

Review

Open Access



Computer vision and machine learning for medical image analysis: recent advances, challenges, and way forward

Eyad Elyan¹, Pattaramon Vuttipittayamongkol², Pamela Johnston¹, Kyle Martin¹, Kyle McPherson¹, Carlos Francisco Moreno-García¹, Chrisina Jayne³, Md. Mostafa Kamal Sarker⁴

¹School of Computing Science and Digital Media, Robert Gordon University, Aberdeen, AB10 7GJ, UK.

²School of Information Technology, Mae Fah Luang University, Chiang Rai 57100, Thailand.

³School of Computing, Engineering & Digital Technologies, Teesside University, Middlesbrough TS1 3BX, UK.

⁴National Subsea Centre, Robert Gordon University, Aberdeen, AB10 7GJ, UK.

Correspondence to: Prof. Eyad Elyan, Robert Gordon University, Garthdee House, Garthdee Road, Aberdeen, AB10 7GJ, Scotland, UK. E-mail: e.elyan@rgu.ac.uk

How to cite this article: Elyan E, Vuttipittayamongkol P, Johnston P, Martin K, McPherson K, Moreno-García CF, Jayne C, Sarker MMK. Computer vision and machine learning for medical image analysis: recent advances, challenges, and way forward. *Art Int Surg* 2022;2:xx. <http://dx.doi.org/10.20517/ais.2021.15>

Received: 23 Dec 2021 **First Decision:** 25 Jan 2022 **Revised:** 1 Mar 2022 **Accepted:** 8 Mar 2022 **Published:** 22 Mar 2022

Academic Editor: Mehmet Mahir Özmen **Copy Editor:** Xi-Jun Chen **Production Editor:** Xi-Jun Chen

Abstract

The recent development in the areas of deep learning and deep convolutional neural networks has significantly progressed and advanced the field of computer vision (CV) and image analysis and understanding. Complex tasks such as classifying and segmenting medical images and localising and recognising objects of interest have become much less challenging. This progress has the potential of accelerating research and deployment of multitudes of medical applications that utilise CV. However, in reality, there are limited practical examples being physically deployed into front-line health facilities. In this paper, we examine the current state of the art in CV as applied to the medical domain. We discuss the main challenges in CV and intelligent data-driven medical applications and suggest future directions to accelerate research, development, and deployment of CV applications in health practices. First, we critically review existing literature in the CV domain that addresses complex vision tasks, including: medical image classification; shape and object recognition from images; and medical segmentation. Second, we present an in-depth discussion of the various challenges that are considered barriers to accelerating research, development, and deployment of intelligent CV methods in real-life medical applications and hospitals. Finally, we conclude by discussing future directions.

Keywords: Medical images, classification, recognition, deep learning, deep convolutional neural networks



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



1. INTRODUCTION

Computer vision (CV) is concerned with giving the computer the ability to process and analyse visual content such as 2D, videos, and 3D images. CV is common across a wide range of applications, including oil and gas [1-3], fishing and agriculture [4], medical image analysis [5-9], robotic surgery [10,11], and others.

CV tasks can be broadly categorised into: image classification, object detection and recognition from images, and image segmentation tasks [12]. Image classification tasks are considered among the most common CV problems [13,14]. These are widely used, especially in the medical domain, and are often formulated as a supervised machine learning (ML) problem [15], where a set of features X (often extracted from the image) is used to predict a certain outcome y referred to as a label. Formally, the aim of a classification task is to find a function $h(x)$ that maps an input image X to y as

$h(x) : X \rightarrow y$, where y represents a set of discrete classes (e.g., classes representing the severity of the disease as mild, moderate, severe). A typical example is to build an automated diagnosis/classification system of pigmented skin diseases using dermoscopic images [16]. A similar task [17] is presented where a classification model to classify patients as either COVID-19 infected or normal (healthy) was proposed based on a collection of chest X-ray images of COVID-19 patients. It should be noted that the key task in both of these examples is to decide if an image contains a specific disease or not, which is considered a *binary classification* problem. The labelled datasets for these kinds of supervised ML problems are represented as feature vectors, as can be seen in Equation 1. X can be extracted features of each image, but in these examples, it is simply the pixels of the image itself. Y is the labels of the images in the dataset.

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & \vdots & \dots & x_{nn} \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ \dots \\ y_m \end{bmatrix} \quad (1)$$

Researchers in the past have developed various ways to extract low-level and high-level features from images. Typical features might include corner points, edges, colour intensity, scale-invariant features such as SIFT and SURF [18,19] and others. SIFT and SURF, in particular, attracted attention from the research community because these features are invariant to image scaling, rotation, pose and illumination, which were considered key challenges in CV and medical images. These features are then used to train ML models to perform a certain supervised classification task. As can be seen in Equation 1, ML provides computers with the ability to learn from historical observation without the need to explicitly program it, or design heuristics to account for the infinite possible combination of features within these observations. There is a wide range of ML algorithms [15], and the choice of a particular algorithm is often informed by several factors, such as the type, size, and complexity of the data and the task. Common ML methods include support vector machines (SVM) [20], ensemble-based methods such as random forests (RF) [21], artificial neural networks (ANN) [22], and others. Before the use of deep learning (DL) and deep convolutional neural networks (CNN) [23] became widespread in 2012, this traditional approach was the most common way of handling CV tasks, including classification, object detection, and object tracking. However, one key disadvantage of such an approach is that the performance of the chosen ML model would rely heavily on the quality of the extracted image features. The quality of these features is often sensitive to various conditions such as light and the object's orientation within the image, as well as noise and other factors. This means that deployed solutions involving hand-crafted features might be considerably less successful than implied by the results using controlled data. That said, methods based on feature extraction and engineering (hand-crafted features) continue to be commonly used across CV tasks such as object detection and localisation, and segmentation. For example, in [24], the authors used a modified version of Viola

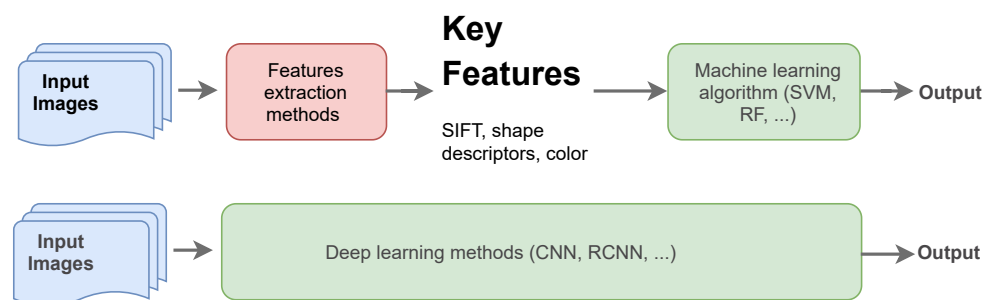


Figure 1. Traditional CV methods (top) vs. deep learning approach (bottom)^[11].

and Jones method^[25] to detect objects from ultrasound images and reported relatively good results. However, progress in solving many CV tasks (e.g., locating the area of interest in an image) continues to challenge the research community. Moreover, research experiments are often carried out in controlled settings, which means that some solutions cannot generalise well across unseen data. It should be noted that generalisation is also considered a problem for DL-based methods, however, with larger volumes of data, transfer learning, data augmentation and other methods, it can be argued that DL-based methods often generalise better than traditional ML-based methods.

The advances in DL and deep CNN research and development since 2012^[23], resulted in significant progress in the CV domain. Thanks to improved hardware to run the algorithms and the availability of large quantities of data, DL-based methods have become more and more widespread. They are now considered the most common and top-performing algorithms in handling many CV tasks. The key difference between traditional methods that use hand-crafted features and DL-based methods, is that the latter is capable of learning the features (underlying representation) of the input images in an end-to-end manner without the need for feature extraction or engineering as illustrated by Gumbs *et al.*^[11] in Figure 1. The ability of DL models to learn and capture the underlying representation of the training data has significantly improved the performance not only in CV but also across various domains such as: Gaming and AI^[26], Natural Language Processing^[27], Health^[28], Cyber Security^[29] and others.

CNN-based methods significantly advanced the field of CV, particularly in the areas of medical image analysis and classification^[5]. These methods have been used since the 1980s. However, due to the increase in computation power and algorithmic development, they are considered now a cornerstone across various vision tasks. CNNs have the ability to capture the underlying representation of the images using partially connected layers and weights sharing. Many CNN architectures consist of a small number of convolution layers, followed by activation functions, and pooling layers for down-sampling of the images, as can be seen in Figure 2. The repeated application of filters (kernels) to the input image results in a map of activations (often called feature maps), which indicate locations of interest in the input image.

However, it should be noted that training modern, deep CNN-based methods require large volumes of data. In supervised ML applications, such data (images, videos, ...) must be fully annotated, and this is not always possible or practical. Therefore, in some cases, transfer learning^[30] is utilised in order to overcome this requirement. In short, transfer learning is a common technique that aims at reusing pre-trained models that were built to solve a particular task to solve new tasks. Pomponiu *et al.*^[6] used a pre-trained CNN to learn the underlying representation of images and extract representative features that help detect skin cancer disease. Then, they fed the extracted features into a k nearest neighbour classifier (kNN) and reported an average ac-

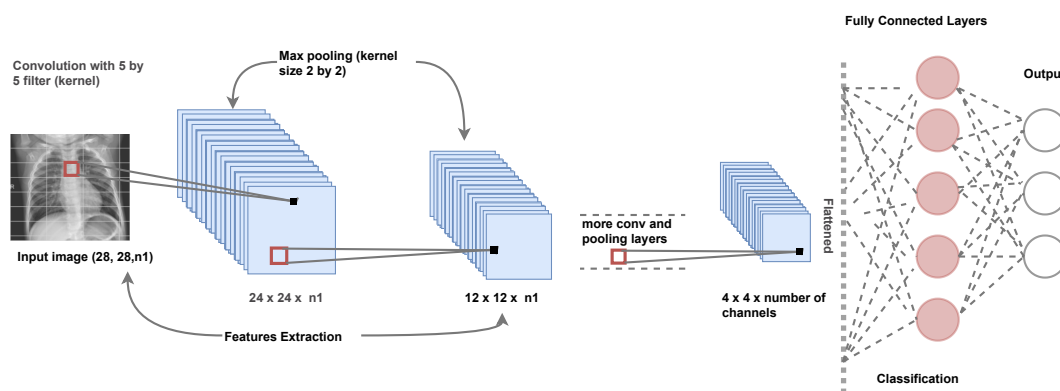


Figure 2. Schematic diagram of CNN model with arbitrary architecture.

accuracy of 93% with similar sensitivity and specificity. Similarly, Esteva *et al.* [7] trained a CNN model using a dataset of 129,450 clinical images represent various skin diseases. The authors also used a pre-trained model to boost performance and reported results comparable with experts. Existing literature shows clearly how CNN-based methods have played a crucial role in solving complex classification tasks in the medical domain. A recent review shows clearly how CNN-based methods advanced research in many areas such as breast cancer, prostate, lung, skin and other types of cancer disease [31]. Similarly, Levine *et al.* [8], in a relatively recent work, showed how DL helped to greatly advance cancer diagnosis, and that DL-based method achieved comparable performance to medical experts in particular classification tasks such as radiology and pathology. That said, it must be noted that even with transfer learning and the availability of large volumes of medical images to train DL models, there is a concern regarding the model's ability to generalise across unseen data or data from different resources. In other words, more work on cross-data evaluation needs to be carried out. Aggarwal *et al.* [32] recently argued in a detailed review paper that the diagnostic accuracy of DL methods on medical images might be overestimated. The authors pointed out the urgent need for a more consistent approach in developing and evaluating DL-based methods in medical settings.

Shape and object detection and recognition from images have also been greatly advanced using CNN and DL-based methods. These tasks are of paramount importance in the medical domain and robotic surgery [28]. Here, the task is not only to classify to check if an image shows some form of abnormality, but also to localise that abnormality within the medical image. Typical DL-based methods for object detection and recognition from images include, Faster Region-based CNN (R-CNN) [33], single shot detectors (SSD) [34], region-based fully convolutional networks (R-FCN), [35], You Only Look Once (YOLO) [36] and others. These methods have shown superior performance across various complex CV tasks in the medical domain, in particular in object localisation and recognition [37–39]. These object recognition methods -in practical terms- vary in terms of accuracy in recognising objects and their suitability for real-time object detection and tracking. For example, YOLO is considered among the fastest object recognition methods. However, R-CNN is often more accurate; for technical comparison between these methods, the reader is referred to [40]. But overall, this progress in the development of object recognition methods has also helped solve some complex CV tasks such as detecting and tracking objects of interest in video footage [41–43], which is an important and very common task in robotic surgery. For example, in [41], a dataset of videos from ten surgeons was used to build a DL model to speed up detection and localisation of tools in robot-assisted surgery (RAS). Similarly, Lee *et al.* [42] used a collection of videos to train a DL model to track surgical instruments to evaluate surgeons' skills in performing procedures by robotic surgery and reported good results. While the provision of such datasets gives opportunities for advancement, it also serves to illustrate the challenge of data availability, and the authors of [42] conceded that their dataset did not provide an exhaustive reference of all possible surgical situations. Various other tasks were also made possible due to these advancements that took place at the algorithmic level, such as processing and

analysing Magnetic Resonance Imaging to detect brain tumours^[44], understanding CT scan images^[9], and localising abnormalities in chest X-ray images^[5]. Interestingly, in^[5], the authors presented a method based on dividing the image into a set of patches (Patch-based Convolutional Neural Network), and using prior knowledge of positive and negative images, their proposed method learned to localise the area of interest (i.e. the area in the image that demonstrated abnormality). It should be noted that many of these tasks were inherently challenging for traditional CV-based methods. However, using the latest development in DL- and CNN-based methods and good-sized datasets, these tasks became achievable.

Medical image and video segmentation^[45,46] has also greatly benefited from the progress in DL-methods' development. Image segmentation is an important CV task for certain medical applications such as robotic-assisted surgery education^[47] and other relevant applications. Examples include: the work presented in^[45] for robotics instrument segmentation, in^[48] for surgery instrument segmentation, and in^[46] for segmenting surgery videos.

It can be argued that progress that took place at the algorithmic level in DL- and CNN-based methods, as well as the availability of high-performing computing machines powered with graphical processing units (GPUs) have significantly progressed research and development in medical image analysis and understanding. For a recent review on how progress helped advance many robotic-assisted surgeries, and other medical-related applications, the reader is referred to^[49]. Additionally, the availability of large volumes of medical images and videos in the public domain also has advanced and accelerated the development of various medical applications that use core CV tasks (e.g., classification, object recognition, segmentation). Examples of these datasets include: MURA^[50], which is a public domain dataset containing 40,561 musculoskeletal radiographs from 14,863 studies representing 11184 different patients; the colon cancer screening dataset^[51] containing 40 colonoscopy videos comprising almost 40,000 image frames), Lung images dataset^[52] comprising CT scans annotated with abnormality information, and others. This progress has the potential of accelerating the deployment of various medical applications that utilise CV. However, existing literature shows that there are limited practical examples being deployed into front-line health facilities^[53]. In this paper, we focus on CV-related literature, recent development, and key challenges, with emphasis on the most common CV vision tasks (classification, segmentation, and object detection). These tasks are considered the key building blocks for any CV solution in the medical domain. The main contributions of the paper can be outlined as follows:

- The paper provides an in-depth critical technical review of the latest developments in medical image analysis and understanding. We critically review existing literature in the CV domain that addresses complex vision tasks, including medical image classification, shape and object recognition, and medical image segmentation
- Comprehensive discussion and critical evaluation of various medical image applications that utilise medical images and CV-based techniques with discussion and evaluation of existing medical datasets in the public domain
- We present an in-depth discussion of the various challenges that are considered barriers to accelerating research, development and deployment of intelligent CV methods in real-life medical applications and hospitals

The rest of this paper is organised as follows: In section 2, we discuss CV and the latest developments in the context of medical images and with focus on key tasks such as classifying, segmenting and recognising shapes and regions of interest. Section 3 focuses on medical image applications and list medical datasets of various types that exist in the public domain, providing a great opportunity to accelerate research and development in this area. Challenges in CV and medical applications are discussed in Section 4, and finally, conclusions and future directions are presented in Section 5.

2. MEDICAL IMAGES

The most common computer vision (CV) tasks in applications that utilise medical images include image classification [6,8], medical image segmentation [45,47,48] and detecting and recognising shapes and objects of interest in medical images [41-43]. Deep learning (DL)-based methods have proven to be superior over other traditional methods in handling such challenging tasks and across a wide range of medical applications such as cancer detection, magnetic resonance imaging (MRI) segmentation, X-ray analysis and others. In this section, we provide a thorough and deep technical review of existing literature that utilise the latest developments in DL- and convolutional neural networks (CNN)-based methods to progress medical image analysis and understanding.

2.1 Classification

In CV, image classification is a fundamental task that also plays an important role in computer-aided diagnosis (CAD) systems over the decades. Traditionally, image classification is used to classify, or label, an image, or the sequence of images, as including one, or more, of a number of predefined diseases, or as without diseases (i.e. normal case) in the CAD system for medical image analysis [54,55]. Common clinical uses of DL-based medical image classification methods include: skin cancer classification in dermoscopic images [56,57], lung cancer identification in CT images [58], breast cancer classification in mammograms [59] and ultrasound [60] images, brain cancer classification in MRI images [61,62], diabetic retinopathy [63,64], eye disease recognition in retinal fundus images [65], and so on. In addition, the classification of histopathology images is also widely used for identifying different types of cancers such as colon cancer [66], prostate cancer [67], breast cancer [68], and ovarian cancer [69]. Recently, DL-based methods have also been popularly utilised to identify COVID infections in the chest X-ray images [55], and although the dataset for this application was limited to 150 patients, the deep neural network achieved 96% sensitivity. In medical image classification, CNN is the state-of-the-art classification approach with the continuous development of DL models, including fine-tuning of existing models for application in the medical domain and the development of new models and algorithms specifically for medical applications. A fine-tuned ResNet-50 [48] architecture is used to classify skin lesions using dermatoscopic images in [56], where the authors observed that the classification accuracy of some models might be highly dependent on the clinical setting and the precise location of the skin lesions in the captured images of the dataset. In [70], an optimal deep neural network (ODNN) with linear discriminant analysis (LDA) is utilised to classify lung cancer in CT images. The conditional generative adversarial network (cGAN) with a simple CNN is proposed in [59] to classify breast cancer subtypes in mammograms, achieving 72% accuracy in classifying the shape of a tumour. In [61], an enhanced ResNet is applied to solve the brain cancer classification problem in MRI images. To classify diabetic retinopathy, an active DL method is proposed in [63]. Recent popular image classification architecture, InceptionResNetV2 has been used to identify retinal exudates and drusen in ultra-widefield fundus images [65]. A multiscale decision aggregation is used in [67], pre-trained Inception-V3 in [68], and a hybrid evolutionary DL in [69] to classify: prostate, breast, and ovarian cancer, respectively.

In the field of thermography, classification is used in a wide range of medical applications, from mass screening for the fever to detection of vascular abnormalities and some cancer diagnoses [71]. Thermography is a non-invasive and a relatively portable method of gathering images to aid diagnosis. The use of thermal cameras to identify the presence of fever in subjects in public spaces had recently risen to the fore with the COVID-19 pandemic when mass screening for fever was recommended as an early preventative measure [72]. Infrared technology had been utilised with some success during the SARS outbreak as a contact-free method of mass-screening for fever detection [73], identifying 305 febrile patients from a total of 72,327 outpatients at a busy hospital. More recently, thermal imaging in public spaces has become ubiquitous, and CV techniques have been applied to improve the accuracy of temperature readings and fever classification from these devices. It has been found that the maximum temperature from the inner canthi is more indicative of fever than the traditional forehead scan [74], and CV techniques have been developed to localise these areas [75].

Thermal image features combined with machine learning have achieved high levels of accuracy in detect-

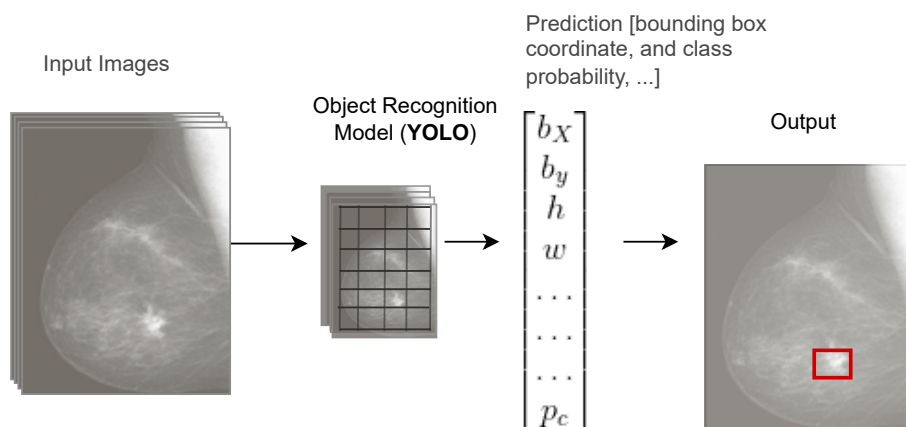


Figure 3. Schematic diagram of object detection methods.

ing the presence of facial paralysis (93% accuracy)^[76,77], carotid artery stenosis (92% accuracy, SVM using features from thermal images)^[78] and localisation of areas of the brain most at risk during ischemic stroke intervention surgery^[79]. Breast cancer is also a very active field of development for thermal images with the non-invasive thermal camera provides a more attractive option than traditional mammography^[80]. DL methods have seen some success in detecting breast cancer, with over 98% classification accuracy on some datasets using CNN^[81,82]. In applications of mental health, emotion detection from thermal images is a growing field, too, which already has a number of large datasets suitable for training CV models^[83]. Given the growing accessibility of thermal cameras, this field might provide an emerging opportunity for remote mental health monitoring and treatment.

2.2 Object detection

Overall, object detection models consist of localisation and identification tasks. The localisation task leads to localising the object position in the image using a bounding box or mask to define which pixels within the image depict the object of interest. The identification task refers to recognise the objects referring to specific pre-defined categories, or to classify the object within the bounding box. Object detection algorithms are commonly used in the medical image analysis domain in order to detect the initial abnormality symptoms of patients. Typical examples include, detection of the lung nodule in the X-ray^[84] or chest CT images^[85], breast lesion detection in mammograms^[86] and ultrasound images^[87].

Generally, there are two approaches that exist in DL-based object detection models: anchor-based methods and anchor-free methods. Anchor-based methods can be divided into single-stage or multistage techniques. However, while the single-stage technique is fast or computationally efficient, detection performance is not better than the multistage technique. The multistage technique has the best detection performance but is computationally costly. Moreover, there are two common single-stage techniques that are extensively used as single-stage detectors with simple architectures: YOLO^[36] and SSD^[34]. In both YOLO and SSD, a feed-forward CNN creates a fixed number of bounding boxes and their respective scores for the appearance of object instances of pre-defined classes in the boxes; non-maximum suppression is applied to generate the final prediction map. YOLO uses only a single-scale feature map whereas SSD uses multiscale feature maps for improving detection performance. Most of these methods are based on the allocation of the bounding box or boxes that contain the object/s of interest. As can be seen in Figure 3, the final outcome of an object detection (e.g., YOLO) is a vector with a set of values representing the coordinates of the center of the bounding box (b_x, b_y), its height and width (h, w), and the probability the bounding box contains the class of interest (p_c), and so on.

In contrast to single-stage object detection techniques, two-stage detection techniques produce a group of regions of interest (ROIs) where a complete object is likely to be found, and recognises each of them. Faster-RCNN^[33] and Mask-RCNN^[88] are commonly used examples of the two-stage detection technique. Both models first produce object proposals by a region proposal network (RPN) and then classify those produced proposals. The difference between these two models is that Mask-RCNN has an additional segmentation branch, whereas Faster-RCNN localises objects by defining the co-ordinates of a bounding box.

Currently, there are many researchers focused on developing anchor-free methods of object detection, and CornerNet^[89] is one of them. CornerNet uses a single CNN that reduces the use of anchor boxes by using the paired key points from the object bounding box top-left and bottom-right corner.

These methods and their extensions have significantly advanced the area of object localisation and recognition. A recent review shows that many similar medical image analysis tasks are carried out using such DL-based methods^[90]. A typical example, is the work presented by Pang *et al.*^[91], where an extension of YOLO was used to identify cholelithiasis and classify gallstones in CT images using a large dataset of 223,846 CT images with gallstone representing 1369 patients. This large dataset contributed to the high accuracy (over 92% in identifying granular gallstones).

It should be noted however, that making use of these object localisation/recognition methods, require large volumes of fully annotated medical images. Such data may not always be readily available, and data annotation, in particular, is considered one of the most labor-intensive and expensive tasks in the medical field as will be in the Challenges section. However, several methods have been presented recently to overcome this hurdle. For example, the work by Schwab *et al.*^[5] who presented a new approach to localise and recognise critical findings in chest X-ray images (CHR), using multi-instance learning method (MIL) that joins classification and detection. The method presented is based on dividing the image into a set of patches (Patch-based CNN). Using the prior knowledge of positive and negative images, the proposed MIL method learns which patches in the positive images are negative. Three different public datasets were used, and competitive results were reported, achieving over 0.9 AUC. Notice here, that there was no requirement to fully annotate all images of the dataset with localisation information, instead, it was sufficient to know only if a complete image was positive or negative.

2.3 Segmentation

Image segmentation refers to a pixel-wise classification task that segment an image into areas with the same attributes. The goal of medical image segmentation is to find the region or contour of a body organ or anatomical parts in images. While object detection methods often produce a bounding box defining the region of interest, segmentation methods will produce a pixel mask for that region. Applications include: segmentation of the whole heart^[92], lung^[93], brain tumours^[94], skin^[95] and breast tumours^[96]. Like other CV tasks, segmentation can also be applied to different medical image modalities. A great breakthrough was achieved in DL-based image segmentation after the introduction of the Fully Convolutional Neural network (FCN)^[97]. U-Net^[98] is the most famous end-to-end FCN model used for medical image segmentation. U-Net and its extensions have also been successfully applied to wide range of medical imaging segmentation tasks, and a detailed review is presented by Litjens *et al.*^[90].

An improved U-Net with a generative adversarial network (GAN) is applied in^[92] to segment the whole heart from the CT images. An automatic, self-attention based lung segmentation model is proposed in^[93] in chest X-rays. In^[94], an attention mechanism is utilised for brain tumour segmentation in MRI multi-modalities brain images. The conditional generative adversarial network (cGAN) is used in^[95] and^[96] to segment skin and breast lesion, respectively.

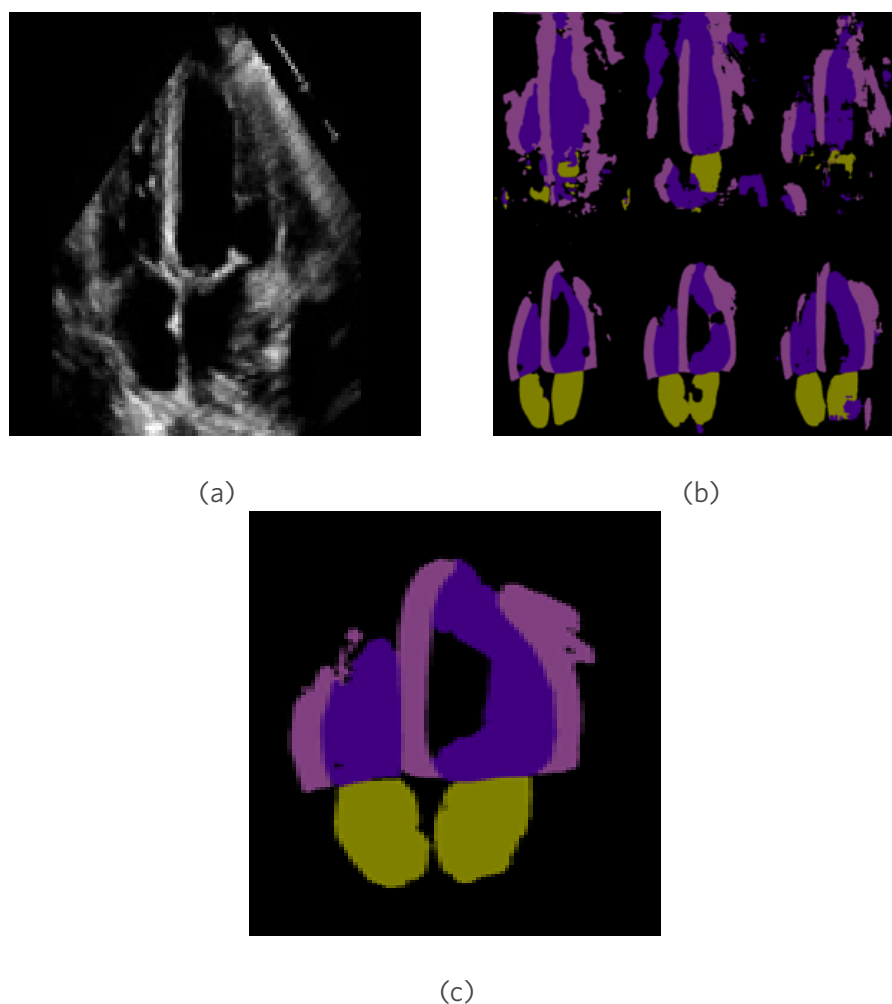


Figure 4. Segmentation of four chamber of heart in echocardiographic image. (a) a sample echocardiographic image of heart with four chambers; (b) segmentation results from the six deep learning segmentation algorithms; and (c) ensemble segmentation result with improved performance over the six initial ones.

Heart related applications have benefited greatly from the recent advances in DL. For example, right ventricle (RV) segmentation to study and detect the anatomically complex shape of the chamber to aid in the diagnosis of several diseases that are directly related to the RV such as coronary heart disease, congenital heart disease, pulmonary hypertension, among others. RV is one of the four chambers of the heart, along with the left ventricle (LV), right atrium (RA) and left atrium (LA). A DL-based framework was proposed by Luo *et al.* [99] to segment the chambers of the heart from magnetic resonance images. More specifically for pulmonary hypertension detection, in a recent study by Dang *et al.* [100] authors implemented a weighted ensemble of DL methods based on comprehensive learning particle swarm optimisation (CLPSO). To this end, they trained a total of six transfer learning models for segmentation, and these were then assembled to get the best possible output. This output was calculated as the weighted sum of segmentation outputs, and the CLPSO algorithm was used to optimise the combined weights. Figure 4a shows a sample echocardiographic image, while Figure 4b presents the six initial segmentation outputs. Finally, Figure 4c shows how the ensemble output outperforms all of the individual segmentation masks. These transfer learning systems were retrained using the CAMUS dataset, which contains 250 images of hearts where only the LV and the LA were segmented. Later on, clinicians from "Hospital 20 de Noviembre" in Mexico city provided 120 sequences with the four chambers segmented, so that the systems could be retrained once again and were capable of localising all four chambers of the heart.

Table 1. Summary of existing medical image analysis datasets with the referred literature

Ref.	Domain	Image modality	Application	Dataset	Methods	#Images	#Classes
[102]		MRI	AD/HC classification	ADNI	DBN with RBMs	300	2
[110]		MRI and PET	AD/MCI/HC classification	ADNI	DPN with an SVM	258	3
[111]	Brain image analysis	MRI	Brain extraction	IBSR, LPBA40 and OASIS,	3D-CNN	188	2
[112]		MRI	Lacune detection	RUNDIMC and the FUTURE	FCN and 3D CNN	1075	2
[103]		T1MRI, fMRI and DTI	Survival prediction	Private	3D CNN and SVM	69	2
[104]	Retinal image analysis	Fundus	Detection of multiple retinal diseases	STARE	DNN	10000	10
[105]		Fundus and Pathology	Segmentation of multiple retinal diseases	Messidor	FCN	1200	4
[113]		Fundus	Segmentation of retinal blood vessel	DRIVE and STARE	cGAN	40	2
[106]		Fundus	Detection and segmentation	REFUGE 2018 challenge	SSD and cGAN	800	3
[55]	Chest image analysis	X-ray	COVID-19 classification	COVIDx	EDANet	1203	3
[107]		X-ray	Detection of abnormal pulmonary patterns	ACS at ASAN	eDenseYOLO	9792	5
[108]		X-ray	Segmentation of malignant pulmonary nodules	Private and National Lung Screening Trial (NLST)	ResNet-34	5485	2
[114]		CT	Detection of lung nodule	LIDC-IDRI	FasterRCNN	3042	2
[85]		CT	Detection and segmentation of lung nodule	LUNA16 and LIDC-IDRI	FasterRCNN and AWEU-Net	5066	2
[115]		MRI	Detecting of coronary artery calcium	Private	DCNNs	1689	2
[66]	Digital pathology images	H&E WSI	Detection of ICOS protein expression in colon cancer	Epi700	Detectron2 and U-Net	97	2
[116]		H&E WSI	Predicts origins for cancers of unknown primary	TCGA	TOAD	32537	18
[117]		H&E WSI	Identify the sub-type of renal cell carcinoma	TCGA, CPTAC, CAMELYON16 and CAMELYON17	CLAM	3750	3
[118]		H&E WSI and WBCs	Nuclei instance segmentation	GLaS, GRAG, MonuSeg, CPM, and WBC	NuClick	3129	7
[59]	Breast image analysis	Mammograms	Breast mass segmentation and shape classification	DDSM and REUS Hospital	cGAN and CNN	851	4
[119]		MRI	Breast tumors classification	BI-RADS	Pre-trained CNNs	927	2
[96]		Ultrasound	Breast tumor segmentation	BUS	CIA cGAN	413	2
[120]		DBT and X-ray	Breast mass segmentation	DBT	U-Net	4047	2
[121]	Cardiac image analysis	MRI	Ventricle segmentation	RVSC	Dilated-Inception Net	48	2
[122]		MRI and CT	Whole heart segmentation	MM-WHS	Different CNNs	120	2
[123]		MRI	Aorta segmentation	UK Biobank	U-Net with RNN	500	2
[124]		MRI and CT	Cardiac substructure segmentation	MM-WHS	cGAN with U-Net	120	2
[125]	Abdominal image analysis	CT	Liver tumor segmentation	LITS	U-Net	865	2
[126]		MRI	Diagnosed polycystic kidney disease	TEMPO	Ensemble U-Net	2400	2
[127]		CT and MRI	Multi-organ segmentation	TCIA and BTCV	DenseVNet	90	8
[128]		Endoscopic	Diagnosis of gastric cancer	Zhongshan Hospital	ResNet50	993	2
[95]	Dermatology	Dermoscopy	Skin lesion segmentation	ISBI 2017 and ISIC 2018	SLSNet	6444	2
[129]		Dermoscopy	Skin lesion classification	ISBI 2016, 2017 and ISIC 2018	Pre-trained CNNs	14439	7
[130]		Photographs	Automated grading of acne vulgaris	Private	Pre-trained CNNs	474	3
[131]		Dermoscopy	Classification of six common skin diseases	Xiangya-Derm	Xy-SkinNet	5660	6

3. APPLICATIONS AND DATASETS

Before the widespread use of deep learning (DL)-based methods, extensive research was carried out using traditional computer vision (CV) methods, where features were engineered or discovered, extracted from images, and then used to train supervised machine learning (ML) algorithms^[101]. However, it is safe to argue now that almost all research related to medical image classification, object recognition and medical image segmentation is currently driven by DL-based methods, especially in the last 10 years^[90].

Advances that took place at the algorithmic levels related to DL-based methods and convolutional neural networks (CNN) have significantly improved accuracy and have been successfully applied across a wide range of medical image applications. Typical applications include brain image analysis^[102], survival prediction using brain medical imaging^[103], retinal image analysis-based applications^[104–106], chest image analysis^[55,107,108] and others. Many of these applications made use of different image modalities, including CT, magnetic resonance imaging (MRI), Ultrasound and X-rays^[109].

In addition to the development at the algorithmic level that took place over the past decade, the other key driving force for the progress in medical image analysis and applications is the availability of datasets in the public domain. This availability enabled the CV research community to test and evaluate many methods for solving complex CV tasks related to medical images. In the literature, several datasets for medical image analysis have been presented based on recent DL methods. These datasets provide a unique opportunity for the research community to advance the research in solving complex vision tasks in the medical domain. Table 1 lists some common datasets in the different domains (e.g. brain, retinal, chest, breast, etc.) with some details.

4. CHALLENGES

The progress that took place in computer vision (CV) is unprecedented, and various inherently challenging CV tasks are now considered solved problems. Typical inherent CV challenges include data variation (e.g. lights, pose), occlusion (overlapping objects in the images/videos), and others. However, despite this significant progress, there remain challenges that need to be addressed to scale up the use of deep learning (DL)-based methods across a wider range of applications in the medical domain. In this paper, we argue that many of these challenges are related to data quality and data availability. However, in the medical domain, some data will

always need to be gathered opportunistically, so it may not be possible to overcome these challenges through data gathering alone. These challenges apply to the full range of medical applications, including complex medical settings such as robotic surgery^[11] and comparatively simple settings such as the detection of fever with thermal cameras. In both of these settings, CV algorithms need to account for the dynamic environment. In robotic surgery, this includes navigation, movement, object recognition (of deforming objects) and actions. In other settings, data variability caused by the environment and equipment used to capture the images can lead to issues. Clearly, data availability and quality play a crucial role in the learning process.

4.1 Data availability and quality

The quality and viability of CV models developed through machine learning (ML) often depend directly upon the quality of available data used to train the models. This is especially important in medical imaging, where specialist imaging technology is required to capture the images, and expert knowledge is required to select, annotate and label data. It should be noted that the performance of many DL-based methods relies on large volumes of images, and in supervised learning settings, these should be fully labelled and annotated.

The annotation of medical images, in particular, continues to be one of the most demanding tasks and requires long hours of medical experts' time. Despite the latest development in CV, annotating medical images as well as images and videos across other domains is still largely carried out manually, often done by drawing bounding boxes around the area of interest, or creating a mask manually, so that such data can be used in the training process. In a complex and dynamic environment, the challenge of collecting and annotating the data (often collections of videos) is even more demanding, labour-intensive and often expensive. Consider, for example, a simple task (for medical experts) such as analysing operative videos to detect steps in laparoscopic sleeve gastrectomy surgery^[132], where the authors had to collect and annotate videos, using two experts, capturing patients' operations (entire surgery). Although good results were reported, it is practically infeasible to scale this approach to capture all data variation in such scenarios. Therefore, more work needs to be done in the area of unsupervised or semi-supervised DL-based methods.

In the medical domain, data must often be captured opportunistically, using equipment, patients and specialists as and when they are available. Strictly controlled and consistent conditions specifically for data gathering purposes are not always possible. This leads to challenges in the generalisability of models. For example, in the field of thermal imaging, even data gathered specifically for the purposes of study exhibits high levels of diversity between datasets^[83,133]. A large dataset specifically for febrile identification from thermal images, does not yet exist^[134]. However, a meta-analysis of existing studies into the use of thermal scanners for febrile identification demonstrated high levels of diversity between the studies^[133], with this partly attributed to differences in equipment, scanner location, demographics of the study population. This diversity means that models trained on one dataset might not generalise to other datasets, let alone situations of widespread use. The International Organization for Standardization has produced a standard specifically for the purpose of mass temperature scanning for fever detection^[135], though some evidence suggests the standard is not yet widely adopted^[136]. A complete lack of standardisation in data gathering protocols in some fields will produce diverse but disjoint datasets, making model generalisation exceptionally difficult.

4.2 Data bias

One of the inherent problems related to CV and ML in general is the data bias or, as commonly known, class-imbalance problem. Classification with imbalanced class distribution poses a challenge for researchers in the field of ML^[137]. An imbalanced dataset is a common term describing a dataset that has a remarkably unequal distribution of classes, as depicted in Figure 5. Such a dataset is likely to cause a bias in the learning process of a ML algorithm. This is because typical learning algorithms are designed to maximize the overall accuracy in classification regardless of the model's per-class accuracies. Hence, in an imbalanced dataset, the learning algorithm will be more compromised for misclassification of minority class instances than majority class ones.

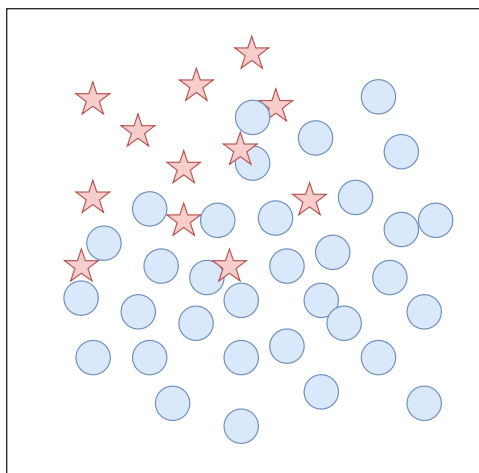


Figure 5. A dataset with imbalanced class distribution.

This can lead to an undesirable scenario when the minority class accuracy is nil while the overall class accuracy still reaches over 90% due to a high imbalance ratio of the minority class to the majority class. This high overall accuracy will be misleading if one is not aware that the predictive model has totally failed to detect anomaly cases. The problem becomes more concerning when the minority class is the class of interest and has a high error cost. This situation has been reported in the literature across a wide variety of problem domains, including medicine^[138,139], oil and gas industry^[140], finance^[141], and banking^[142].

In the medical domain, imbalanced datasets are often seen due to the limited availability of samples, generally patient data, belonging to the group of interest^[143]. For example, the data of patients with benign tumours may greatly outnumber the data of cases with malignant tumours, which is a natural phenomenon of many existing types of tumours that the overwhelming majority of these tumours are benign^[144-146]. Predictive analysis of other diseases such as heart disease, cerebral stroke, Parkinson's, and epilepsy are also good examples of imbalanced data classification tasks^[143,147,148]. Results achieved from the analysis will be crucial and have a high impact on society since these are major threats to public health globally. On this account, a number of research works to serve this purpose have been being carried out. Nonetheless, little has been realised about the issue of disproportional class distribution in the data^[138]. This is evidenced in the review of Kalantari *et al.*^[149], where only 1 out of 71 presented supervised-learning-based approaches for medical application addressed the class imbalance issue.

This is a problem in the medical domain that is unlikely to be solved by sourcing more data samples. To deal with imbalanced datasets, data resampling is widely-used^[150]. Resampling methods are applied to change the class distribution in order to mitigate the imbalance effect on the learning algorithm's performance. Figure 6a and Figure 6b shows the two main resampling approaches, namely, oversampling and undersampling, respectively. Oversampling is the practice of synthesising additional samples in the minority class whereas undersampling is the practice of reducing the number of majority class instances. Algorithm-level solutions, which involve modifying existing learning algorithms to address the imbalance problem in a dataset, are also implemented. One main approach of these solutions is to adjust the cost function according to the imbalance ratio. Others are such as employing ensemble-based methods, one-sided selection, and neural-network-based solutions. For further discussion of recent methods, the reader is referred to^[150].

It should be pointed out, however, that with medical images and medical datasets in general, undersampling methods may not be a favourable technique, because it often results in an information loss, which may have a negative impact on the overall accuracy^[151].

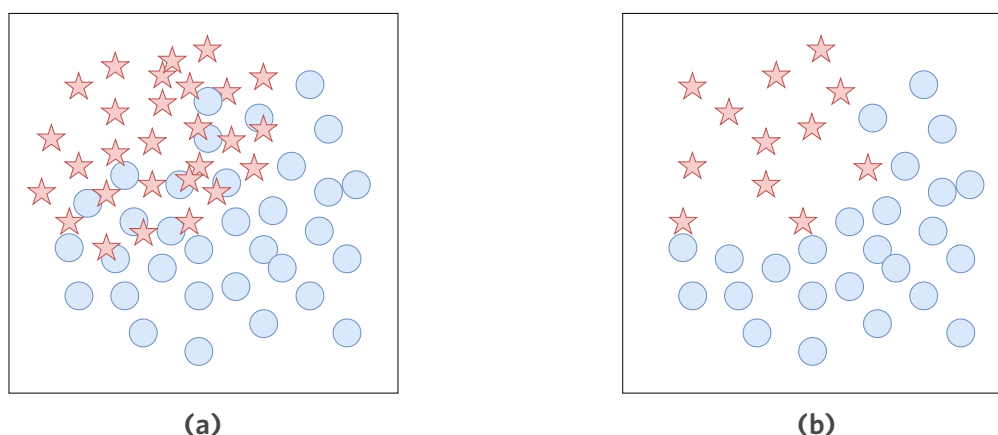


Figure 6. The dataset with (a) oversampling and (b) undersampling applied.

There have been a number of research articles on handling the class imbalance in medical datasets. The approaches range from simply utilising existing solutions to designing new techniques. However, many works have focused on tabulated medical data^[152-155], which is easier to deal with compared to medical images.

Krawczyk *et al.*^[156] used an existing method that combined ensemble and evolutionary undersampling to improve image classification of breast cancer malignancy grading. Bria *et al.*^[157] applied an ad hoc designed cascade of decision trees to handle class imbalance in small lesion detection on medical images. Recent creation of methods to address the class balance problem specifically in image and video analysis can be seen in the following examples: Yuan *et al.*^[158] introduced a regularized ensemble framework of DL for multi-class video classification of bowel cancer detection; Zhou *et al.*^[159] proposed an imbalanced reward signal to be used in a reinforcement learning model for medical image classification.

The emerging generative adversarial networks (GANs) have played an important role in data augmentation in the minority class, especially in CV^[160,161]. Qasim *et al.*^[162] presented a new GAN-based method that generated synthetic images to improve medical image segmentation. Similarly, Rezaei *et al.*^[163] designed new architectures of GANs for minority class oversampling in Magnetic Resonance Imaging for brain disease diagnosis.

As can be seen in the works mentioned above, using complex and advanced techniques seems to be a requirement for solving the imbalanced class distribution in CV tasks. This makes class imbalance a very challenging problem in medical image analysis, which results in limited growth of research in this domain.

4.3 Explainability of computer vision algorithms using medical data

As well as issues in data quality and quantity and the inherent class imbalance within some datasets, there is the question of trust in the algorithms themselves. There are increasing legal regulations and social responsibilities for ML models to be explainable (such as^[164]). Recent studies have highlighted that the lack of robust explanation capabilities in existing algorithms is a challenge to be resolved before AI can see further widespread application within the medical domain^[165,166]. Within this field, explaining the outcomes of CV algorithms is challenging in general, but this is exacerbated within medical image analysis due to (1) the variety of stakeholders involved; and (2) the complexity of data and models to be explained make explaining their outcomes difficult.

Typically, the need for an explanation is subjective and highly dependent on an individual user's context^[167]. Within the medical domain, there is a broad range of stakeholder, each of whom possesses very different

requirements and expectations which must be satisfied by an explanation^[168]. The list of stakeholders include: patients; clinicians; care providers; regulatory and/or governance bodies; and algorithm developers, among others. While there are existing models to link users based upon aspects such as their experience with AI or domain knowledge^[169], in medicine (and similar fields) there is the added complexity of non-overlapping expertise in the form of clinical specialisations^[170]. As a result, ensuring that an explanatory algorithm can comprehensively meet the explanation needs of all stakeholders is a challenge to be resolved.

Explaining the outcome of CV models is currently achieved in two ways: by augmenting existing black-box CV models with post-hoc algorithms that *explain* their decision-making; or by developing *interpretable* models where the decision-making process is reasonably transparent or understandable^[171]. In regards to the former method, most off-the-shelf explanation methods are designed to be problem- and model-agnostic to satisfy a wide range of use-cases^[172,173]. There is then a disconnect between the explanation algorithm and the model it is designed to explain, and as a result, the explanation provided may not be truly reflective of the actual decision-making process (i.e. they are not faithful to each other). For example, recent studies suggest that the widely-used saliency mapping method is not always a reliable source of explanation^[171,174]. On the other hand, explanations derived from interpretable models are guaranteed to be faithful. However, the barrier to entry is slightly higher due to the complexity of understanding the volume of available information and the risk of ‘information overload’^[175]. There is growing work on the use of interpretable CV models for medical imaging^[176-178], but the field is still in its infancy.

4.4. Dynamic environment challenges

Much of the published literature related to medical image analysis and understanding and other domains have used datasets of images and videos that were largely compiled in a controlled environment. This may include controlling light conditions, movement, quality of the images, the position of the camera and subjects, equipment used to capture the data and so on. However, in a very dynamic setting, such controls may not be possible. This still poses a challenge for the CV research community.

A typical controlled environment is a surgical environment and applications related to robotics surgery^[49]. However, as controlled as a surgical environment may be in isolation, there may not be any degree of standardisation between separate operating theatres and data gathering equipment. In such scenarios, the performances and generalisation of the DL models will largely depend on the quality and diversity of the data, regardless of the CV task. However, it should be noted, that in such a scenario, CV tasks become more complicated, where accurate object detection and tracking under various conditions become paramount.

The ability to construct a 3D representation out of 2D visual content (video streams and images) continues to be a challenging problem for the research community. Consider, for example, the need for estimating depth information for endoscopic surgery images, which is an important task to facilitate navigation in a surgery setting. In the DL era, if we can obtain large volumes of good quality videos with the corresponding depth maps, then such a task may be very possible^[179]. However, this is quite impossible in a medical setting due to the dynamic and diverse nature of such an environment. The depth information will be unique to individual patients, but models which predict the depth information will need to be able to generalise across many different, unique patients. Labelling and annotating the amount of data required to achieve generalisable models is labour-intensive and often very expensive, as discussed above. Similar to 3D representation, 3D image registration, as well as learning from different data modalities to improve generalisation of the deep models are still considered as key challenges in the CV research community.

Another closely related challenge is that the quality of the data collected (images, videos) maybe unintentionally degraded in an uncontrolled environment, and this, in turn, will have a negative impact on the performance of any DL model. Existing literature suggests that deep models performance is similar to human performance

on lower quality and distorted images^[180].

5. CONCLUSION AND FUTURE DIRECTION

The progress that has taken place in the area of medical image analysis and understanding over the past decade is considered unprecedented, and can be measured by orders of magnitude. Complex computer vision (CV) tasks such as classifying images, localising and segmenting areas of interest, and detecting and tracking objects from video streams became relatively easy to achieve. This development can be largely attributed to the development that took place at the algorithm levels, especially the development of convolutional neural networks-based methods, the progress in computing power, and finally the availability of large volumes of medical images and related data in the public domain. In this paper, we have reviewed and summarised some of the key technologies and underlying methods behind this progress, and outlined the vast range of medical applications that have greatly benefited from the latest developments in CV and image processing and analysis.

This paper also outlined key challenges and barriers to scaling up the practical use of AI-driven solutions across a wider range of medical applications. Data was found to be the fundamental building block in developing these solutions. There is clear evidence in the literature that with high-quality data, good performance can always be achieved. However, in reality, preparing the data can be very labour-intensive, time-consuming and often expensive. Key tasks such as image classification require accurately labelled data, and in the medical domain, this needs to be carried out by more than one medical expert, to ensure that minimal bias is injected into the data. Similarly, to build a model that is capable of tracking an object of interest, in a surgery video, enough videos and images need to be labelled and fully annotated (e.g., drawing a bounding box or a mask around the region of interest). It can be said that most of these data labelling/annotation tasks are still based on manual or semi-automated approaches. However, algorithms currently under development have the ability to work from partially annotated data and intelligently and automatically annotate images successfully.

Even if data is gathered widely and labour-intensive data annotation is carried out perfectly, there may still exist an inherent class imbalance within the data itself. Recent developments in the literature imply that there are algorithmic ways to successfully overcome this challenge, too. Future work may include building on existing data generation methods such as generative adversarial networks (GANs) to produce more diverse datasets without the need to gather and label/annotate individual samples. GANs have proven to be capable of generating realistic and diverse images from noise and provide a possible solution to generate data conditioned on a particular class of interest and solve the class imbalance problem. Another direction that should be considered is the development of deep learning (DL)-based methods so that they become capable of learning from smaller datasets, which means that efforts to label and annotate datasets can be at least reduced.

There also remains much to be done in terms of ensuring interpretability is built into machine learning models from the ground up. This will help to engender trust within medical practitioners and patients, and ensure that accurate, explainable DL models can be developed for immediate widespread deployment in the medical domain.

Finally, there is an urgent need for experts from the medical and AI domains to work together on an ongoing basis. This will ensure that expert knowledge remains at the heart of the process in creating accurate, understandable and, most importantly, *usable* CV techniques which can advance medical care globally and take us into the future.

DECLARATIONS

Authors' contributions

Made substantial contributions to conception and design of the study, the structure, narrative and the review, critical discussion and evaluation of existing literature: Elyan E, Sarker M, Vuttipittayamongkol P, Johnston P, Martin K, McPherson K, Moreno-García C, Jayne C

Availability of data and materials

Not applicable.

Financial support and sponsorship

None.

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2022.

REFERENCES

1. Elyan E, Jamieson L, Ali-Gombe A. Deep learning for symbols detection and classification in engineering drawings. *Neural Netw* 2020;129:91-102. [DOI](#)
2. Moreno-García CF, Elyan E, Jayne C. New trends on digitisation of complex engineering drawings. *Neural Comput & Applic* 2019;31:1695-712. [DOI](#)
3. Moreno-García CF, Elyan E, Jayne C. Heuristics-based detection to improve text/graphics segmentation in complex engineering drawings. In: Boracchi G, Iliadis L, Jayne C, Likas A, editors. *Engineering applications of neural networks*. Cham: Springer International Publishing; 2017. pp. 87-98.
4. Ali-Gombe A, Elyan E, Jayne C. Fish classification in context of noisy images. In: Boracchi G, Iliadis L, Jayne C, Likas A, editors. *Engineering applications of neural networks*. Cham: Springer International Publishing; 2017. pp. 216-26.
5. Schwab E, Gooßen A, Deshpande H, Saalbach A. Localization of critical findings in chest X-ray without local annotations using multi-instance learning; 2020. [DOI](#)
6. Pomponiu V, Nejati H, Cheung NM. Deepmole: Deep neural networks for skin mole lesion classification. In: 2016 IEEE International Conference on Image Processing (ICIP); 2016. pp. 2623–27. [DOI](#)
7. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115-8. [DOI](#)
8. Levine AB, Schlosser C, Grewal J, Coope R, Jones SJM, Yip S. Rise of the machines: advances in deep learning for cancer diagnosis. *Trends Cancer* 2019;5:157-69. [DOI](#)
9. Schlemper J, Oktay O, Schaap M, et al. Attention gated networks: learning to leverage salient regions in medical images. *Med Image Anal* 2019;53:197-207. [DOI](#)
10. Vyborny CJ, Giger ML. Computer vision and artificial intelligence in mammography. *AJR Am J Roentgenol* 1994;162:699-708. [DOI](#)
11. Gumbs AA, Frigerio I, Spolverato G, et al. Artificial intelligence surgery: how do we get to autonomous actions in surgery? *Sensors (Basel)* 2021;21:5526. [DOI](#)
12. Gonzalez RC, Woods RE. *Digital image processing (3rd Edition)*. USA: Prentice-Hall, Inc.; 2006.
13. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015. pp. 1-9. [DOI](#)
14. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017;60:84-90. [DOI](#)
15. Bishop CM. *Pattern recognition and machine learning*. Springer; 2006.
16. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common

- pigmented skin lesions. *Sci Data* 2018;5:180161. DOI
17. Ismael AM, Şengür A. Deep learning approaches for COVID-19 detection based on chest X-ray images. *Expert Syst Appl* 2021;164:114054. DOI
 18. Lowe DG. Distinctive image features from scale-invariant keypoints. *Inter J Comput Vis* 2004 Nov;60:91–110. DOI
 19. Bay H, Tuytelaars T, Gool LV. SURF: Speeded up robust features. In: *Computer Vision – ECCV 2006*. Springer Berlin Heidelberg; 2006. pp. 404–17. DOI
 20. Hearst M, Dumais S, Osuna E, Platt J, Scholkopf B. Support vector machines. *IEEE Intell Syst Their Appl* 1998;13:18-28. DOI
 21. Breiman L. Random forests. *Machine Learning* 2001;45:5–32. DOI
 22. Amato F, López A, Peña-Méndez EM, Vañhara P, Hampl A, Havel J. Artificial neural networks in medical diagnosis. *J Appl Biomed* 2013;11:47-58. DOI
 23. Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press; 2016. <http://www.deeplearningbook.org>.
 24. Masek J, Burget R, Karasek J, Uher V, Güney S. Evolutionary improved object detector for ultrasound images. In: *2013 36th International Conference on Telecommunications and Signal Processing (TSP)*; 2013. pp. 586–90. DOI
 25. Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. vol. 1; 2001. pp. I–I. DOI
 26. Holcomb SD, Porter WK, Ault SV, Mao G, Wang J. Overview on deepmind and its alphago zero AI. In: *Proceedings of the 2018 International Conference on Big Data and Education. ICBDE '18*. New York, NY, USA: ACM; 2018. pp. 67–71. DOI
 27. Yang Z, Yang D, Dyer C, et al. Hierarchical attention networks for document classification. In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. San Diego, California: Association for Computational Linguistics; 2016. pp. 1480–89. DOI
 28. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med* 2019;25:24-9. DOI
 29. Shone N, Ngoc TN, Phai VD, Shi Q. A deep learning approach to network intrusion detection. *IEEE Trans Emerg Top Comput Intell* 2018;2:41-50. DOI
 30. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng* 2010;22:1345-59. DOI
 31. Hu Z, Tang J, Wang Z, Zhang K, Zhang L, Sun Q. Deep learning for image-based cancer detection and diagnosis – a survey. *Pattern Recognition* 2018;83:134-49. DOI
 32. Aggarwal R, Sounderajah V, Martin G, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *NPJ Digit Med* 2021;4:65. DOI
 33. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. In: *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1. NIPS'15*. Cambridge, MA, USA: MIT Press; 2015. pp. 91–99. Available from: <http://dl.acm.org/citation.cfm?id=2969239.2969250>.
 34. Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector. *CoRR* 2015;abs/1512.02325. Available from: <http://arxiv.org/abs/1512.02325>.
 35. Dai J, Li Y, He K, Sun J. R-FCN: Object Detection via region-based fully convolutional networks. *CoRR* 2016;abs/1605.06409. Available from: <http://arxiv.org/abs/1605.06409>.
 36. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2016. pp. 779–88. DOI
 37. Twinanda AP, Shehata S, Mutter D, Marescaux J, de Mathelin M, Padoy N. EndoNet: a deep architecture for recognition tasks on laparoscopic videos. *IEEE Trans Med Imaging* 2017;36:86–97. DOI
 38. Yu F, Silva Croso G, Kim TS, et al. Assessment of automated identification of phases in videos of cataract surgery using machine learning and deep learning techniques. *JAMA Netw Open* 2019;2:e191860. DOI
 39. Khalid S, Goldenberg M, Grantcharov T, Taati B, Rudzicz F. Evaluation of deep learning models for identifying surgical actions and measuring performance. *JAMA Netw Open* 2020;3:e201664. DOI
 40. Tan L, Huangfu T, Wu L, Chen W. Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification. *BMC Med Inform Decis Mak* 2021;21:324. DOI
 41. Sarikaya D, Corso JJ, Guru KA. Detection and localization of robotic tools in robot-assisted surgery videos using deep neural networks for region proposal and detection. *IEEE Trans Med Imaging* 2017;36:1542-9. DOI
 42. Lee D, Yu HW, Kwon H, Kong HJ, Lee KE, Kim HC. Evaluation of surgical skills during robotic surgery by deep learning-based multiple surgical instrument tracking in training and actual operations. *J Clin Med* 2020;9:1964. DOI
 43. Marban A, Srinivasan V, Samek W, Fernández J, Casals A. Estimating position and velocity in 3D space from monocular video sequences using a deep neural network. In: *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*; 2017. pp. 1460–69. DOI
 44. Wadhwa A, Bhardwaj A, Singh Verma V. A review on brain tumor segmentation of MRI images. *Magn Reson Imaging* 2019;61:247-59. DOI
 45. Shvets AA, Rakhlin A, Kalinin AA, Igloukov VI. Automatic instrument segmentation in robot-assisted surgery using deep learning. In: *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*; 2018. pp. 624–28. DOI
 46. Garcia-Peraza-Herrera LC, Li W, Gruijthuijsen C, et al. Real-time segmentation of non-rigid surgical tools based on deep learning and tracking In: Peters T, Yang GZ, Navab N, et al., editors. *Computer-Assisted and Robotic Endoscopy*. Cham: Springer International Publishing; 2017. pp. 84–95.
 47. Ward TM, Mascagni P, Madani A, Padoy N, Perretta S, Hashimoto DA. Surgical data science and artificial intelligence for surgical

- education. *J Surg Oncol* 2021;124:221-30. DOI
48. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016. DOI
49. Kennedy-Metz LR, Mascagni P, Torralba A, et al. Computer vision in the operating room: opportunities and caveats. *IEEE Trans Med Robot Bionics* 2021;3:2-10. DOI
50. Rajpurkar P, Irvin J, Bagul A, et al. MURA: Large dataset for abnormality detection in musculoskeletal radiographs; 2018. DOI
51. Tajbakhsh N, Shin JY, Gurudu SR, et al. Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Med Imaging* 2016;35:1299-312. DOI
52. Armato III, Samuel G, McLennan G, Bidaut L, et al. Data from LIDC-IDRI. The Cancer Imaging Archive; 2015. DOI
53. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019;17:195. DOI
54. Sarvamangala DR, Kulkarni RV. Convolutional neural networks in medical image understanding: a survey. *Evol Intell* 2021;1:1-22. DOI
55. Sarker MMK, Makhlof Y, Banu SF, Chambon S, Radeva P, Puig D. Web-based efficient dual attention networks to detect COVID-19 from X-ray images. *Electron Lett* 2020;56:1298-301. DOI
56. Yap J, Yolland W, Tschandl P. Multimodal skin lesion classification using deep learning. *Experimental Dermatology* 2018;27:1261-67.
57. Kassem MA, Hosny KM, Damaševićius R, Eltoukhy MM. Machine learning and deep learning methods for skin lesion classification and diagnosis: a systematic review. *Diagnostics (Basel)* 2021;11:1390. DOI
58. Tian P, He B, Mu W, et al. Assessing PD-L1 expression in non-small cell lung cancer and predicting responses to immune checkpoint inhibitors using deep learning on computed tomography images. *Theranostics* 2021;11:2098.
59. Singh VK, Romani S, Rashwan HA, et al. Conditional generative adversarial and convolutional networks for X-ray breast mass segmentation and shape classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2018. pp. 833-40. DOI
60. Hijab A, Rushdi MA, Gomaa MM, Eldeib A. Breast cancer classification in ultrasound images using transfer learning. In: 2019 Fifth International Conference on Advances in Biomedical Engineering (ICABME). IEEE; 2019. pp. 1-4. DOI
61. Abdelaziz Ismael SA, Mohammed A, Hefny H. An enhanced deep learning approach for brain cancer MRI images classification using residual networks. *Artif Intell Med* 2020;102:101779. DOI
62. Hashemzahi R, Mahdavi SJS, Kheirabadi M, Kamel SR. Detection of brain tumors from MRI images base on deep learning using hybrid model CNN and NADE. *Biocybernetics and Biomedical Engineering* 2020;40:1225-32. DOI
63. Qureshi I, Ma J, Abbas Q. Diabetic retinopathy detection and stage classification in eye fundus images using active deep learning. *Multimed Tools Appl* 2021;80:11691-721. DOI
64. Martínez-Murcia FJ, Ortiz A, Ramírez J, Górriz JM, Cruz R. Deep residual transfer learning for automatic diagnosis and grading of diabetic retinopathy. *Neurocomputing* 2021;452:424-34. DOI
65. Li Z, Guo C, Nie D, et al. Automated detection of retinal exudates and drusen in ultra-widefield fundus images based on deep learning. *Eye (Lond)* 2021;1-6. DOI
66. Sarker MMK, Makhlof Y, Craig SG, et al. A means of assessing deep learning-based detection of ICOS protein expression in colon cancer. *Cancers (Basel)* 2021;13:3825. DOI
67. Karimi D, Nir G, Fazli L, Black PC, Goldenberg L, Salcudean SE. Deep learning-based gleason grading of prostate cancer from histopathology images-role of multiscale decision aggregation and data augmentation. *IEEE J Biomed Health Inform* 2020;24:1413-26. DOI
68. Gamble P, Jaroensri R, Wang H, et al. Determining breast cancer biomarker status and associated morphological features using deep learning. *Commun Med* 2021;1:1-12. DOI
69. Ghoniem RM, Algarni AD, Refky B, Ewees AA. Multi-modal evolutionary deep learning model for ovarian cancer diagnosis. *Symmetry* 2021;13:643. DOI
70. Lakshmanaprabu S, Mohanty SN, Shankar K, Arunkumar N, Ramirez G. Optimal deep learning model for classification of lung cancer on CT images. *Future Generation Computer Systems* 2019;92:374-82. DOI
71. Lahiri BB, Bagavathiappan S, Jayakumar T, Philip J. Medical applications of infrared thermography: a review. *Infrared Phys Technol* 2012;55:221-35. DOI
72. Manigandan S, Wu MT, Ponnusamy VK, Raghavendra VB, Pugazhendhi A, Brindhadevi K. A systematic review on recent trends in transmission, diagnosis, prevention and imaging features of COVID-19. *Process Biochem* 2020;98:233-40. DOI
73. Chiu WT, Lin PW, Chiou HY, et al. Infrared thermography to mass-screen suspected SARS patients with fever. *Asia Pac J Public Health* 2005;17:26-8. DOI
74. Zhou Y, Ghassemi P, Chen M, et al. Clinical evaluation of fever-screening thermography: impact of consensus guidelines and facial measurement location. *J Biomed Opt* 2020;25:097002. DOI
75. Ferrari C, Berlincioni L, Bertini M, Del Bimbo A. Inner eye canthus localization for human body temperature screening. In: 2020 25th International Conference on Pattern Recognition (ICPR). IEEE; 2021. pp. 8833-40. DOI
76. Liu X, Dong S, An M, Bai L, Luan J. Quantitative assessment of facial paralysis using infrared thermal imaging. In: 2015 8th International Conference on Biomedical Engineering and Informatics (BMEI). IEEE; 2015. pp. 106-10. DOI
77. Liu X, Feng J, Zhang R, Luan J, Wu Z. Quantitative assessment of Bell's palsy-related facial thermal asymmetry using infrared thermography: a preliminary study. *J Therm Biol* 2021;100:103070. DOI
78. Saxena A, Ng EYK, Lim ST. Infrared (IR) thermography as a potential screening modality for carotid artery stenosis. *Comput Biol Med*

- 2019;113:103419. DOI
79. Hoffmann N, Koch E, Steiner G, Petersohn U, Kirsch M. Learning thermal process representations for intraoperative analysis of cortical perfusion during ischemic strokes. In: *Deep Learning and Data Labeling for Medical Applications*. Springer; 2016. pp. 152–60. DOI
80. Singh D, Singh AK. Role of image thermography in early breast cancer detection- Past, present and future. *Comput Methods Programs Biomed* 2020;183:105074. DOI
81. Ekici S, Jawzal H. Breast cancer diagnosis using thermography and convolutional neural networks. *Med Hypotheses* 2020;137:109542. DOI
82. Zuluaga-gomez J, Al Masry Z, Benaggoune K, Meraghni S, Zerhouni N. A CNN-based methodology for breast cancer diagnosis using thermal images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* 2021;9:131–45. DOI
83. Ordun C, Raff E, Purushotham S. The use of AI for thermal emotion recognition: A review of problems and limitations in standard design and data. *arXiv preprint arXiv:200910589* 2020.
84. Takemiya R, Kido S, Hirano Y, Mabu S. Detection of pulmonary nodules on chest x-ray images using R-CNN. In: *International Forum on Medical Imaging in Asia 2019*. vol. 11050. International Society for Optics and Photonics; 2019. p. 110500W.
85. Banu SF, Sarker M, Kamal M, et al. AWEU-Net: an attention-aware weight excitation u-net for lung nodule segmentation. *Appl Sci* 2021;11:10132. DOI
86. Reiazi R, Paydar R, Ardakani AA, et al. Mammography lesion detection using faster R-CNN detector. In: *CS & IT Conference Proceedings*. vol. 8. CS & IT Conference Proceedings; 2018. DOI
87. Yap MH, Goyal M, Osman F, et al. Breast ultrasound region of interest detection and lesion localisation. *Artif Intell Med* 2020;107:101880. DOI
88. He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. In: *Proceedings of the IEEE international conference on computer vision*; 2017. pp. 2961-69.
89. Law H, Deng J. Cornernet: Detecting objects as paired keypoints. In: *Proceedings of the European conference on computer vision (ECCV)*; 2018. pp. 734–50.
90. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42:60-88. DOI
91. Pang S, Ding T, Qiao S, et al. A novel YOLOv3-arch model for identifying cholelithiasis and classifying gallstones on CT images. *PLoS One* 2019;14:e0217647 DOI
92. Le K, Lou Z, Huo W, Tian X. Auto whole heart segmentation from CT images using an improved Unet-GAN. *J Phys : Conf Ser* 2021;1769:012016. DOI
93. Kim M, Lee BD. Automatic lung segmentation on chest X-rays using self-attention deep neural network. *Sensors (Basel)* 2021;21:369. DOI
94. Ranjbarzadeh R, Bagherian Kasgari A, Jafarzadeh Ghousechi S, Anari S, Naseri M, Bendeche M. Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images. *Sci Rep* 2021;11:10930. DOI
95. Sarker MMK, Rashwan HA, Akram F, et al. SLSNet: Skin lesion segmentation using a lightweight generative adversarial network. *Expert Systems with Applications* 2021;183:115433. DOI
96. Singh VK, Abdel-nasser M, Akram F, et al. Breast tumor segmentation in ultrasound images using contextual-information-aware deep adversarial learning framework. *Expert Systems with Applications* 2020;162:113870. DOI
97. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2015. pp. 3431–40.
98. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer; 2015. pp. 234–41. DOI
99. Luo G, An R, Wang K, Dong S, Zhang H. A deep learning network for right ventricle segmentation in short-axis MRI. In: *2016 Computing in Cardiology Conference (CinC)*; 2016. pp. 485–88. DOI
100. Dang T, Nguyen TT, Moreno-Garcia CF, Elyan E, McCall J. Weighted ensemble of deep learning models based on comprehensive learning particle swarm optimization for medical image segmentation. In: *IEEE Congress on Evolutionary Computing*. IEEE; 2021. pp. 744–51. DOI
101. Miranda E, Aryuni M, Irwansyah E. A survey of medical image classification techniques. In: *2016 International Conference on Information Management and Technology (ICIMTech)*; 2016. pp. 56–61. DOI
102. Brosch T, Tam R, Initiative ADN, et al. Manifold learning of brain MRIs by deep learning. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2013. pp. 633–40. DOI
103. Nie D, Zhang H, Adeli E, Liu L, Shen D. 3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients. In: *International conference on medical image computing and computer-assisted intervention*. Springer; 2016. pp. 212–20. DOI
104. Choi JY, Yoo TK, Seo JG, et al. Multi-categorical deep learning neural network to classify retinal images: A pilot study employing small database. *PLoS one* 2017;12:e0187336. DOI
105. Badar M, Shahzad M, Fraz M. Simultaneous segmentation of multiple retinal pathologies using fully convolutional deep neural network. In: *Annual Conference on Medical Image Understanding and Analysis*. Cham: Springer; 2018. pp. 313–24. DOI
106. Singh VK, Rashwan HA, Saleh A, et al. REFUGE CHALLENGE 2018-Task 2: deep optic disc and cup segmentation in fundus images using U-Net and multi-scale feature matching networks. *arXiv preprint arXiv:180711433* 2018.
107. Cho Y, Kim YG, Lee SM, Seo JB, Kim N. Reproducibility of abnormality detection on chest radiographs using convolutional neural network in paired radiographs obtained within a short-term interval. *Sci Rep* 2020;10:17417. DOI
108. Yoo H, Kim KH, Singh R, Digumarthy SR, Kalra MK. Validation of a deep learning algorithm for the detection of malignant pulmonary

- nodules in chest radiographs. *JAMA Netw Open* 2020;3:e2017135. DOI
109. Wang J, Zhu H, Wang S, Zhang Y. A review of deep learning on medical image analysis. *Mobile Netw Appl* 2021;26:351-80. DOI
110. Hosseini-Asl E, Gimel'farb G, El-Baz A. Alzheimer's disease diagnostics by a deeply supervised adaptable 3D convolutional network. *arXiv preprint arXiv:160700556* 2016.
111. Kleesiek J, Urban G, Hubert A, et al. Deep MRI brain extraction: A 3D convolutional neural network for skull stripping. *Neuroimage* 2016;129:460-9. DOI
112. Ghafoorian M, Karssemeijer N, Heskes T, et al. Deep multi-scale location-aware 3D convolutional neural networks for automated detection of lacunes of presumed vascular origin. *Neuroimage Clin* 2017;14:391-9. DOI
113. Akram F, Singh VK, Rashwan HA, et al. Adversarial learning with multiscale features and kernel factorization for retinal blood vessel segmentation. *arXiv preprint arXiv:190702742* 2019.
114. Su Y, Li D, Chen X. Lung nodule detection based on faster R-CNN framework. *Comput Methods Programs Biomed* 2021;200:105866. DOI
115. Gupta S, Blankstein R. Detecting coronary artery calcium on chest radiographs: can we teach an old dog new tricks? *Radiol Cardiothorac Imaging* 2021;3:e210123. DOI
116. Lu MY, Chen TY, Williamson DFK, et al. AI-based pathology predicts origins for cancers of unknown primary. *Nature* 2021;594:106-10. DOI
117. Lu MY, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng* 2021;5:555-70. DOI
118. Alemi Koochbanani N, Jahanifar M, Zamani Tajadin N, Rajpoot N. NuClick: a deep learning framework for interactive segmentation of microscopic images. *Med Image Anal* 2020;65:101771. DOI
119. Hu Q, Whitney HM, Giger ML. A deep learning methodology for improved breast cancer diagnosis using multiparametric MRI. *Sci Rep* 2020;10:10536. DOI
120. Lai X, Yang W, Li R. DBT masses automatic segmentation using U-net neural networks. *Comput Math Methods Med* 2020;2020:7156165. DOI
121. Li J, Yu ZL, Gu Z, Liu H, Li Y. Dilated-inception net: multi-scale feature aggregation for cardiac right ventricle segmentation. *IEEE Transactions on Biomedical Engineering* 2019;66:3499-508.
122. Zhuang X, Li L, Payer C, et al. Evaluation of algorithms for multi-modality whole heart segmentation: an open-access grand challenge. *Med Image Anal* 2019;58:101537. DOI
123. Bai W, Suzuki H, Qin C, et al. Recurrent neural networks for aortic image sequence segmentation with sparse annotations. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2018. pp. 586-94. DOI
124. Joyce T, Chartsias A, Tsaftaris S. Deep multi-class segmentation without ground-truth labels. In: Medical Imaging with Deep Learning: Amsterdam; 2018.
125. Gruber N, Antholzer S, Jaschke W, Kremser C, Haltmeier M. A joint deep learning approach for automated liver and tumor segmentation. In: 2019 13th International conference on Sampling Theory and Applications (SampTA). IEEE; 2019. pp. 1-5. DOI
126. Kline TL, Korfiatis P, Edwards ME, et al. Performance of an artificial multi-observer deep neural network for fully automated segmentation of polycystic kidneys. *J Digit Imaging* 2017;30:442-8. DOI
127. Gibson E, Giganti F, Hu Y, et al. Automatic multi-organ segmentation on abdominal CT with dense V-networks. *IEEE Trans Med Imaging* 2018;37:1822-34. DOI
128. Zhu Y, Wang QC, Xu MD, 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.] DOI
129. Mahbod A, Schaefer G, Wang C, Dorffner G, Ecker R, Ellinger I. Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification. *Comput Methods Programs Biomed* 2020;193:105475. DOI
130. Lim ZV, Akram F, Ngo CP, et al. Automated grading of acne vulgaris by deep learning with convolutional neural networks. *Skin Res Technol* 2020;26:187-92. DOI
131. Huang K, Jiang Z, Li Y, et al. The classification of six common skin diseases based on xiangya-derm: development of a chinese database for artificial intelligence. *J Med Internet Res* 2021;23:e26025. DOI
132. Hashimoto DA, Rosman G, Witkowski ER, et al. Computer vision analysis of intraoperative video: automated recognition of operative steps in laparoscopic sleeve gastrectomy. *Ann Surg* 2019;270:414-21. DOI
133. Aggarwal N, Garg M, Dwarakanathan V, et al. Diagnostic accuracy of non-contact infrared thermometers and thermal scanners: a systematic review and meta-analysis. *J Travel Med* 2020;27:taaa193. DOI
134. Ulhaq A, Born J, Khan A, Gomes DPS, Chakraborty S, Paul M. Covid-19 control by computer vision approaches: a survey. *IEEE Access* 2020;8:179437-56. DOI
135. 80601-2-59:2017 I. Medical electrical equipment — Part 2-59: Particular requirements for the basic safety and essential performance of screening thermographs for human febrile temperature screening. Geneva, CH: International Organization for Standardization; 2017.
136. Howell KJ, Mercer JB, Smith RE. Infrared thermography for mass fever screening: repeating the mistakes of the past. *J Med Virol* 2020;30:5-6.
137. Vuttipittayamongkol P, Elyan E. Neighbourhood-based undersampling approach for handling imbalanced and overlapped data. *Information Sciences* 2020;509:47-70. DOI
138. Vuttipittayamongkol P, Elyan E. Overlap-based undersampling method for classification of imbalanced medical datasets. In: IFIP International Conference on Artificial Intelligence Applications and Innovations. Cham: Springer International Publishing; 2020. pp.

- 358-69. DOI
139. Maji P, Mondal HK, Roy AP, Poddar S, Mohanty SP. iKardo: an intelligent ECG device for automatic critical beat identification for smart healthcare. *IEEE Trans Consumer Electron* 2021;67:235-43. DOI
140. Vuttipittayamongkol P, Tung A, Elyan E. A data-driven decision support tool for offshore oil and gas decommissioning. *IEEE Access* 2021;9:137063–82. DOI
141. Le T, Hoang Son L, Vo M, Lee M, Baik S. A cluster-based boosting algorithm for bankruptcy prediction in a highly imbalanced dataset. *Symmetry* 2018;10:250. DOI
142. Li Z, Huang M, Liu G, Jiang C. A hybrid method with dynamic weighted entropy for handling the problem of class imbalance with overlap in credit card fraud detection. *Expert Systems with Applications* 2021;175:114750. DOI
143. Vuttipittayamongkol P, Elyan E. Improved overlap-based undersampling for imbalanced dataset classification with application to epilepsy and parkinson's disease. *Int J Neural Syst* 2020;30:2050043. DOI
144. Friedman E, Patino MO, Abdel Razek AAK. MR imaging of salivary gland tumors. *Magn Reson Imaging Clin N Am* 2022;30:135-49. DOI
145. Sun S, Mo JQ, Levy ML, Crawford J. Atypical giant suprasellar prolactinoma presenting with visual field changes in the absence of symptoms of hyperprolactinemia. *Cureus* 2021;13:e19632. DOI
146. Lang SM, Mills AM, Cantrell LA. Malignant Brenner tumor of the ovary: review and case report. *Gynecol Oncol Rep* 2017;22:26-31. DOI
147. Liu T, Fan W, Wu C. A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset. *Artif Intell Med* 2019;101:101723. DOI
148. Zhu M, Xia J, Jin X, et al. Class weights random forest algorithm for processing class imbalanced medical data. *IEEE Access* 2018;6:4641-52. DOI
149. Kalantari A, Kamsin A, Shamshirband S, Gani A, Alinejad-rokny H, Chronopoulos AT. Computational intelligence approaches for classification of medical data: State-of-the-art, future challenges and research directions. *Neurocomputing* 2018;276:2-22. DOI
150. Vuttipittayamongkol P, Elyan E, Petrovski A. On the class overlap problem in imbalanced data classification. *Knowledge-Based Systems* 2021;212:106631. DOI
151. E Elyan, C F Moreno-García, C Jayne. CDSMOTE: class decomposition and synthetic minority class oversampling technique for imbalanced-data classification. *Neural Comput & Applic* 2021;33:2839-51. DOI
152. Luo H, Liao J, Yan X, LiU L. Oversampling by a constraint-based causal network in medical imbalanced data classification. In: 2021 IEEE International Conference on Multimedia and Expo (ICME). IEEE; 2021. pp. 1–6. DOI
153. Khushi M, Shaikat K, Alam TM, et al. A comparative performance analysis of data resampling methods on imbalance medical data. *IEEE Access* 2021;9:109960-75. DOI
154. Rahman MM, Davis DN. Addressing the class imbalance problem in medical datasets. *Inter J Machine Learning Comput* 2013;3:224.
155. Huang C, Huang X, Fang Y, et al. Sample imbalance disease classification model based on association rule feature selection. *Pattern Recognition Letters* 2020;133:280-6. DOI
156. Krawczyk B, Galar M, Jeleń Ł, Herrera F. Evolutionary undersampling boosting for imbalanced classification of breast cancer malignancy. *Applied Soft Computing* 2016;38:714-26. DOI
157. Bria A, Marrocco C, Tortorella F. Addressing class imbalance in deep learning for small lesion detection on medical images. *Comput Biol Med* 2020;120:103735. DOI
158. Yuan X, Xie L, Abouelenien M. A regularized ensemble framework of deep learning for cancer detection from multi-class, imbalanced training data. *Pattern Recognition* 2018;77:160-72. DOI
159. Zhou Q, Ren C, Qi S. An imbalanced R-STDP learning rule in spiking neural networks for medical image classification. *IEEE Access* 2020;8:224162–77. DOI
160. Ali-Gombe A, Elyan E. MFC-GAN: class-imbalanced dataset classification using multiple fake class generative adversarial network. *Neurocomputing* 2019;361:212-21. DOI
161. Ali-Gombe A, Elyan E, Jayne C. Multiple fake classes GAN for data augmentation in face image dataset. In: 2019 International joint conference on neural networks (IJCNN). IEEE; 2019. pp. 1–8. DOI
162. Qasim AB, Ezhov I, Shit S, Set al. Red-GAN: Attacking class imbalance via conditioned generation. Yet another medical imaging perspective. In: Medical Imaging with Deep Learning. PMLR; 2020. pp. 655–68. DOI
163. Rezaei M, Uemura T, Näppi J, Yoshida H, Lippert C, et al. Generative synthetic adversarial network for internal bias correction and handling class imbalance problem in medical image diagnosis. In: Medical Imaging 2020: Computer-Aided Diagnosis. vol. 11314. International Society for Optics and Photonics; 2020. p. 113140E.
164. European Parliament and Council. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation);2016.
165. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng* 2018;2:719-31. DOI
166. Tizhoosh HR, Pantanowitz L. Artificial intelligence and digital pathology: challenges and opportunities. *J Pathol Inform* 2018;9:38. DOI
167. Miller T. Explanation in artificial intelligence: insights from the social sciences. *Artificial Intelligence* 2019;267:1-38. DOI
168. Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak* 2020;20:310. DOI

169. Mohseni S, Zarei N, Ragan ED. A multidisciplinary survey and framework for design and evaluation of explainable AI systems. *ACM Trans Interact Intell Syst* 2021;11:1-45. [DOI](#)
170. Stewart JE, Rybicki FJ, Dwivedi G. Medical specialties involved in artificial intelligence research: is there a leader. *Tas Med J* 2020;2:20-7.
171. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* 2019;1:206-15. [DOI](#)
172. Ribeiro MT, Singh S, Guestrin C. "Why should i trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining; 2016. pp. 1135-44.
173. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* 2017;30:4765-74.
174. Adebayo J, Gilmer J, Muelly M, et al. Sanity checks for saliency maps. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems; 2018. pp. 9525-36.
175. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet Digital Health* 2021;3:e745-50. [DOI](#)
176. Barnett AJ, Schwartz FR, Tao C, et al. Interpretable mammographic image classification using case-based reasoning and deep learning. *arXiv preprint arXiv:210705605* 2021.
177. Zhang Z, Xie Y, Xing F, McGough M, Yang L. Mdnnet: A semantically and visually interpretable medical image diagnosis network. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. pp. 6428-36.
178. Sun J, Darbehani F, Zaidi M, Wang B. Saunet: shape attentive u-net for interpretable medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2020. pp. 797-806.
179. Zhou XY, Guo Y, Shen M, Yang GZ. Application of artificial intelligence in surgery. *Frontiers of Medicine* 2020;14:417-30. [DOI](#)
180. Dodge S, Karam L. A study and comparison of human and deep learning recognition performance under visual distortions. In: 2017 26th International Conference on Computer Communication and Networks (ICCCN); 2017. pp. 1-7. [DOI](#)