS one*.
72. Burgh, H.K., Schmidt, R., Westeneng, H.J., Reus, M.A., Berg, L.H., & Heuvel, M.P. (2017). Deep learning predictions of survival based on MRI in amyotrophic lateral sclerosis. *NeuroImage: Clinical*.
73. Lee, C.S., Baughman, D.M., & Lee, A.Y. (2016). Deep learning is effective for the classification of OCT images of normal versus Age-related Macular Degeneration. Ophthalmology. Retina, 1 4, 322-327.

74. Betancur, J., Commandeur, F., Motlagh, M.E., Sharir, T., Einstein, A.J., Bokhari, S.I., Fish, M., Ruddy, T.D., Kaufmann, P.A., Sinusas, A.J., Miller, E.J., Bateman, T.M., Dorbala, S., Carli, M.F., Germano, G., Otaki, Y., Tamarappoo, B., Dey, D., Berman, D.S., & Slomka, P.J. (2018). Deep Learning for Prediction of Obstructive Disease From Fast Myocardial Perfusion SPECT: A Multicenter Study. *JACC. Cardiovascular imaging, 11 11*, 1654-1663.
75. Quellec, G., Charrière, K., Boudi, Y., Cochener, B. & Lamard, M. (2017). Deep image mining for diabetic retinopathy screening. *Medical Image Analysis*, 39, 178-193.
76. Cheng, J., Ni, D., Chou, Y., Qin, J., Tiu, C.P., Chang, Y., Huang, C., Shen, D., & Chen, C. (2016). Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans. *Scientific reports*.
77. Song, Y., Zhang, Y., Yan, X., Liu, H., Zhou, M., Hu, B., & Yang, G. (2018). Computer-aided diagnosis of prostate cancer using a deep convolutional neural network from multiparametric MRI. *Journal of magnetic resonance imaging: JMRI, 48 6*, 1570-1577.
78. Li, S., Wei, J., Chan, H., Helvie, M.A., Roubidoux, M.A., Lu, Y., Zhou, C., Hadjiiski, L.M., & Samala, R.K. (2018). Computer-aided assessment of breast density: comparison of supervised deep learning and feature-based statistical learning. *Physics in medicine and biology, 63 2*, 025005.
79. Ngo, T.A., Lu, Z., & Carneiro, G. (2017). Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance. *Medical image analysis, 35*, 159-171.
80. Zhang, J., Xia, Y., Wu, Q., & Xie, Y. (2017). Classification of Medical Images and Illustrations in the Biomedical Literature Using Synergic Deep Learning. *ArXiv, abs/1706.09092*.
81. Araújo, T., Aresta, G., Castro, E.V., Rouco, J., Aguiar, P.D., Eloy, C., Polónia, A., & Campilho, A. (2017). Classification of breast cancer histology images using Convolutional Neural Networks. *PloS one*.
82. Han, Z., Wei, B., Zheng, Y., Yin, Y., Li, K., & Li, S. (2017). Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model. *Scientific Reports*.
83. Zhao, Y., Dong, Q., Zhang, S., Zhang, W., Chen, H., Jiang, X., Guo, L., Hu, X., Han, J., & Liu, T. (2018). Automatic Recognition of fMRI-Derived Functional Networks Using 3-D Convolutional Neural Networks. *IEEE Transactions on Biomedical Engineering, 65*, 1975-1984.
84. Tiulpin, A., Thevenot, J., Rahtu, E., Lehenkari, P., & Saarakkala, S. (2018). Automatic Knee Osteoarthritis Diagnosis from Plain Radiographs: A Deep Learning-Based Approach. *Scientific Reports*.
85. Lee, J., & Nishikawa, R.M. (2018). Automated mammographic breast density estimation using a fully convolutional network. *Medical physics, 45 3*, 1178-1190.
86. Esses, S.J., Lu, X., Zhao, T., Shanbhogue, K.P., Dane, B., Bruno, M., & Chandarana, H. (2018). Automated image quality evaluation of T2 -weighted liver MRI utilizing deep learning architecture. *Journal of magnetic resonance imaging: JMRI, 47 3*, 723-728.
87. Serj, M.F., Lavi, B., Hoff, G., & Valls, D.P. (2018). A Deep Convolutional Neural Network for Lung Cancer Diagnostic. *CoRR, abs/1804.08170*.
88. Yap, M.H., Pons, G., Martí, J., Ganau, S., Sentís, M., Zwiggelaar, R., Davison, A.K., & Martí, R. (2018). Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks. *IEEE Journal of Biomedical and Health Informatics, 22*, 1218-1226.
89. Du, X., Kurmann, T., Chang, P., Allan, M., Ourselin, S., Sznitman, R., Kelly, J.D., & Stoyanov, D. (2018). Articulated Multi-Instrument 2-D Pose Estimation Using Fully Convolutional Networks. *IEEE Transactions on Medical Imaging*.

90. Kim, K.H., Choi, S.H., & Park, S. (2017). Improving Arterial Spin Labeling by Using Deep Learning. *Radiology, 287 2*, 658-666.
91. Song, Y., Zhang, L., Chen, S., Ni, D., Lei, B.Y., & Wang, T. (2015). Accurate Segmentation of Cervical Cytoplasm and Nuclei Based on Multiscale Convolutional Network and Graph Partitioning. *IEEE Transactions on Biomedical Engineering, 62*, 2421-2433.
92. Zhang, Z., Liang, X., Dong, X., Xie, Y., & Cao, G. (2018). A Sparse-View CT Reconstruction Method Based on Combination of DenseNet and Deconvolution. IEEE Transactions on Medical Imaging, 37, 1407-1417.
93. Xue, Y., Zhang, R., Deng, Y., Chen, K., & Jiang, T. (2017). A preliminary examination of the diagnostic value of deep learning in hip osteoarthritis. *PloS one*.
94. Khosravan, N., Celik, H., Turkbey, B., Jones, E., Wood, B.J., & Bagci, U. (2019). A collaborative computer aided diagnosis (C-CAD) system with eye-tracking, sparse attentional model, and deep learning. *Medical image analysis, 51*, 101-115.
95. Shin, H. C., Lu, L., Kim, L., Seff, A., Yao, J., & Summers, R. M. (2016). Interleaved Text/Image Deep Mining on a Large-Scale Radiology Database for Automated Image Interpretation. Journal of Machine Learning Research, 17(1-31), 2.
96. Liu, M., Wu, W., Gu, Z., Yu, Z., Qi, F. & Li, Y. (2018). Deep learning based on Batch Normalization for P300 signal detection. *Neurocomputing*, 275, 288-297.
97. Ohsugi H., Hitoshi Tabuchi H., Enno H. & Ishitobi N. (2017) Accuracy of deep learning, a machine learning technology, using ultra-wide-field fundus ophthalmoscopy for detecting rhegmatogenous retinal detachment. Scientific Reports,7, 9425
98. Chen, C. M., Huang, Y. S., Fang, P. W., Liang, C. W., & Chang, R. F. (2020). A computer-aided diagnosis system for differentiation and delineation of malignant regions on whole-slide prostate histopathology image using spatial statistics and multidimensional DenseNet. *Medical Physics*, *47*(3), 1021-1033
99. Jang, R., Kim, N., Jang, M., Lee, K. H., Lee, S. M., Lee, K. H., ... & Seo, J. B. (2020). Assessment of the Robustness of Convolutional Neural Networks in Labeling Noise by Using Chest X-Ray Images From Multiple Centers. *JMIR Medical Informatics*, *8*(8), e18089.
100. Sujit, S. J., Coronado, I., Kamali, A., Narayana, P. A., & Gabr, R. E. (2019). Automated image quality evaluation of structural brain MRI using an ensemble of deep learning networks. *Journal of Magnetic Resonance Imaging*, *50*(4), 1260-1267.
101. Saha, S. K., Fernando, B., Cuadros, J., Xiao, D., & Kanagasingam, Y. (2018). Automated quality assessment of colour fundus images for diabetic retinopathy screening in telemedicine. *Journal of digital imaging*, *31*(6), 869-878.
102. Zhang, L., Fabbri, D., Upender, R., & Kent, D. (2019). Automated sleep stage scoring of the Sleep Heart Health Study using deep neural networks. *Sleep*, *42*(11), zsz159.
103. Kong, Y., Kong, X., He, C., Liu, C., Wang, L., Su, L., ... & Cheng, R. (2020). Constructing an automatic diagnosis and severity-classification model for acromegaly using facial photographs by deep learning. *Journal of hematology & oncology*, *13*(1), 1-4.
104. Das, A., Rad, P., Choo, K. K. R., Nouhi, B., Lish, J., & Martel, J. (2019). Distributed machine learning cloud teleophthalmology IoT for predicting AMD disease progression. *Future Generation Computer Systems*, *93*, 486-498
105. Kim, Y. D., Noh, K. J., Byun, S. J., Lee, S., Kim, T., Sunwoo, L., & Park, S. J. (2020). Effects of Hypertension, Diabetes, and Smoking on Age and Sex Prediction from Retinal Fundus Images. *Scientific reports*, *10*(1), 1-14.
106. Pan, I., Agarwal, S., & Merck, D. (2019). Generalizable inter-institutional classification of abnormal chest radiographs using efficient convolutional neural networks. *Journal of digital imaging*, *32*(5), 888-896

107. Zhang, C., Wu, S., Lu, Z., Shen, Y., Wang, J., Huang, P., ... & Xue, J. (2020). Hybrid Adversarial-Discriminative Network for Leukocyte Classification in Leukemia. *Medical Physics*.
108. Cheon, S., Kim, J., & Lim, J. (2019). The use of deep learning to predict stroke patient mortality. *International journal of environmental research and public health*, *16*(11), 1876.
109. Yu, R., Zheng, Y., Zhang, R., Jiang, Y., & Poon, C. C. (2019). Using a multi-task recurrent neural network with attention mechanisms to predict hospital mortality of patients. *IEEE Journal of Biomedical and Health Informatics*, *24*(2), 486-492.
110. Smith-Bindman, R., Kwan, M. L., Marlow, E. C., Theis, M. K., Bolch, W., Cheng, S. Y., & Pole, J. D. (2019). Trends in use of medical imaging in US health care systems and in Ontario, Canada, 2000-2016. *Jama*, *322*(9), 843-856.
111. Lee, Y., Kwon, J. M., Lee, Y., Park, H., Cho, H., & Park, J. (2018). Deep learning in the medical domain: predicting cardiac arrest using deep learning. *Acute and critical care*, *33*(3), 117.
112. Ravì, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G. Z. (2016). Deep learning for health informatics. IEEE journal of biomedical and health informatics, 21(1), 4-21.

Appendix A: MeSH terms for Deep Learning and Medicine

MeSH "Deep Learning" includes the following terms: Learning, Deep/ Hierarchical Learning/ Learning, Hierarchical.
MeSH "Medicine" includes the following terms: Specialties, Medical/ Specialties, Medical/ Medical Specialties/ Specialty, Medical/ Addiction Medicine/ Adolescent Medicine/ Aerospace Medicine/ Allergy and Immunology/ Anesthesiology/ Bariatric Medicine/ Behavioral Medicine/ Clinical Medicine/ Evidence-Based Medicine/ Precision Medicine/ Community Medicine/ Dermatology/ Disaster Medicine/ Emergency Medicine/ Pediatric Emergency Medicine/ Forensic Medicine/ General Practice/ Family Practice/ Genetics, Medical/ Geography, Medical/ Geriatrics/ Global Health/ Hospital Medicine/ Integrative Medicine Internal Medicine (Cardiology, Endocrinology, Gastroenterology, Hematology, Infectious Disease Medicine, Medical Oncology, Nephrology, Pulmonary Medicine, Rheumatology, Sleep Medicine Specialty)/ Military Medicine/ Molecular Medicine/ Naval Medicine/ Neurology/ Neuropathology/ Neurotology/ Osteopathic Medicine/ Palliative Medicine/ Pathology (Forensic Pathology, Neuropathology, Pathology, Clinical, Pathology, Molecular, Pathology, Surgical, Telepathology)/ Pediatrics (Neonatology, Pediatric Emergency Medicine, Perinatology, Perioperative Medicine)/ Physical and Rehabilitation Medicine/ Rehabilitation /Psychiatry (Adolescent Psychiatry, Biological Psychiatry, Child Psychiatry, Community Psychiatry, Forensic Psychiatry, Geriatric Psychiatry, Military Psychiatry, Neuropsychiatry)/ Public Health (Epidemiology, Preventive Medicine)/ Radiology (Nuclear Medicine, Radiation Oncology, Radiology, Interventional)/ Regenerative Medicine/ Reproductive Medicine (Andrology, Gynecology)/ Social Medicine/ Specialties, Surgical (Colorectal Surgery, General Surgery, Gynecology, Neurosurgery, Obstetrics, Ophthalmology, Orthognathic Surgery, Orthopedics, Otolaryngology, Surgery, Plastic, Surgical Oncology, Thoracic Surgery)/ Traumatology/ Urology/ Sports Medicine/ Telemedicine/ Theranostic Nanomedicine/ Travel Medicine/ Tropical Medicine/ Vaccinology/ Venereology/ Wilderness Medicine.

Appendix B: Selected papers of the review not cited in the text.

[1] Zhao, W., Yang, J., Sun, Y., Li, C., Wu, W., Jin, L., ... & Hua, Y. (2018). 3D deep learning from CT scans predicts tumor invasiveness of subcentimeter pulmonary adenocarcinomas. *Cancer research*, *78*(24), 6881-6889.
[2] Zhao, X., Wang, X., Xia, W., Li, Q., Zhou, L., Li, Q., ... & Wang, W. (2020). A cross-modal 3D deep learning for accurate lymph node metastasis prediction in clinical stage T1 lung adenocarcinoma. *Lung Cancer*.
[3] Shen, Y., Li, X., Liang, X., Xu, H., Li, C., Yu, Y., & Qiu, B. (2020). A deep-learning-based approach for adenoid hypertrophy diagnosis. *Medical Physics*.
[4] Hoseini, F., Shahbahrami, A., & Bayat, P. (2018). An efficient implementation of deep convolutional neural networks for MRI segmentation. *Journal of digital imaging*, *31*(5), 738-747.
[5] Pranav, R., Park, A., Irvin, J., Chute, C., Bereket, M., Domenico, M., ... & Patel, B. N. (2020). AppendiXNet: Deep Learning for Diagnosis of Appendicitis from A Small Dataset of CT Exams Using Video Pretraining. *Scientific Reports (Nature Publisher Group)*, *10*(1).
[6] Yıldırım, Ö., Pławiak, P., Tan, R. S., & Acharya, U. R. (2018). Arrhythmia detection using deep convolutional neural network with long duration ECG signals. *Computers in biology and medicine*, *102*, 411-420.
[7] Schwyzer, M., Martini, K., Benz, D. C., Burger, I. A., Ferraro, D. A., Kudura, K., & Messerli, M. (2020). Artificial intelligence for detecting small FDG-positive lung nodules in digital PET/CT: impact of image reconstructions on diagnostic performance. *European Radiology*, *30*(4), 2031-2040.
[8] Vakanski, A., Xian, M., & Freer, P. E. (2020). Attention-Enriched Deep Learning Model for Breast Tumor Segmentation in Ultrasound Images. *Ultrasound in Medicine & Biology*, *46*(10), 2819-2833.
[9] Vakanski, A., Xian, M., & Freer, P. E. (2020). Attention-Enriched Deep Learning Model for Breast Tumor Segmentation in Ultrasound Images. *Ultrasound in Medicine & Biology*, *46*(10), 2819-2833.
[10] Candemir, S., White, R. D., Demirer, M., Gupta, V., Bigelow, M. T., Prevedello, L. M., & Erdal, B. S. (2020). Automated coronary artery atherosclerosis detection and weakly supervised localization on coronary CT angiography with a deep 3-dimensional convolutional neural network. *Computerized Medical Imaging and Graphics*, 101721.
[11] Stoean, C., Stoean, R., Atencia, M., Abdar, M., Velázquez-Pérez, L., Khosravi, A., & Joya, G. (2020). Automated Detection of Presymptomatic Conditions in Spinocerebellar Ataxia Type 2 Using Monte Carlo Dropout and Deep Neural Network Techniques with
[12] van den Heuvel, T. L., Petros, H., Santini, S., de Korte, C. L., & van Ginneken, B. (2019). Automated fetal head detection and circumference estimation from free-hand ultrasound sweeps using deep learning in resource-limited countries. *Ultrasound in medicine & biology*, *45*(3), 773-785.
[13] Zhuge, Y., Ning, H., Mathen, P., Cheng, J. Y., Krauze, A. V., Camphausen, K., & Miller, R. W. (2020). Automated glioma grading on conventional MRI images using deep convolutional neural networks. *Medical Physics*.

[14] Ma, X., Wei, J., Zhou, C., Helvie, M. A., Chan, H. P., Hadjiiski, L. M., & Lu, Y. (2019). Automated pectoral muscle identification on MLO-view mammograms: Comparison of deep neural network to conventional computer vision. *Medical physics*, *46*(5), 2103-2114.
[15] Zabihollahy, F., Schieda, N., Krishna Jeyaraj, S., & Ukwatta, E. (2019). Automated segmentation of prostate zonal anatomy on T2-weighted (T2W) and apparent diffusion coefficient (ADC) map MR images using U-Nets. *Medical physics*, *46*(7), 3078-3090.
[16] Ferreira, P. F., Martin, R. R., Scott, A. D., Khalique, Z., Yang, G., Nielles-Vallespin, S., ... & Firmin, D. N. (2020). Automating in vivo cardiac diffusion tensor postprocessing with deep learning–based segmentation. *Magnetic Resonance in Medicine*.
[17] Zhong, T., Huang, X., Tang, F., Liang, S., Deng, X., & Zhang, Y. (2019). Boosting-based cascaded convolutional neural networks for the segmentation of CT organs-at-risk in nasopharyngeal carcinoma. *Medical physics*, *46*(12), 5602-5611.
[18] Byra, M., Galperin, M., Ojeda-Fournier, H., Olson, L., O'Boyle, M., Comstock, C., & Andre, M. (2019). Breast mass classification in sonography with transfer learning using a deep convolutional neural network and color conversion. *Medical physics*, *46*(2), 746-755.
[19] Agrawal, V., Udupa, J., Tong, Y., & Torigian, D. (2020). BRR-Net: A tandem architectural CNN–RNN for automatic body region localization in CT images. *Medical Physics*.
[20] Bi, X., Li, S., Xiao, B., Li, Y., Wang, G., & Ma, X. (2020). Computer aided Alzheimer's disease diagnosis by an unsupervised deep learning technology. *Neurocomputing*, *392*, 296-304.
[21] Song, Y., Zhang, Y. D., Yan, X., Liu, H., Zhou, M., Hu, B., & Yang, G. (2018). Computer-aided diagnosis of prostate cancer using a deep convolutional neural network from multiparametric MRI. *Journal of Magnetic Resonance Imaging*, *48*(6), 1570-1577.
[22] Stember, J. N., Chang, P., Stember, D. M., Liu, M., Grinband, J., Filippi, C. G., & Jambawalikar, S. (2019). Convolutional neural networks for the detection and measurement of cerebral aneurysms on magnetic resonance angiography. *Journal of digital imaging*, *32*(5), 808-815
[23] Ko, H., Chung, H., Kang, W. S., Kim, K. W., Shin, Y., Kang, S. J., ... & Lee, J. (2020). COVID-19 pneumonia diagnosis using a simple 2D deep learning framework with a single chest CT Image: Model Development and Validation. *Journal of Medical Internet Research*, *22*(6), e19569.
[24] Kulkarni, P. M., Robinson, E. J., Pradhan, J. S., Gartrell-Corrado, R. D., Rohr, B. R., Trager, M. H., & Rizk, E. M. (2020). Deep learning based on standard H&E images of primary melanoma tumors identifies patients at risk for visceral recurrence and death. *Clinical Cancer Research*, *26*(5), 1126-1134.
[25] Bychkov, D., Linder, N., Turkki, R., Nordling, S., Kovanen, P. E., Verrill, C., ... & Lundin, J. (2018). Deep learning based tissue analysis predicts outcome in colorectal cancer. *Scientific reports*, *8*(1), 1-11.
[26] Antico, M., Sasazawa, F., Dunnhofer, M., Camps, S. M., Jaiprakash, A. T., Pandey, A. K., & Fontanarosa, D. (2020). Deep learning-based femoral cartilage automatic segmentation in ultrasound imaging for guidance in robotic knee arthroscopy. *Ultrasound in Medicine & Biology*, *46*(2), 422-435.
[27] Scannell, C. M., Veta, M., Villa, A. D., Sammut, E. C., Lee, J., Breeuwer, M., & Chiribiri, A. (2020). Deep-Learning-Based Preprocessing for Quantitative Myocardial Perfusion MRI. *Journal of Magnetic Resonance Imaging*, *51*(6), 1689-1696.
[28] Yuan, Y., Qin, W., Ibragimov, B., Zhang, G., Han, B., Meng, M. Q. H., & Xing, L. (2019). Densely Connected Neural Network With Unbalanced Discriminant and Category Sensitive

Constraints for Polyp Recognition. *IEEE Transactions on Automation Science and Engineering*, *17*(2), 574-583.
[29] Papandrianos, N., Papageorgiou, E., Anagnostis, A., & Papageorgiou, K. (2020). Efficient Bone Metastasis Diagnosis in Bone Scintigraphy Using a Fast Convolutional Neural Network Architecture. *Diagnostics*, *10*(8), 532.
[30] Park, H., Lee, H. J., Kim, H. G., Ro, Y. M., Shin, D., Lee, S. R., ... & Kong, M. (2019). Endometrium segmentation on transvaginal ultrasound image using key-point discriminator. *Medical physics*, *46*(9), 3974-3984.
[31] Estrada, S., Lu, R., Conjeti, S., Orozco-Ruiz, X., Panos-Willuhn, J., Breteler, M. M., & Reuter, M. (2020). FatSegNet: A fully automated deep learning pipeline for adipose tissue segmentation on abdominal dixon MRI. *Magnetic Resonance in Medicine*, *83*(4), 1471-1483.
[32] Wang, J., Chen, X., Lu, H., Zhang, L., Pan, J., Bao, Y., ... & Qian, D. (2020). Feature-shared adaptive-boost deep learning for invasiveness classification of pulmonary subsolid nodules in CT images. *Medical Physics*, *47*(4), 1738-1749.
[33] Berhane, H., Scott, M., Elbaz, M., Jarvis, K., McCarthy, P., Carr, J., ... & Rigsby, C. K. (2020). Fully automated 3D aortic segmentation of 4D flow MRI for hemodynamic analysis using deep learning. *Magnetic resonance in medicine*.
[34] Ornek, A. H., Ceylan, M., & Ervural, S. (2019). Health status detection of neonates using infrared thermography and deep convolutional neural networks. *Infrared Physics & Technology*, *103*, 103044.
[35] Porter, E., Fuentes, P., Siddiqui, Z., Thompson, A., Levitin, R., Solis, D., ... & Guerrero, T. (2020). Hippocampus Segmentation on non-Contrast CT using Deep Learning. *Medical Physics*.
[36] Kudva, V., Prasad, K., & Guruvare, S. (2019). Hybrid Transfer Learning for Classification of Uterine Cervix Images for Cervical Cancer Screening. *Journal of Digital Imaging*, 1-13.
[37] Meisel, C., & Bailey, K. A. (2019). Identifying signal-dependent information about the preictal state: A comparison across ECoG, EEG and EKG using deep learning. *EBioMedicine*, *45*, 422-431
[38] Verburg, E., Wolterink, J. M., de Waard, S. N., Išgum, I., van Gils, C. H., Veldhuis, W. B., & Gilhuijs, K. G. (2019). Knowledge-based and deep learning-based automated chest wall segmentation in magnetic resonance images of extremely dense breasts. *Medical physics*, *46*(10), 4405-4416.
[39] Hussein, S., Kandel, P., Bolan, C. W., Wallace, M. B., & Bagci, U. (2019). Lung and pancreatic tumor characterization in the deep learning era: novel supervised and unsupervised learning approaches. *IEEE transactions on medical imaging*, *38*(8), 1777-1787.
[40] Wang, Q., Shen, F., Shen, L., Huang, J., & Sheng, W. (2019). Lung nodule detection in CT images using a raw patch-based convolutional neural network. *Journal of digital imaging*, *32*(6), 971-979.
[41] Park, B., Park, H., Lee, S. M., Seo, J. B., & Kim, N. (2019). Lung segmentation on HRCT and volumetric CT for diffuse interstitial lung disease using deep convolutional neural networks. *Journal of Digital Imaging*, *32*(6), 1019-1026.
[42] Mutasa, S., Chang, P. D., Ruzal-Shapiro, C., & Ayyala, R. (2018). MABAL: a novel deep-learning architecture for machine-assisted bone age labeling. *Journal of digital imaging*, *31*(4), 513-519.

[43] Apiparakoon, T., Rakratchatakul, N., Chantadisai, M., Vutrapongwatana, U., Kingpetch, K., Sirisalipoch, S., ... & Chuangsuwanich, E. (2020). MaligNet: Semisupervised Learning for Bone Lesion Instance Segmentation Using Bone Scintigraphy. *IEEE Access*, *8*, 27047-27066.

[44] Nie, D., Lu, J., Zhang, H., Adeli, E., Wang, J., Yu, Z., ... & Shen, D. (2019). Multi-channel 3D deep feature learning for survival time prediction of brain tumor patients using multi-modal neuroimages. *Scientific reports*, *9*(1), 1-14.

[45] Kats, L., Vered, M., Blumer, S., & Kats, E. (2020). Neural Network Detection and Segmentation of Mental Foramen in Panoramic Imaging. *Journal of Clinical Pediatric Dentistry*, *44*(3), 168-173.

[46] Yu-Heng, L., Wei-Ning, C., Te-Cheng, H., Lin, C., Tsao, Y., & Semon, W. (2020). Overall survival prediction of non-small cell lung cancer by integrating microarray and clinical data with deep learning. *Scientific Reports (Nature Publisher Group)*, *10*(1).

[47] Fu, Y., Lei, Y., Wang, T., Tian, S., Patel, P., Jani, A. B., ... & Yang, X. (2020). Pelvic multi-organ segmentation on cone-beam CT for prostate adaptive radiotherapy. *Medical Physics*.

[48] Nobashi, T., Zacharias, C., Ellis, J. K., Ferri, V., Koran, M. E., Franc, B. L., ... & Davidzon, G. A. (2020). Performance comparison of individual and ensemble CNN models for the classification of brain 18F-FDG-PET scans. *Journal of digital imaging*, *33*(2), 447-455.

[49] Dreizin, D., Zhou, Y., Zhang, Y., Tirada, N., & Yuille, A. L. (2020). Performance of a deep learning algorithm for automated segmentation and quantification of traumatic pelvic hematomas on CT. *Journal of Digital Imaging*, *33*(1), 243-251.

[50] Yeh, H. Y., Chao, C. T., Lai, Y. P., & Chen, H. W. (2020). Predicting the Associations between Meridians and Chinese Traditional Medicine Using a Cost-Sensitive Graph Convolutional Neural Network. *International Journal of Environmental Research and Public Health*, *17*(3), 740.

[51] Yuan, Y., Qin, W., Buyyounouski, M., Ibragimov, B., Hancock, S., Han, B., & Xing, L. (2019). Prostate cancer classification with multiparametric MRI transfer learning model. *Medical physics*, *46*(2), 756-765.

[52] Chibuta, S., & Acar, A. C. (2020). Real-time Malaria Parasite Screening in Thick Blood Smears for Low-Resource Setting. *Journal of Digital Imaging*, 1-13.

[53] Østvik, A., Smistad, E., Aase, S. A., Haugen, B. O., & Lovstakken, L. (2019). Real-time standard view classification in transthoracic echocardiography using convolutional neural networks. *Ultrasound in medicine & biology*, *45*(2), 374-384

[54] Zeiser, F. A., da Costa, C. A., Zonta, T., Marques, N. M., Roehe, A. V., Moreno, M., & da Rosa Righi, R. (2020). Segmentation of Masses on Mammograms Using Data Augmentation and Deep Learning. *Journal of Digital Imaging*, 1-11.

[55] Cheon, S., Kim, J., & Lim, J. (2019). The use of deep learning to predict stroke patient mortality. *International journal of environmental research and public health*, *16*(11), 1876.

[56] Barragán-Montero, A. M., Nguyen, D., Lu, W., Lin, M. H., Norouzi-Kandalan, R., Geets, X., ... & Jiang, S. (2019). Three-dimensional dose prediction for lung IMRT patients with deep neural networks: robust learning from heterogeneous beam configurations. *Medical physics*, *46*(8), 3679-3691.

[57] Zhou, C., Fan, H., & Li, Z. (2019). Tonguenet: Accurate localization and segmentation for tongue images using deep neural networks. *IEEE Access*, *7*, 148779-148789.

[58] Yu, R., Zheng, Y., Zhang, R., Jiang, Y., & Poon, C. C. (2019). Using a multi-task recurrent neural network with attention mechanisms to predict hospital mortality of patients. *IEEE Journal of Biomedical and Health Informatics*, *24*(2), 486-492.
[59] Bohara, G., Sadeghnejad Barkousaraie, A., Jiang, S., & Nguyen, D. (2020). Using deep learning to predict beam-tunable Pareto optimal dose distribution for intensity-modulated radiation therapy. *Medical Physics*.
[60] Zhou, J., Luo, L. Y., Dou, Q., Chen, H., Chen, C., Li, G. J., ... & Heng, P. A. (2019). Weakly supervised 3D deep learning for breast cancer classification and localization of the lesions in MR images. *Journal of Magnetic Resonance Imaging*, *50*(4), 1144-1151.
[61] Narayana, P. A., Coronado, I., Sujit, S. J., Wolinsky, J. S., Lublin, F. D., & Gabr, R. E. (2020). Deep-learning-based neural tissue segmentation of MRI in multiple sclerosis: Effect of training set size. *Journal of Magnetic Resonance Imaging*, *51*(5), 1487-1496