

REVIEW

Open Access



Transfer learning for medical image classification: a literature review

Hee E. Kim^{1*} , Alejandro Cosa-Linan¹, Nandhini Santhanam¹, Mahboubeh Jannesari¹, Mate E. Maros^{1†} and Thomas Ganslandt^{1,2†}

Abstract

Background: Transfer learning (TL) with convolutional neural networks aims to improve performances on a new task by leveraging the knowledge of similar tasks learned in advance. It has made a major contribution to medical image analysis as it overcomes the data scarcity problem as well as it saves time and hardware resources. However, transfer learning has been arbitrarily configured in the majority of studies. This review paper attempts to provide guidance for selecting a model and TL approaches for the medical image classification task.

Methods: 425 peer-reviewed articles were retrieved from two databases, PubMed and Web of Science, published in English, up until December 31, 2020. Articles were assessed by two independent reviewers, with the aid of a third reviewer in the case of discrepancies. We followed the PRISMA guidelines for the paper selection and 121 studies were regarded as eligible for the scope of this review. We investigated articles focused on selecting backbone models and TL approaches including feature extractor, feature extractor hybrid, fine-tuning and fine-tuning from scratch.

Results: The majority of studies ($n = 57$) empirically evaluated multiple models followed by deep models ($n = 33$) and shallow ($n = 24$) models. Inception, one of the deep models, was the most employed in literature ($n = 26$). With respect to the TL, the majority of studies ($n = 46$) empirically benchmarked multiple approaches to identify the optimal configuration. The rest of the studies applied only a single approach for which feature extractor ($n = 38$) and fine-tuning from scratch ($n = 27$) were the two most favored approaches. Only a few studies applied feature extractor hybrid ($n = 7$) and fine-tuning ($n = 3$) with pretrained models.

Conclusion: The investigated studies demonstrated the efficacy of transfer learning despite the data scarcity. We encourage data scientists and practitioners to use deep models (e.g. ResNet or Inception) as feature extractors, which can save computational costs and time without degrading the predictive power.

Keywords: Deep learning, Transfer learning, Fine-tuning, Convolutional neural network, Medical image analysis

Introduction

Medical image analysis is a robust subject of research, with millions of studies having been published in the last decades. Some recent examples include computer-aided tissue detection in whole slide images (WSI) and the diagnosis of COVID-19 pneumonia from chest images. Traditionally, sophisticated image feature extraction or discriminant handcrafted features (e.g. histograms of oriented gradients (HOG) features [1] or local binary pattern (LBP) features [2]) have dominated

*Correspondence: HeeEun.Kim@medma.uni-heidelberg.de

†Mate E. Maros and Thomas Ganslandt have contributed equally to this work

¹ Department of Biomedical Informatics at the Center for Preventive Medicine and Digital Health (CPD-BW), Medical Faculty Mannheim, Heidelberg University, Theodor-Kutzer-Ufer 1-3, 68167 Mannheim, Germany

Full list of author information is available at the end of the article



the field of image analysis, but the recent emergence of deep learning (DL) algorithms has inaugurated a shift towards non-handcrafted engineering, permitting automated image analysis. In particular, convolutional neural networks (CNN) have become the workhorse DL algorithm for image analysis. In recent data challenges for medical image analysis, all of the top-ranked teams utilized CNN. For instance, the top-ten ranked solutions, excepting one team, had utilized CNN in the CAMELYON17 challenge for automated detection and classification of breast cancer metastases in whole slide images [3]. It has also been demonstrated that the features extracted from DL surpassed that of the handcrafted methods by Shi et al. [4].

However, DL algorithms including CNN require—under preferable circumstances—a large amount of data for training; hence follows the data scarcity problem. Particularly, the limited size of medical cohorts and the cost of expert-annotated data sets are some well-known challenges. Many research endeavors have tried to overcome this problem with transfer learning (TL) or domain adaptation [5] techniques. These aim to achieve high performance on target tasks by leveraging knowledge learned from source tasks. A pioneering review paper of TL was contributed by Pan and Yang [6] in 2010, and they classified TL techniques from a labeling aspect, while Weiss et al. [7] summarized TL studies based on homogeneous and heterogeneous approaches. Most recently in 2020, Zhuang et al. [8] reviewed more than forty representative TL approaches from the perspectives of data and models. Unsupervised TL is an emerging subject and has recently received increasing attention from researchers. Wilson and Cook [9] surveyed a large number of articles of unsupervised deep domain adaptation. Most recently, generative adversarial networks (GANs)-based frameworks [10–12] gained momentum, a particularly promising approach is DANN [13]. Furthermore, multiple kernel active learning [14] and collaborative unsupervised methods [15] have also been utilized for unsupervised TL.

Some studies conducted a comprehensive review focused primarily on DL in the medical domain. Litjens et al. [16] reviewed DL for medical image analysis by summarizing over 300 articles, while Chowdhury et al. [17] reviewed the state-of-the-art research on self-supervised learning in medicine. On the other hand, others surveyed articles focusing on TL with a specific case study such as microorganism counting [18], cervical cytopathology [19], neuroimaging biomarkers of Alzheimer's disease [20] and magnetic resonance brain imaging in general [21].

In this paper, we aimed to conduct a survey on TL with pretrained CNN models for medical image analysis

across use cases, data subjects and data modalities. Our major contributions are as follows:

- (i) An overview of contributions to the various case studies is presented;
- (ii) Actionable recommendations on how to leverage TL for medical image classification are provided;
- (iii) Publicly available medical datasets are compiled with URL as a supplementary material.

The rest of this paper is organized as follows. Section 2 covers the background knowledge and the most common notations used in the following sections. In Sect. 3, we describe the protocol for the literature selection. In Sect. 4, the results obtained are analyzed and compared. Critical discussions are presented in Sect. 5. Finally, we end with a conclusion and the lessons learned in Sect. 6. Figure 1 is the main diagram which presents the whole manuscript.

Background

Transfer learning

Transfer learning (TL) stems from cognitive research, which uses the idea, that knowledge is transferred across related tasks to improve performances on a new task. It is well-known that humans are able to solve similar tasks by leveraging previous knowledge. The formal definition of TL is defined by Pan and Yang with notions of domains and tasks. “A domain consists of a feature space \mathcal{X} and marginal probability distribution $P(X)$, where $X = \{x_1, \dots, x_n\} \in \mathcal{X}$. Given a specific domain denoted by $D = \{\mathcal{X}, P(X)\}$, a task is denoted by $T = \{\mathcal{Y}, f(\cdot)\}$ where \mathcal{Y} is a label space and $f(\cdot)$ is an objective predictive function. A task is learned from the pair $\{x_i, y_i\}$ where $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$. Given a source domain D_S and learning task T_S , a target domain D_T and learning task T_T , transfer learning aims to improve the learning of the target predictive function $f_T(\cdot)$ in D_T by using the knowledge in D_S and T_S ” [6].

Analogously, one can learn how to drive a motorbike T_T (transferred task) based on one's cycling skill T_S (source task) where driving two-wheel vehicles is regarded as the same domain $D_S = D_T$. This does not mean that one will not learn how to drive a motorbike without riding a bike, but it takes less effort to practice driving the motorbike by adapting one's cycling skills. Similarly, learning the parameters of a network from scratch will require larger annotated datasets and a longer training time to achieve an acceptable performance.

Convolutional neural networks using imageNet

Convolutional neural networks (CNN) are a special type of deep learning that processes grid-like topology data such as image data. Unlike the standard neural network

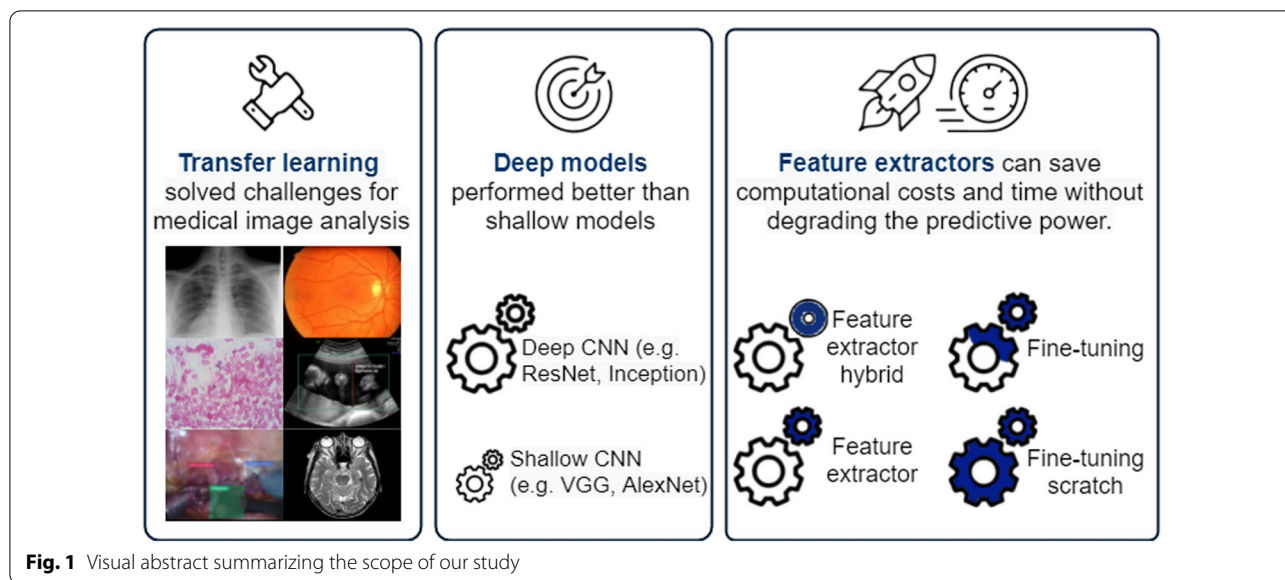


Fig. 1 Visual abstract summarizing the scope of our study

consisting of fully connected layers only, CNN consists of at least one convolutional layer. Several pretrained CNN models are publicly accessible online with downloadable parameters. They were pretrained with millions of natural images on the ImageNet dataset (ImageNet large scale visual recognition challenge; ILSVRC) [22].

In this paper, CNN models are denoted as backbone models. Table 1 summarizes the five most popular models in chronological order from top to bottom. LeNet [23] and AlexNet [24] are the first generations of CNN models developed in 1998 and 2012 respectively. Both are relatively shallow compared to other models that are developed recently. After AlexNet won the ImageNet large scale visual recognition challenge (ILSVRC) in 2012, designing novel networks became an emerging topic among researchers. VGG [25], also referred to as OxfordNet, is recognized as the first deep model, while GoogLeNet [26], also known as Inception1, set the new state of the art in the ILSVRC 2014. Inception introduced the novel block concept that

employs a set of filters with different sizes, and its deep networks were constructed by concatenating the multiple outputs. However, in the architecture of very deep networks, the parameters of the earlier layers are poorly updated during training because they are too far from the output layer. This problem is known as the vanishing gradient problem which was successfully addressed by ResNet [27] by introducing residual blocks with skip connections between layers.

The number of parameters of one filter is calculated by $(a * b * c) + 1$, where $a * b$ is the filter dimension, c is the number of filters in the previous layer and added 1 is the bias. The total number of parameters is the summation of the parameters of each filter. In the classifier head, all models use the Softmax function except LeNet-5, which utilizes the hyperbolic tangent function. The Softmax function fits well with the classification problem because it can convert feature vectors to the probability distribution for each class candidate.

Table 1 Overview of five backbone models

Model type	Model	Released year	Parameters (all)	Parameters (FE only)	Trainable layers (FE + FC layers)	Dataset
Shallow and linear	LeNet5	1998	60,000	1,716	4 (2 + 2)	MNIST
	AlexNet	2012	62.3 M	3.7 M	8 (5 + 3)	ImageNet
	VGG16	2014	134.2 M	14.7 M	16 (13 + 3)	
Deep	GoogLeNet	2014	5.3 M	5.3 M	22 (21 + 1)	
	ResNet50	2015	25.6 M	23.5 M	51 (50 + 1)	

FE: feature extraction, FC: fully connected layers; MNIST database: Modified National Institute of Standards and Technology database of handwritten digits with 60,000 training and 10,000 test images, ImageNet database: organized according to the WordNet hierarchy with over 14 million hand-annotated images for visual object recognition research

Transfer learning with convolutional neural networks

TL with CNN is the idea that knowledge can be transferred at the parametric level. Well-trained CNN models utilize the parameters of the convolutional layers for a new task in the medical domain. Specifically, in TL with CNN for medical image classification, a medical image classification (target task) can be learned by leveraging the generic features learned from the natural image classification (source task) where labels are available in both domains. For simplicity, the terminology of TL in the remainder of the paper refers to homogeneous TL (i.e. both domains are image analysis) with pretrained CNN models using ImageNet data for medical image classification in a supervisory manner.

Roughly, there are two TL approaches to leveraging CNN models: either feature extractor or fine-tuning. The feature extractor approach freezes the convolutional layers, whereas the fine-tuning approach updates parameters during model fitting. Each can be further divided into two subcategories; hence, four TL approaches are defined and surveyed in this paper. They are intuitively visualized in Fig. 2. Feature extractor hybrid (Fig. 2a) discards the FC layers and attaches a machine learning algorithm such as SVM or Random Forest classifier into the feature extractor, whereas the skeleton of the given networks remains the same in the other types (Fig. 2b-d). Fine-tuning from scratch is the most time-intensive approach because it updates the entire ensemble of parameters during the training process.

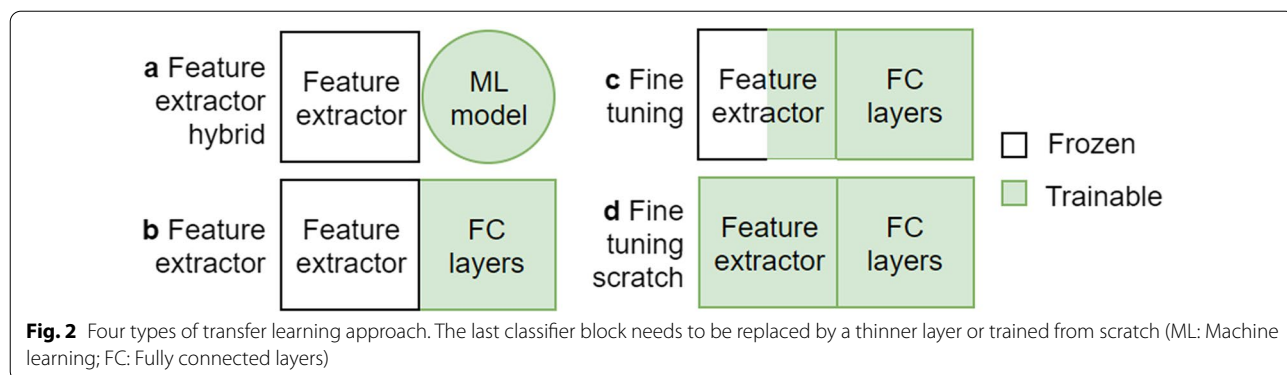
Methods

Publications were retrieved from two peer-reviewed databases (PubMed database on January 2, 2021, and Web of Science database on January 22, 2021). Papers were selected based on the following four conditions: (1) convolutional or CNN should appear in the title or abstract; (2) image data analysis should be considered; (3) “transfer learning” or “pretrained” should appear in the title or abstract; finally, (4) only experimental studies

were considered. The time constraint is specified only for the latest date, which is December 31, 2020. The exact search strings used for these two databases are denoted in Appendix A. Duplicates were merged before screening assessment. The first author screened the title, abstract and methods in order to exclude studies proposing a novel CNN model. Typically, this type of study stacked up multiple CNN models or concatenated CNN models and handcrafted features, and then compared its efficacy with other CNN models. Non-classification tasks, and those publications which fell outside the aforementioned date range, were also excluded. For the eligibility assessment, full texts were examined by two researchers. A third, independent researcher was involved in decision-making in the case of discrepancy between the two researchers.

Methodology analysis

Eight properties of 121 research articles were surveyed, investigated, compared and summarized in this paper. Five are quantitative properties and three are qualitative properties. They are specified as follows: (1) Off-the-shelf CNN model type (AlexNet, CaffeNet, Inception1, Inception2, Inception3, Inception4, Inception-Resnet, LeNet, MobileNet, ResNet, VGG16, VGG19, DenseNet, Xception, many or else); (2) Model performances (accuracy, AUC, sensitivity and specificity); (3) Transfer learning type (feature extractor, feature extractor hybrid, fine-tuning, fine-tuning or many); (4) Fine-tuning ratio; (5) Data modality (endoscopy, CT/CAT scan, mammographic, microscopy, MRI, OCT, PET, photography, sonography, SPECT, X-ray/radiography or many); (6) Data subject (abdominopelvic cavity, alimentary system, bones, cardiovascular system, endocrine glands, genital systems, joints, lymphoid system, muscles, nervous system, tissue specimen, respiratory system, sense organs, the integument, thoracic cavity, urinary system, many or else); (7) Data quantity; and (8) The number of classes. They fall



into one of three categories, namely model, transfer learning or data.

Results

Figure 3 shows the PRISMA flow diagram of paper selection. We initially retrieved 467 papers from PubMed and Web of Science. 42 duplicates were merged from two databases, and then 425 studies were assessed for screening. 189 studies were excluded during the screening phase, and then full texts of 236 studies were assessed for the next stage. 114 studies were disqualified from inclusion, resulting in 121 studies. These selected studies were further investigated and organized with respect to their backbone model and TL type. The data characteristics and model performance were also analyzed to gain insights regarding how to employ TL.

Figure 4a shows that studies of TL for medical image classification have emerged since 2016 with a 4-year delay after AlexNet [24] won the ImageNet Challenge in 2012. Since then the number of publications grew rapidly for consecutive years. Studies published in 2020 seem shrinking compared to the number of publications in

2019, because the process of indexing a publication may take anywhere from three to six months.

Backbone model

The majority of the studies (n=57) evaluated several backbone models empirically as depicted in Fig. 4b. For example, Rahaman and his colleagues [28] contributed an intensive benchmark study by evaluating fifteen models, namely: VGG16, VGG19, ResNet50, ResNet101, ResNet152, ResNet50V2, ResNet101V2, ResNet152V2, Inception3, InceptionResNet2, MobileNet1, DenseNet121, DenseNet169, DenseNet201 and XceptionNet. They concluded that VGG19 presented the highest accuracy of 89.3%. This result is exceptional because other studies reported that deeper models (e.g. Inception and ResNet) performed better than the shallow models (e.g. VGG and AlexNet). Five studies [29–33] compared Inception and VGG and reported that Inception performed better, and Ovalle-Magallanes et al. [34] also concluded that Inception3 outperformed compared to ResNet50 and VGG16. Finally, Talo et al. [35] reported that ResNet50 achieved the best classification accuracy compared to AlexNet, VGG16, ResNet18 and ResNet34.

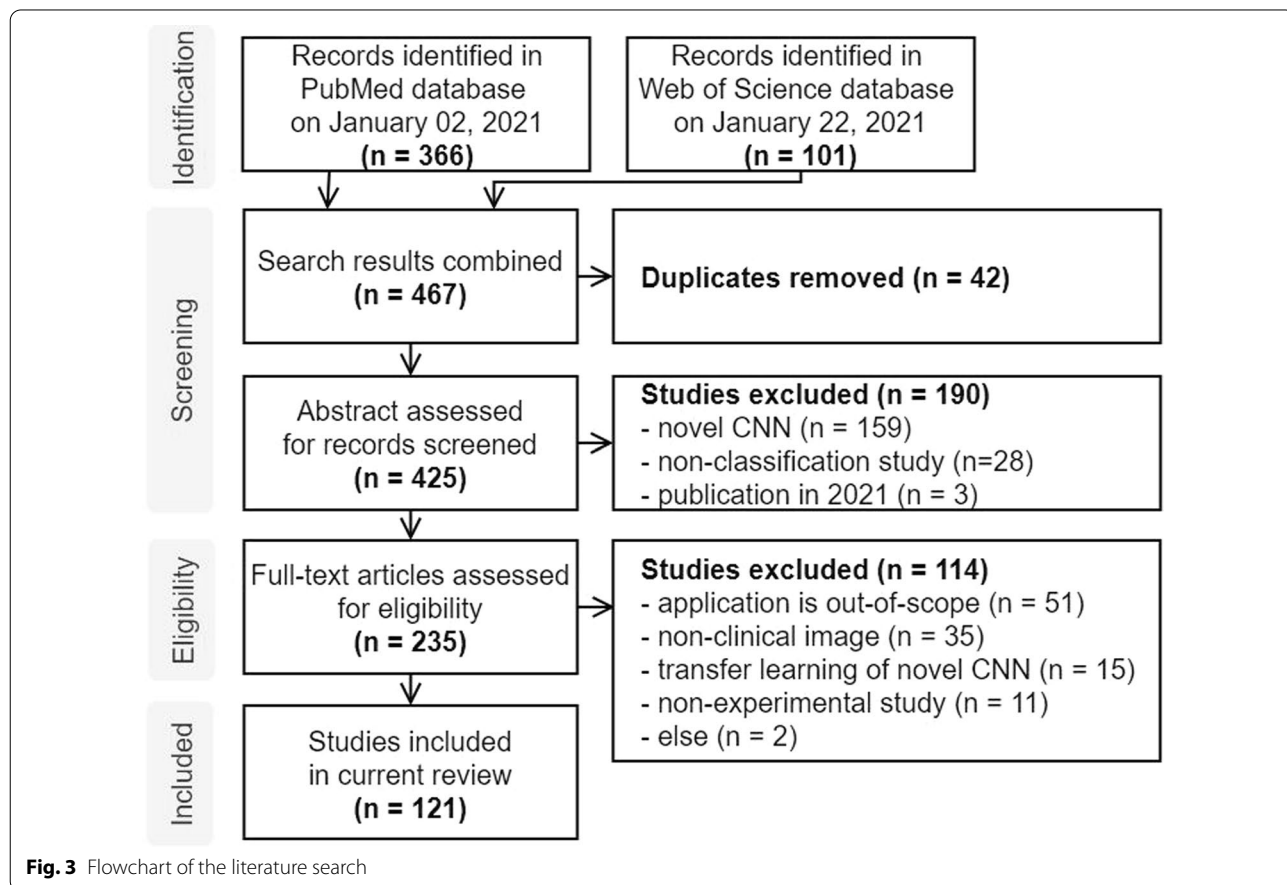
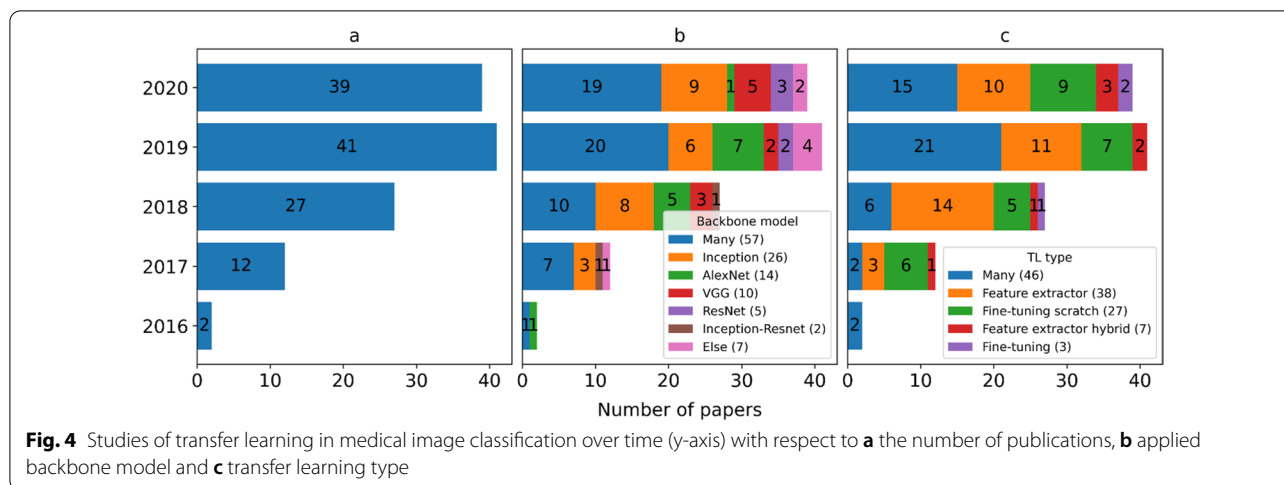


Fig. 3 Flowchart of the literature search



Besides the benchmark studies, the most prevalent model was the Inception (n=26) that consists of the least parameters shown in Table 1. AlexNet (n=14) and VGG (n=10) were the next commonly used models although they are shallower than ResNet (n=5) and Inception-Resnet (n=2). Finally, only a few studies (n=7) used a specific model such as LeNet5, DenseNet, CheXNet, DarkNet, OverFeat or CaffeNet.

Transfer learning

Similar to the backbone model, the majority of models (n=46) evaluated numerous TL approaches, which are illustrated in Fig. 4c. Many researchers aimed to search for the optimal choice of TL approach. Typically, grid search was applied. Shin and his colleagues [36] extensively evaluated three components by varying three CNN models (CifarNet, AlexNet and GoogLeNet) with three TL approaches (feature extractor, fine-tuning from scratch with and without random initialization), and the fine-tuned GoogLeNet from scratch without random initialization was identified as the best performing model.

The most popular TL approach was feature extractor (n=38) followed by fine-tuning from scratch (n=27), feature extractor hybrid (n=7) and fine-tuning (n=3). Feature extractor takes the advantage of saving computational costs by a large degree compared to the others. Likewise, the feature extractor hybrid can profit from the same advantage by removing the FC layers and adding less expansive machine learning algorithms. This is particularly beneficial for CNN models with heavy FC layers like AlexNet and VGG. Fine-tuning from scratch was the second most popular approach despite it being the most resource-expensive type because it updates the entire model. Fine-tuning is less expensive compared to the fine-tuning from scratch as it partially updates the

parameters of the convolutional layers. Additional file 2: Table 2 in Appendix B presents an overview of four TL approaches which were organized based on three dimensions: data modality, data subject and TL type.

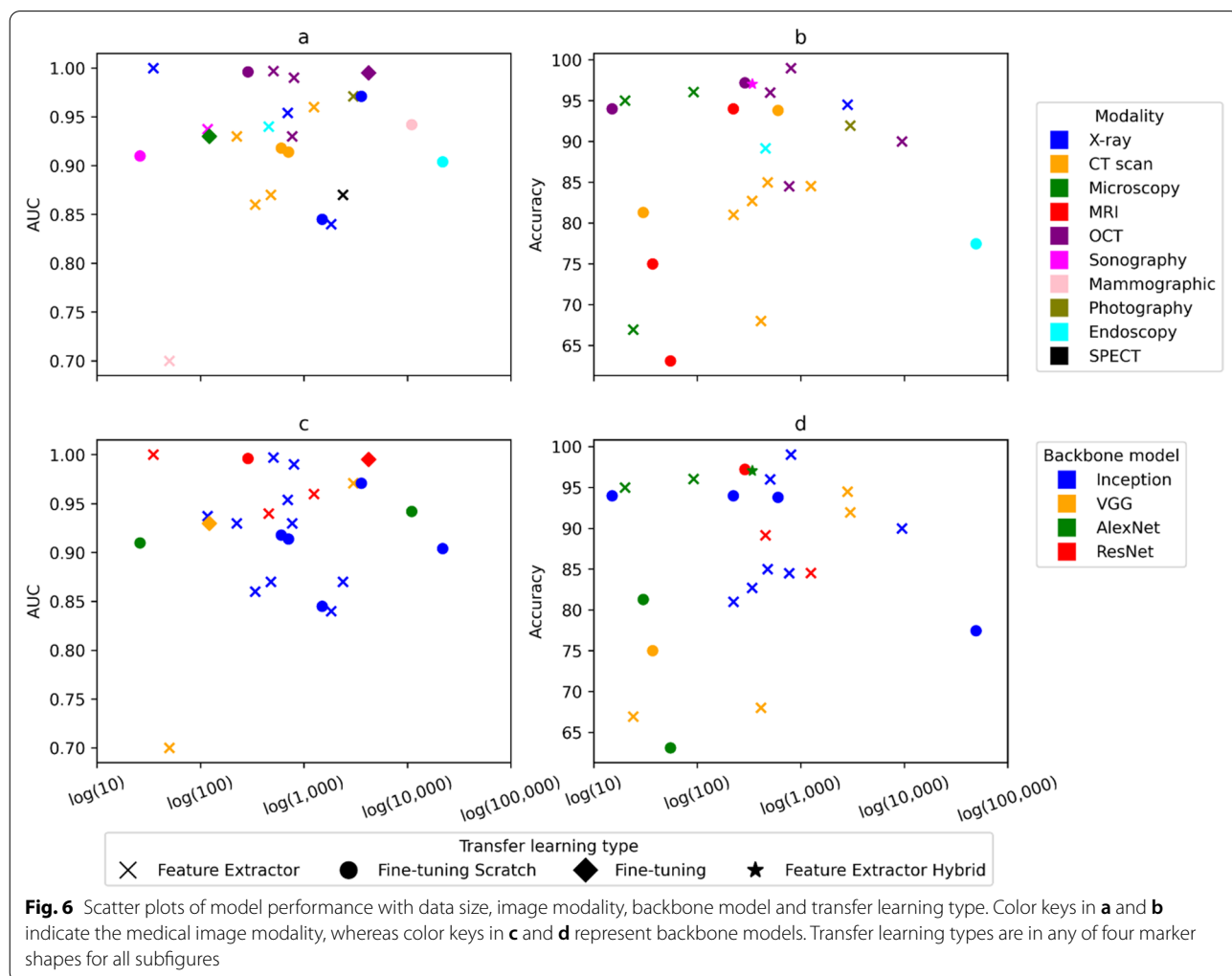
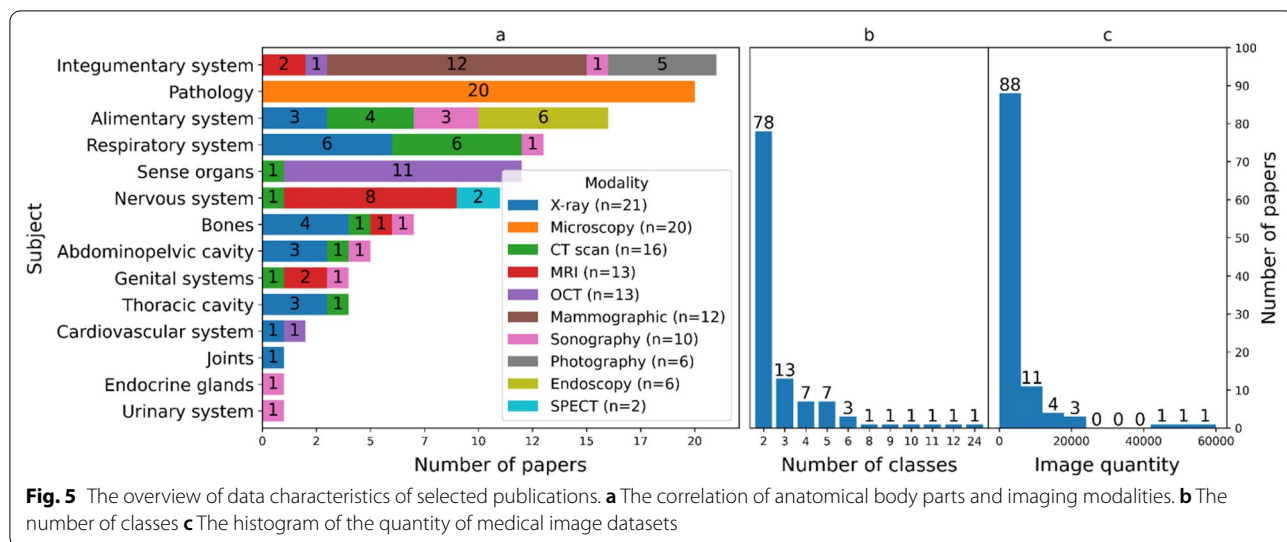
Data characteristics

As the summary of data characteristics is depicted in Fig. 5, a variety of human anatomical regions has been studied. Most of the studied regions were breast cancer exams and skin cancer lesions. Likewise, a wide variety of imaging modalities contained a unique attribute of medical image analysis. For instance, computed tomography (CT) scans and magnetic resonance imaging (MRI) are capable of generating 3D image data, while digital microscopy can generate terabytes of whole slide image (WSI) of tissue specimens.

Figure 5b shows that the majority of studies consist of binary classes, while Fig. 5c shows that the majority of studies have fallen into the first bin which ranges from 0 to 600. Minor publications are not depicted in Fig. 5 for the following reasons: the experiment was conducted with multiple subjects (human body parts); multiple tasks; multiple databases; or the subject is non-human body images (e.g. surgical tools).

Performance visualization

Figure 6 shows scatter plots of model performance, TL type and two data characteristics: data size and image modality. The Y coordinates adhere to two metrics, namely area under the receiver operating characteristic curve (AUC) and accuracy. Eleven studies used both metrics, so they are displayed on both scatter plots. The X coordinate is the normalized data quantity, otherwise it is not fair to compare the classification performance with two classes versus ten classes. The data quantities of



three modalities—CT, MRI and Microscopy—reflect the number of patients.

For the fair comparison, studies employed only a single model, TL type and image modality are depicted ($n=41$). Benchmark studies were excluded; otherwise, one study would generate several overlapping data points and potentially lead to bias. The excluded studies are either with multiple models ($n=57$), with multiple TL types ($n=14$) or with minor models like LeNet ($n=9$).

According to Spearman's rank correlation analyses, there were no relevant associations observed between the size of the data set and performance metrics. Data size and AUC (Fig. 6a, c) showed no relevant correlation ($r_{sp}=0.05$, $p=0.03$). Similarly, only a weak positive trend ($r_{sp}=0.13$, $p=0.17$) could be detected between the size of the dataset and accuracy (Fig. 6b, d). There was also no association between other variables such as modality, TL type and backbone model. For instance, the data points of models, such as feature extractors that were fitted into optical coherence tomography (OCT) images (purple crosses, Fig. 6a, b) showed that larger data quantities did not necessarily guarantee better performance. Notably, data points in cross shapes (models as feature extractors) showed decent results even though only a few fully connected layers were being retrained.

Discussion

In this survey of selected literature, we have summarized 121 research articles applying TL to medical image analysis and found that the most frequently used model was Inception. Inception is a deep model, nevertheless, it consists of the least parameters (Table 1) owing to the 1×1 filter [37]. This 1×1 filter acts as a fully connected layer in Inception and ResNet and it lowers the computational burden to a great degree [38]. To our surprise, AlexNet and VGG were the next popular models. At first glance, this result seemed counterintuitive because ResNet is a more powerful model with fewer parameters compared to AlexNet or VGG. For instance, ResNet50 achieved a top-5 error of 6.7% on ILSVRC, which was 2.6% lower than VGG16 with 5.2 times fewer parameters and 9.7% lower than AlexNet with 2.4 times fewer parameters [27]. However, this assumption is valid only if the model was fine-tuned from scratch. The number of parameters significantly drops when the model is utilized as a feature extractor as shown in Table 1. He et al. [39] performed an in-depth evaluation of the impact of various settings for refining the training of multiple backbone models, focusing primarily on the ResNet architecture. Another assumption was that AlexNet and VGG are easy to understand because the network morphology is linear and made up of stacked layers. This stands against more

complex concepts such as skip connections, bottlenecks, convolutional blocks introduced in Inception or ResNet.

With respect to TL approaches, the majority of studies empirically tested as many possible combinations of CNN models with as many as possible TL approaches. Compared to previously suggested best practices [40], some studies determined fine-tuning arbitrarily and ambiguously. For instance, [41] froze all layers except the last 12 layers without justification, while [42, 43] did not clearly describe the fine-tuning configuration. Lee et al. [44] partitioned VGG16/19 into 5 blocks, unfroze blocks sequentially and identified the model fine-tuned with two blocks that achieved the highest performance. Similarly, fine-tuned CaffeNet by unfreezing each layer sequentially [45]. The best results were obtained by the model with one retrained layer for the detection task and with two retrained layers for the classification task.

Fine-tuning from scratch ($n=27$) was a prevalent TL approach in the literature, however, we recommend using this approach carefully for two reasons: firstly, it does not improve the model performance as shown in Fig. 6 and secondly, it is the computationally most expensive choice because it updates large gradients for entire layers. Therefore, we encourage one to begin with the feature extractor approach, then incrementally fine-tune the convolutional layers. We recommend updating all layers (fine-tuning from scratch), if the feature extractor does not reflect the characteristics of the new medical images.

There was no consensus among studies concerning the global optimum configuration for fine-tuning. [46] concluded that fine-tuning the last fully connected layers of Inception3, ResNet50, and DenseNet121 outperformed fine-tuning from scratch in all cases. On the other hand, Yu et al. [47] found that retraining from scratch of DenseNet201 achieved the highest diagnostic accuracy. We speculate that one of the causes is the variety of data subjects and imaging modalities addressed in Sect. 4.3. Hence, investigating the medical data characteristics (e.g. anatomical sites, imaging modalities, data size, label size and more) and TL with CNN models would be interesting to investigate, yet it is understudied in the current literature. Morid et al. [48] stated that deep CNN models may be more effective for the following image modalities: X-ray, endoscopic and ultrasound images, while shallow CNN models may be optimal for processing these image modalities: OCT and photography for skin lesions and fundus. Nonetheless, more research is needed to further confirm these hypotheses.

TL with random initialization often appeared in the literature [49–52]. These studies used the architecture of CNN models only and initialized the training with random weights. One could argue that there is no transfer of

knowledge if the entire weights and biases are initialized, but this is still considered as TL in the literature.

It is also worth noting that only a few studies [53, 54] employed native 3D-CNN. Both studies reported that 3D-CNN outperformed 2D-CNN and 2.5-CNN models, however, Zhang et al. [53] set the number of the frames to 16 and Xiong et al. [54] reduced the resolution up to 21*21*21 voxels due to the limitation of computer resources. The majority of the studies constructed 2D-CNN or 2.5D-CNN from 3D inputs. In order to reduce the processing burden, only a sample of image slices from 3D inputs was taken. We expect that the number of studies employing 3D models will increase in the future as high-performance DL is an emerging research topic.

We confirmed (Fig. 5c) that only a limited amount of data was available in most studies for medical image analysis. Many studies took advantage of using publicly accessible medical datasets from grand challenges (<https://grand-challenge.org/challenges>). This is a particularly beneficial scientific practice because novel solutions are shared online allowing for better reproducibility. We summarized 78 publicly available medical datasets in Additional file 3: Suppl. Table 3 (Appendix C), which were organized based on the following five attributes: data modality, anatomical part/region, task type, data name, published year and the link.

Although most evaluated papers included only brief information about their hardware setup, no details were provided about training or test time performance. As most medical data sets are small, usually consumer-grade GPUs in custom workstations or seldom server-grade GPUs (P100 or V100) were sufficient for TL. Previous survey studies have investigated how DL can be optimized and sped up on GPUs [55] or by using specifically designed hardware accelerators like field-programmable gate arrays (FPGA) for neural network inference [56]. We could not investigate these aspects of efficient TL because execution time was rarely reported in the surveyed literature.

This study is limited to surveying only TL for medical image classification. However, many interesting task-oriented TL studies were published in the past few years, with a particular focus on object detection and image segmentation [57], as reflected by the amount of public data sets (see also Additional file 3: Appendix C., Table 3). We only investigated off-the-shelf CNN models pretrained on ImageNet and intentionally left out custom CNN architectures, although these can potentially outperform TL-based models on certain tasks [58, 59]. Also, we did not evaluate aspects of potential model improvements leveraged by the differences of the source- and the target domain of the training data

used for TL [60]. Similarly, we did not evaluate vision transformers (ViT) [61], which are emerging for image data analysis. For instance, Liu et al. [62] compared 22 backbone models and four ViT models and concluded that one of the ViT models exhibited the highest accuracy trained on cropped cytopathology cell images. Recently, Chen et al. [63] proposed a novel architecture that is a parallel design of MobileNet and ViT, in view of achieving not only more efficient computation but also better model performance.

Conclusion

We aimed to provide actionable insights to the readers and ML practitioners, on how to select backbone CNN models and tune them properly with consideration of medical data characteristics. While we encourage readers to methodically search for the optimal choice of model and TL setup, it is a good starting point to employ deep CNN models (preferably ResNet or Inception) as feature extractors. We recommend updating only the last fully connected layers of the chosen model on the medical image dataset. In case the model performance needs to be refined, the model should be fine-tuned by incrementally unfreezing convolutional layers from top to bottom layers with a low learning rate. Following these basic steps can save computational costs and time without degrading the predictive power. Finally, publicly accessible medical image datasets were compiled in a structured table describing the modality, anatomical region, task type and publication year as well as the URL for accession.

Abbreviations

AUC: Area under the receiver operating characteristic curve; CT: Computed tomography; CNN: Convolutional neural networks; DL: Deep learning; FC: Fully connected; FPGA: Field-programmable gate arrays; GPU: Graphics processing unit; HOG: Histograms of oriented gradients; ILSVRC: ImageNet large scale visual recognition challenge; LBP: Local binary pattern; MRI: Magnetic resonance imaging; OCT: Optical coherence tomography; TL: Transfer learning; TPU: Tensor processing unit; ViT: Vision transformer; WSI: Whole slide image.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12880-022-00793-7>.

Additional file 1. Search terms.

Additional file 2. Summary table of studies.

Additional file 3. Summary table of public medical datasets.

Acknowledgements

The authors would like to thank Joseph Babcock (Catholic University of Paris) and Jonathan Griffiths (Academic Writing Support Center, Heidelberg University) for proofreading and Fabian Siegel MD and Frederik Trinkmann MD (Medical Faculty Mannheim, Heidelberg University) for comments on the manuscript. We would like to thank the reviewer for their constructive feedback.

Author contributions

H.E.K. conceptualized the study. H.E.K. and A.C.L. created the search query and article collection. A.C.L., N.S., M.J., M.E.M. and H.K. screened and evaluated the selected papers. H.E.K. analyzed the data and created figures. H.E.K., M.E.M. and T.G. interpreted the data. M.E.M. advised technical aspects of the study. H.E.K., M.E.M. and T.G. wrote the manuscript. M.E.M. and T.G. supervised the study. All authors critically reviewed the manuscript and approved the final version.

Funding

Open Access funding enabled and organized by Projekt DEAL. A.C.L., N.S., M.E.M. and T.G. were supported by funding from the German Ministry for Education and Research (BMBF) within the framework of the Medical Informatics Initiative (MIRACUM Consortium: Medical Informatics for Research and Care in University Medicine; 01ZZ1801E).

Availability of data and materials

The dataset analyzed in this study are shown in Appendix B. In-depth information is available on reasonable request from the corresponding author (HeeEun.Kim@medma.uni-heidelberg.de).

Declarations

Ethics approval and consent to participate

Not applicable. This manuscript is exempt from ethics approval because it does not use any animal or human subject data or tissue.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no conflict of interest.

Author details

¹Department of Biomedical Informatics at the Center for Preventive Medicine and Digital Health (CPD-BW), Medical Faculty Mannheim, Heidelberg University, Theodor-Kutzer-Ufer 1-3, 68167 Mannheim, Germany. ²Chair of Medical Informatics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Wetterkreuz 15, 91058 Erlangen, Germany.

Received: 25 August 2021 Accepted: 30 March 2022

Published online: 13 April 2022

References

- Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). IEEE; 2005. pp. 886–93.
- He D-C, Wang L. Texture unit, texture spectrum, and texture analysis. *IEEE Trans Geosci Remote Sens.* 1990;28:509–12.
- CAMELYON17—Grand Challenge. [grand-challenge.org. https://camelyon17.grand-challenge.org/evaluation/challenge/leaderboard/](https://camelyon17.grand-challenge.org/evaluation/challenge/leaderboard/). Accessed 3 Apr 2021.
- Shi B, Grimm LJ, Mazurowski MA, Baker JA, Marks JR, King LM, et al. Prediction of occult invasive disease in ductal carcinoma in situ using deep learning features. *J Am Coll Radiol.* 2018;15(3 Pt B):527–34.
- Wang Z, Du B, Guo Y. Domain adaptation with neural embedding matching. *IEEE Trans Neural Netw Learn Syst.* 2019;31:2387–97.
- Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng.* 2010;22:1345–59.
- Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. *J Big data.* 2016;3:1–40.
- Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, et al. A comprehensive survey on transfer learning. *Proc IEEE.* 2020;109:43–76.
- Wilson G, Cook DJ. A survey of unsupervised deep domain adaptation. *ACM Trans Intell Syst Technol (TIST).* 2020;11:1–46.
- Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: *Advances in neural information processing systems.* 2014;27.
- Zhu J-Y, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE international conference on computer vision.* 2017. pp. 2223–32.
- Zhang T, Cheng J, Fu H, Gu Z, Xiao Y, Zhou K, et al. Noise adaptation generative adversarial network for medical image analysis. *IEEE Trans Med Imaging.* 2019;39:1149–59.
- Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, et al. Domain-adversarial training of neural networks. *J Mach Learn Res.* 2016;17:2096–2030.
- Wang Z, Du B, Tu W, Zhang L, Tao D. Incorporating distribution matching into uncertainty for multiple kernel active learning. *IEEE Trans Knowl Data Eng.* 2019;33:128–42.
- Zhang Y, Wei Y, Wu Q, Zhao P, Niu S, Huang J, et al. Collaborative unsupervised domain adaptation for medical image diagnosis. *IEEE Trans Image Process.* 2020;29:7834–44.
- Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciampi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017;42:60–88.
- Chowdhury A, Rosenthal J, Waring J, Umeton R. Applying self-supervised learning to medicine: review of the state of the art and medical implementations. In: *Informatics. Multidisciplinary Digital Publishing Institute;* 2021. p. 59.
- Zhang J, Li C, Rahaman MM, Yao Y, Ma P, Zhang J, et al. A comprehensive review of image analysis methods for microorganism counting: from classical image processing to deep learning approaches. *Artif Intell Rev.* 2021;1–70.
- Rahaman MM, Li C, Wu X, Yao Y, Hu Z, Jiang T, et al. A survey for cervical cytopathology image analysis using deep learning. *IEEE Access.* 2020;8:61687–710.
- Agarwal D, Marques G, de la Torre-Diez I, Franco Martin MA, García Zapirain B, Martín RF. Transfer learning for Alzheimer's disease through neuroimaging biomarkers: a systematic review. *Sensors.* 2021;21:7259.
- Valverde JM, Imani V, Abdollahzadeh A, De Feo R, Prakash M, Ciszek R, et al. Transfer learning in magnetic resonance brain imaging: a systematic review. *J Imaging.* 2021;7:66.
- ImageNet. <https://www.image-net.org/update-mar-11-2021.php>. Accessed 18 May 2021.
- Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE.* 1998;86:2278–324.
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ, editors. *Advances in neural information processing systems 25.* Curran Associates, Inc.; 2012. p. 1097–105.
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv:14091556 [cs].* 2015.
- Hegde RB, Prasad K, Hebbar H, Singh BMK. Feature extraction using traditional image processing and convolutional neural network methods to classify white blood cells: a study. *Australas Phys Eng Sci Med.* 2019;42:627–38.
- He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR). Las Vegas, NV, USA: IEEE; 2016. p. 770–8.
- Rahaman MM, Li C, Yao Y, Kulwa F, Rahman MA, Wang Q, et al. Identification of COVID-19 samples from chest X-Ray images using deep learning: a comparison of transfer learning approaches. *XST.* 2020;28:821–39.
- Burdick J, Marques O, Weinthal J, Furt B. Rethinking skin lesion segmentation in a convolutional classifier. *J Digit Imaging.* 2018;31:435–40.
- Chen Q, Hu S, Long P, Lu F, Shi Y, Li Y. A transfer learning approach for malignant prostate lesion detection on multiparametric MRI. *Technol Cancer Res Treat.* 2019;18:1533033819858363.
- Lakhani P. Deep convolutional neural networks for endotracheal tube position and X-ray image classification: challenges and opportunities. *J Digit Imaging.* 2017;30:460–8.
- Yang H, Zhang J, Liu Q, Wang Y. Multimodal MRI-based classification of migraine: using deep learning convolutional neural network. *Biomed Eng Online.* 2018;17:138.
- Yu S, Liu L, Wang Z, Dai G, Xie Y. Transferring deep neural networks for the differentiation of mammographic breast lesions. *Sci China Technol Sci.* 2019;62:441–7.

34. Ovalle-Magallanes E, Avina-Cervantes JG, Cruz-Aceves I, Ruiz-Pinales J. Transfer learning for stenosis detection in X-ray coronary angiography. *Mathematics*. 2020;8:1510.
35. Taló M, Yildirim O, Baloglu UB, Aydın G, Acharya UR. Convolutional neural networks for multi-class brain disease detection using MRI images. *Comput Med Imaging Graph*. 2019;78:101673.
36. Shin H-C, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging*. 2016;35:1285–98.
37. Lin M, Chen Q, Yan S. Network in network. arXiv:13124400 [cs]. 2014.
38. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. arXiv:14094842 [cs]. 2014.
39. He T, Zhang Z, Zhang H, Zhang Z, Xie J, Li M. Bag of tricks for image classification with convolutional neural networks. arXiv:181201187 [cs]. 2018.
40. Chollet F. Deep learning with Python. Simon and Schuster; 2021.
41. Hemelings R, Elen B, Barbosa-Breda J, Lemmens S, Meire M, Pourjavan S, et al. Accurate prediction of glaucoma from colour fundus images with a convolutional neural network that relies on active and transfer learning. *Acta Ophthalmol*. 2020;98:e94–100.
42. Valkonen M, Isola J, Ilinen O, Muhonen V, Saxlin A, Tolonen T, et al. Cytokeratin-supervised deep learning for automatic recognition of epithelial cells in breast cancers stained for ER, PR, and Ki-67. *IEEE Trans Med Imaging*. 2019;39:534–42.
43. Han SS, Park GH, Lim W, Kim MS, Na JI, Park I, et al. 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*. 2018;13:e0191493.
44. Lee K-S, Kim JY, Jeon E, Choi WS, Kim NH, Lee KY. Evaluation of scalability and degree of fine-tuning of deep convolutional neural networks for COVID-19 screening on chest X-ray images using explainable deep learning algorithm. *J Person Med*. 2020;10:213.
45. Zhang R, Zheng Y, Mak TWC, Yu R, Wong SH, Lau JY, et al. Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain. *IEEE J Biomed Health Inform*. 2016;21:41–7.
46. Singh V, Danda V, Gorniak R, Flanders A, Lakhani P. Assessment of critical feeding tube malpositions on radiographs using deep learning. *J Digit Imaging*. 2019;32:651–5.
47. Yu X, Zeng N, Liu S, Zhang Y-D. Utilization of DenseNet201 for diagnosis of breast abnormality. *Mach Vis Appl*. 2019;30:1135–44.
48. Morid MA, Borjali A, Del Fiol G. A scoping review of transfer learning research on medical image analysis using ImageNet. arXiv:200413175 [cs, eess]. 2020. <https://doi.org/10.1016/j.combiomed.2020.104115>.
49. Karri SPK, Chakraborty D, Chatterjee J. Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration. *Biomed Opt Express*. 2017;8:579–92.
50. Kim Y-G, Kim S, Cho CE, Song IH, Lee HJ, Ahn S, et al. Effectiveness of transfer learning for enhancing tumor classification with a convolutional neural network on frozen sections. *Sci Rep*. 2020;10:21899.
51. Lee H, Tajmir S, Lee J, Zissen M, Yeshiwas BA, Alkasab TK, et al. Fully automated deep learning system for bone age assessment. *J Digit Imaging*. 2017;30:427–41.
52. Tang Y-X, Tang Y-B, Peng Y, Yan K, Bagheri M, Redd BA, et al. Automated abnormality classification of chest radiographs using deep convolutional neural networks. *NPJ Digit Med*. 2020;3:70.
53. Zhang X, Zhang Y, Han EY, Jacobs N, Han Q, Wang X, et al. Classification of whole mammogram and tomosynthesis images using deep convolutional neural networks. *IEEE Trans Nanobiosci*. 2018;17:237–42.
54. Xiong J, Li X, Lu L, Schwartz LH, Fu X, Zhao J, et al. Implementation strategy of a CNN model affects the performance of CT assessment of EGFR mutation status in lung cancer patients. *IEEE Access*. 2019;7:64583–91.
55. Mittal S, Vaishay S. A survey of techniques for optimizing deep learning on GPUs. *J Syst Archit*. 2019;99:101635.
56. Guo K, Zeng S, Yu J, Wang Y, Yang H. A survey of FPGA-based neural network accelerator. arXiv:171208934 [cs]. 2018.
57. Sun C, Li C, Zhang J, Rahaman MM, Ai S, Chen H, et al. Gastric histopathology image segmentation using a hierarchical conditional random field. *Biocybern Biomed Eng*. 2020;40:1535–55.
58. Rahaman MM, Li C, Yao Y, Kulwa F, Wu X, Li X, et al. DeepCervix: a deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques. arXiv preprint arXiv:210212191. 2021.
59. Alzubaidi L, Al-Amidie M, Al-Asadi A, Humaidi AJ, Al-Shamma O, Fadhel MA, et al. Novel transfer learning approach for medical imaging with limited labeled data. *Cancers*. 2021;13:1590.
60. Alzubaidi L, Fadhel MA, Al-Shamma O, Zhang J, Santamaría J, Duan Y, et al. Towards a better understanding of transfer learning for medical imaging: a case study. *Appl Sci*. 2020;10:4523.
61. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: transformers for image recognition at scale. arXiv:201011929 [cs]. 2021.
62. Liu W, Li C, Rahamana MM, Jiang T, Sun H, Wu X, et al. Is the aspect ratio of cells important in deep learning? A robust comparison of deep learning methods for multi-scale cytopathology cell image classification: from convolutional neural networks to visual transformers. arXiv:210507402 [cs]. 2021.
63. Chen Y, Dai X, Chen D, Liu M, Dong X, Yuan L, et al. Mobile-former: bridging mobilenet and transformer. arXiv:210805895 [cs]. 2021.
64. Huang J, Habib A-R, Mendis D, Chong J, Smith M, Duvnjak M, et al. An artificial intelligence algorithm that differentiates anterior ethmoidal artery location on sinus computed tomography scans. *J Laryngol Otol*. 2020;134:52–5.
65. Yamada A, Oyama K, Fujita S, Yoshizawa E, Ichinohe F, Komatsu D, et al. Dynamic contrast-enhanced computed tomography diagnosis of primary liver cancers using transfer learning of pretrained convolutional neural networks: is registration of multiphase images necessary? *Int J CARS*. 2019;14:1295–301.
66. Peng J, Kang S, Ning Z, Deng H, Shen J, Xu Y, et al. Residual convolutional neural network for predicting response of transarterial chemoembolization in hepatocellular carcinoma from CT imaging. *Eur Radiol*. 2020;30:413–24.
67. Hadj Said M, Le Roux M-K, Catherine J-H, Lan R. Development of an artificial intelligence model to identify a dental implant from a radiograph. *Int J Oral Maxillofac Implants*. 2020;35.
68. Lee J-H, Kim D-H, Jeong S-N. Diagnosis of cystic lesions using panoramic and cone beam computed tomographic images based on deep learning neural network. *Oral Dis*. 2020;26:152–8.
69. Parmar P, Habib AR, Mendis D, Daniel A, Duvnjak M, Ho J, et al. An artificial intelligence algorithm that identifies middle turbinate pneumatization (concha bullosa) on sinus computed tomography scans. *J Laryngol Otol*. 2020;134:328–31.
70. Kajikawa T, Kadoya N, Ito K, Takayama Y, Chiba T, Tomori S, et al. Automated prediction of dosimetric eligibility of patients with prostate cancer undergoing intensity-modulated radiation therapy using a convolutional neural network. *Radiol Phys Technol*. 2018;11:320–7.
71. Dawud AM, Yurtkan K, Oztoprak H. Application of deep learning in neuroradiology: brain haemorrhage classification using transfer learning. *Comput Intell Neurosci*. 2019;2019.
72. Zhao X, Qi S, Zhang B, Ma H, Qian W, Yao Y, et al. Deep CNN models for pulmonary nodule classification: model modification, model integration, and transfer learning. *J Xray Sci Technol*. 2019;27:615–29.
73. da Nobrega RVM, Rebouças Filho PP, Rodrigues MB, da Silva SP, Junior CMD, de Albuquerque VHC. Lung nodule malignancy classification in chest computed tomography images using transfer learning and convolutional neural networks. *Neural Comput Appl*. 2018;1–18.
74. Zhang S, Sun F, Wang N, Zhang C, Yu Q, Zhang M, et al. Computer-aided diagnosis (CAD) of pulmonary nodule of thoracic CT image using transfer learning. *J Digit Imaging*. 2019;32:995–1007.
75. Nibali A, He Z, Wollersheim D. Pulmonary nodule classification with deep residual networks. *Int J Comput Assist Radiol Surg*. 2017;12:1799–808.
76. Pham TD. A comprehensive study on classification of COVID-19 on computed tomography with pretrained convolutional neural networks. *Sci Rep*. 2020;10:16942.

77. Gao J, Jiang Q, Zhou B, Chen D. Lung nodule detection using convolutional neural networks with transfer learning on CT images. *Combinatorial Chemistry & High Throughput Screening*. 2020.
78. Chowdhury NI, Smith TL, Chandra RK, Turner JH. Automated classification of osteomeatal complex inflammation on computed tomography using convolutional neural networks. In: *International forum of allergy & rhinology*. Wiley Online Library; 2019. pp. 46–52.
79. Nishio M, Sugiyama O, Yakami M, Ueno S, Kubo T, Kuroda T, et al. Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning. *PLoS ONE*. 2018;13:e0200721.
80. Zachariah R, Samaraseena J, Luba D, Duh E, Dao T, Requa J, et al. Prediction of polyp pathology using convolutional neural networks achieves “resect and discard” thresholds. *Am J Gastroenterol*. 2020;115:138–44.
81. Zhu Y, Wang Q-C, Xu M-D, Zhang Z, Cheng J, Zhong Y-S, et al. Application of convolutional neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy. *Gastrointest Endosc*. 2019;89:806–815.e1.
82. Cho B-J, Bang CS, Park SW, Yang YJ, Seo SI, Lim H, et al. Automated classification of gastric neoplasms in endoscopic images using a convolutional neural network. *Endoscopy*. 2019;51:1121–9.
83. Shichijo S, Nomura S, Aoyama K, Nishikawa Y, Miura M, Shinagawa T, et al. Application of convolutional neural networks in the diagnosis of helicobacter pylori infection based on endoscopic images. *EBioMedicine*. 2017;25:106–11.
84. Shichijo S, Endo Y, Aoyama K, Takeuchi Y, Ozawa T, Takiyama H, et al. Application of convolutional neural networks for evaluating helicobacter pylori infection status on the basis of endoscopic images. *Scand J Gastroenterol*. 2019;54:158–63.
85. Patrini I, Ruperti M, Moccia S, Mattos LS, Frontoni E, De Momi E. Transfer learning for informative-frame selection in laryngoscopic videos through learned features. *Med Boil Eng Comput*. 2020;1–14.
86. Samala RK, Chan H-P, Hadjiiski L, Helvie MA. Risks of feature leakage and sample size dependencies in deep feature extraction for breast mass classification. *Med Phys*. 2020.
87. Mohamed AA, Berg WA, Peng H, Luo Y, Jankowitz RC, Wu S. A deep learning method for classifying mammographic breast density categories. *Med Phys*. 2018;45:314–21.
88. Perek S, Kiryati N, Zimmerman-Moreno G, Sklair-Levy M, Konen E, Mayer A. Classification of contrast-enhanced spectral mammography (CESM) images. *Int J Comput Assist Radiol Surg*. 2019;14:249–57.
89. Samala RK, Chan H-P, Hadjiiski L, Helvie MA, Richter CD, Cha KH. Breast cancer diagnosis in digital breast tomosynthesis: effects of training sample size on multi-stage transfer learning using deep neural nets. *IEEE Trans Med Imaging*. 2018;38:686–96.
90. Huynh BQ, Li H, Giger ML. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *J Med Imaging*. 2016;3:034501.
91. Chougrad H, Zouaki H, Alheyane O. Deep convolutional neural networks for breast cancer screening. *Comput Methods Programs Biomed*. 2018;157:19–30.
92. Samala RK, Chan H-P, Hadjiiski LM, Helvie MA, Richter CD. Generalization error analysis for deep convolutional neural network with transfer learning in breast cancer diagnosis. *Phys Med Biol*. 2020;65:105002.
93. Shen L, Margolies LR, Rothstein JH, Fluder E, McBride R, Sieh W. Deep learning to improve breast cancer detection on screening mammography. *Sci Rep*. 2019;9:1–12.
94. Shafiqe S, Tehsin S. Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks. *Technol Cancer Res Treat*. 2018;17:1533033818802789.
95. Yu Y, Wang J, Ng CW, Ma Y, Mo S, Fong ELS, et al. Deep learning enables automated scoring of liver fibrosis stages. *Sci Rep*. 2018;8:16016.
96. Huttunen MJ, Hassan A, McCloskey CW, Fasih S, Upham J, Vanderhyden BC, et al. Automated classification of multiphoton microscopy images of ovarian tissue using deep learning. *J Biomed Opt*. 2018;23:066002.
97. Talo M. Automated classification of histopathology images using transfer learning. *Artif Intell Med*. 2019;101:101743.
98. Mazo C, Bernal J, Trujillo M, Alegre E. Transfer learning for classification of cardiovascular tissues in histological images. *Comput Methods Programs Biomed*. 2018;165:69–76.
99. Riordon J, McCallum C, Sinton D. Deep learning for the classification of human sperm. *Comput Biol Med*. 2019;111:103342.
100. Marsh JN, Matlock MK, Kudose S, Liu T-C, Stappenbeck TS, Gaut JP, et al. Deep learning global glomerulosclerosis in transplant kidney frozen sections. *IEEE Trans Med Imaging*. 2018;37:2718–28.
101. Kanavati F, Toyokawa G, Momosaki S, Rambeau M, Kozuma Y, Shoji F, et al. Weakly-supervised learning for lung carcinoma classification using deep learning. *Sci Rep*. 2020;10:9297.
102. Kather JN, Krisam J, Charoentong P, Luedde T, Herpel E, Weis C-A, et al. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLOS Med*. 2019;16:e1002730.
103. He Y, Guo J, Ding X, van Ooijen PM, Zhang Y, Chen A, et al. Convolutional neural network to predict the local recurrence of giant cell tumor of bone after curettage based on pre-surgery magnetic resonance images. *Eur Radiol*. 2019;29:5441–51.
104. Yuan Y, Qin W, Buyyounouski M, Ibragimov B, Hancock S, Han B, et al. Prostate cancer classification with multiparametric MRI transfer learning model. *Med Phys*. 2019;46:756–65.
105. Borkowski K, Rossi C, Ciritsis A, Marcon M, Hejduk P, Stieb S, et al. Fully automatic classification of breast MRI background parenchymal enhancement using a transfer learning approach. *Medicine (Baltimore)*. 2020;99.
106. Zhu Z, Harowicz M, Zhang J, Saha A, Grimm LJ, Hwang ES, et al. Deep learning analysis of breast MRIs for prediction of occult invasive disease in ductal carcinoma in situ. *Comput Biol Med*. 2019;115:103498.
107. Fukuma R, Yanagisawa T, Kinoshita M, Shinozaki T, Arita H, Kawaguchi A, et al. Prediction of IDH and TERT promoter mutations in low-grade glioma from magnetic resonance images using a convolutional neural network. *Sci Rep*. 2019;9:1–8.
108. Banzato T, Causin F, Della Puppa A, Cester G, Mazzali L, Zotti A. Accuracy of deep learning to differentiate the histopathological grading of meningiomas on MR images: a preliminary study. *J Magn Reson Imaging*. 2019;50:1152–9.
109. Swati ZNK, Zhao Q, Kabir M, Ali F, Ali Z, Ahmed S, et al. Brain tumor classification for MR images using transfer learning and fine-tuning. *Comput Med Imaging Graph*. 2019;75:34–46.
110. Yang Y, Yan L-F, Zhang X, Han Y, Nan H-Y, Hu Y-C, et al. Glioma grading on conventional MR images: a deep learning study with transfer learning. *Front Neurosci*. 2018;12:804.
111. Deepak S, Ameer PM. Brain tumor classification using deep CNN features via transfer learning. *Comput Biol Med*. 2019;111:103345.
112. Singla N, Dubey K, Srivastava V. Automated assessment of breast cancer margin in optical coherence tomography images via pretrained convolutional neural network. *J Biophoton*. 2019;12:e201800255.
113. Gessert N, Lutz M, Heyder M, Latus S, Leistner DM, Abdelwahed YS, et al. Automatic plaque detection in IVOC pullbacks using convolutional neural networks. *IEEE Trans Med Imaging*. 2018;38:426–34.
114. Ahn JM, Kim S, Ahn K-S, Cho S-H, Lee KB, Kim US. A deep learning model for the detection of both advanced and early glaucoma using fundus photography. *PLoS ONE*. 2018;13:e0207982.
115. Treder M, Laueremann JL, Eter N. Automated detection of exudative age-related macular degeneration in spectral domain optical coherence tomography using deep learning. *Graefes Arch Clin Exp Ophthalmol*. 2018;256:259–65.
116. Zheng C, Xie X, Huang L, Chen B, Yang J, Lu J, et al. Detecting glaucoma based on spectral domain optical coherence tomography imaging of peripapillary retinal nerve fiber layer: a comparison study between hand-crafted features and deep learning model. *Graefes Arch Clin Exp Ophthalmol*. 2020;258:577–85.
117. Zago GT, Andreão RV, Dorizzi B, Salles EOT. Retinal image quality assessment using deep learning. *Comput Biol Med*. 2018;103:64–70.
118. Burlina P, Pacheco KD, Joshi N, Freund DE, Bressler NM. Comparing humans and deep learning performance for grading AMD: a study in using universal deep features and transfer learning for automated AMD analysis. *Comput Biol Med*. 2017;82:80–6.
119. Liu TA, Ting DS, Paul HY, Wei J, Zhu H, Subramanian PS, et al. Deep learning and transfer learning for optic disc laterality detection: Implications for machine learning in neuro-ophthalmology. *J Neuroophthalmol*. 2020;40:178–84.

120. Choi JY, Yoo TK, Seo JG, Kwak J, Um TT, Rim TH. Multi-categorical deep learning neural network to classify retinal images: a pilot study employing small database. *PLoS ONE*. 2017;12:e0187336.
121. Gómez-Valverde JJ, Antón A, Liefers B, Herranz A, Santos A, et al. Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning. *Biomed Opt Express*. 2019;10:892–913.
122. Xu BY, Chiang M, Chaudhary S, Kulkarni S, Pardeshi AA, Varma R. Deep learning classifiers for automated detection of gonioscopic angle closure based on anterior segment OCT images. *Am J Ophthalmol*. 2019;208:273–80.
123. Shen X, Zhang J, Yan C, Zhou H. An automatic diagnosis method of facial acne vulgaris based on convolutional neural network. *Sci Rep*. 2018;8:1–10.
124. Cirillo MD, Mirdell R, Sjöberg F, Pham TD. Time-independent prediction of burn depth using deep convolutional neural networks. *J Burn Care Res*. 2019;40:857–63.
125. Huang K, He X, Jin Z, Wu L, Zhao X, Wu Z, et al. Assistant diagnosis of basal cell carcinoma and seborrheic keratosis in chinese population using convolutional neural network. *J Healthcare Eng*. 2020;2020.
126. Sun Y, Shan C, Tan T, Tong T, Wang W, Pourtaherian A. Detecting discomfort in infants through facial expressions. *Physiol Meas*. 2019;40:115006.
127. Cheng PM, Malhi HS. Transfer learning with convolutional neural networks for classification of abdominal ultrasound images. *J Digit Imaging*. 2017;30:234–43.
128. Xue L-Y, Jiang Z-Y, Fu T-T, Wang Q-M, Zhu Y-L, Dai M, et al. Transfer learning radiomics based on multimodal ultrasound imaging for staging liver fibrosis. *Eur Radiol*. 2020;1–11.
129. Byra M, Styczynski G, Szmigielski C, Kalinowski P, Michalowski L, Paluszkiwicz R, et al. Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images. *Int J Comput Assisted Radiol Surg*. 2018;13:1895–903.
130. Banzato T, Bonsembiante F, Aresu L, Gelain ME, Burti S, Zotti A. Use of transfer learning to detect diffuse degenerative hepatic diseases from ultrasound images in dogs: a methodological study. *Vet J*. 2018;233:35–40.
131. Hetherington J, Lessoway V, Gunka V, Abolmaesumi P, Rohling R. SLIDE: automatic spine level identification system using a deep convolutional neural network. *Int J CARS*. 2017;12:1189–98.
132. Chi J, Walia E, Babyn P, Wang J, Groot G, Eramian M. Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. *J Digit Imaging*. 2017;30:477–86.
133. Sridar P, Kumar A, Quinton A, Nanan R, Kim J, Krishnakumar R. Decision fusion-based fetal ultrasound image plane classification using convolutional neural networks. *Ultrasound Med Biol*. 2019;45:1259–73.
134. Byra M, Galperin M, Ojeda-Fournier H, Olson L, O'Boyle M, Comstock C, et al. Breast mass classification in sonography with transfer learning using a deep convolutional neural network and color conversion. *Med Phys*. 2019;46:746–55.
135. Chen C-H, Lee Y-W, Huang Y-S, Lan W-R, Chang R-F, Tu C-Y, et al. Computer-aided diagnosis of endobronchial ultrasound images using convolutional neural network. *Comput Methods Programs Biomed*. 2019;177:175–82.
136. Zheng Q, Furth SL, Tasian GE, Fan Y. Computer-aided diagnosis of congenital abnormalities of the kidney and urinary tract in children based on ultrasound imaging data by integrating texture image features and deep transfer learning image features. *J Pediatr Urol*. 2019;15:75–e1.
137. Kim DH, Wit H, Thurston M. Artificial intelligence in the diagnosis of Parkinson's disease from ioflupane-123 single-photon emission computed tomography dopamine transporter scans using transfer learning. *Nucl Med Commun*. 2018;39:887–93.
138. Papathanasiou ND, Spyridonidis T, Apostolopoulos DJ. Automatic characterization of myocardial perfusion imaging polar maps employing deep learning and data augmentation. *Hell J Nucl Med*. 2020;23:125–32.
139. Cheng PM, Tejura TK, Tran KN, Whang G. Detection of high-grade small bowel obstruction on conventional radiography with convolutional neural networks. *Abdom Radiol*. 2018;43:1120–7.
140. Devnath L, Luo S, Summons P, Wang D. Automated detection of pneumoconiosis with multilevel deep features learned from chest X-Ray radiographs. *Comput Biol Med*. 2021;129:104125.
141. Kim J-E, Nam N-E, Shim J-S, Jung Y-H, Cho B-H, Hwang JJ. Transfer learning via deep neural networks for implant fixture system classification using periapical radiographs. *J Clin Med*. 2020;9:1117.
142. Lee J-H, Kim D-H, Jeong S-N, Choi S-H. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J Dent*. 2018;77:106–11.
143. Lee J-H, Jeong S-N. Efficacy of deep convolutional neural network algorithm for the identification and classification of dental implant systems, using panoramic and periapical radiographs: a pilot study. *Medicine*. 2020;99.
144. Paul HY, Kim TK, Wei J, Shin J, Hui FK, Sair HI, et al. Automated semantic labeling of pediatric musculoskeletal radiographs using deep learning. *Pediatr Radiol*. 2019;49:1066–70.
145. Kim DH, MacKinnon T. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. *Clin Radiol*. 2018;73:439–45.
146. Cheng C-T, Ho T-Y, Lee T-Y, Chang C-C, Chou C-C, Chen C-C, et al. Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs. *Eur Radiol*. 2019;29:5469–77.
147. Abidin AZ, Deng B, DSouza AM, Nagarajan MB, Coan P, Wismüller A. Deep transfer learning for characterizing chondrocyte patterns in phase contrast X-Ray computed tomography images of the human patellar cartilage. *Comput Biol Med*. 2018;95:24–33.
148. Heidari M, Mirniaharikandehei S, Khuzani AZ, Danala G, Qiu Y, Zheng B. Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms. *Int J Med Inf*. 2020;144:104284.
149. Albahli S, Albattah W. Deep transfer learning for COVID-19 prediction: case study for limited data problems. *Curr Med Imaging*. 2020.
150. Minaee S, Kafieh R, Sonka M, Yazdani S, Jamalipour SG. Deep-COVID: predicting COVID-19 from chest X-ray images using deep transfer learning. *Med Image Anal*. 2020;65:101794.
151. Apostolopoulos ID, Mpesiana TA. Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Phys Eng Sci Med*. 2020;43:635–40.
152. Romero M, Interian Y, Solberg T, Valdes G. Targeted transfer learning to improve performance in small medical physics datasets. *Med Phys*. 2020;47:6246–56.
153. Clancy K, Aboutalib S, Mohamed A, Sumkin J, Wu S. Deep learning pre-training strategy for mammogram image classification: an evaluation study. *J Digit Imaging*. 2020;33:1257–65.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

