REVIEW ARTICLE

The use of machine learning methods for computed tomography image classification in the COVID-19 pandemic: a review

O uso de métodos de aprendizado de máquina para classificação de imagens de tomografia computadorizada na pandemia da COVID-19: uma revisão

El uso de métodos de aprendizaje de máquina para clasificación de imágenes de tomografía computarizada en la pandemia de COVID-19: una revisión

Jacek Sieredziński[1] ORCID 0009-0001-3940-1675
Daniel Zaborski[2] ORCID 0000-0003-3590-3229

[1]109 Military Hospital and Clinic; Piotra Skargi 9-11; 70-965, Szczecin, Poland.
[2]Laboratory of Biostatistics, West Pomeranian University of Technology; Klemensa Janickiego 29; 71-270, Szczecin, Poland.

Address: Klemensa Janickiego 29; 71-270, Szczecin, Poland.
E-mail: daniel.zaborski@zut.edu.pl

ABSTRACT

Background and Objectives: COVID-19 has been declared a pandemic by the World Health Organization, representing a major challenge worldwide. An early diagnosis method for COVID-19 is based on CT scans, which can be analyzed using artificial intelligence to save medical, logistical, and human resources. Therefore, this study aimed to present the current state of the art in the application of machine learning to classify computed tomography images in the COVID-19 pandemic. **Content:** The review briefly describes the types of machine learning methods for COVID-19 detection, the stages of deep learning model construction (segmentation, augmentation), and selected aspects of explainable artificial intelligence. Finally, the application results are discussed and the most common performance indicators for individual models are given. **Conclusion:** Models and algorithms developed during the peak of the COVID-19 pandemic can be reused in the event of future outbreaks of this or similar infectious diseases.

Keywords: *COVID-19. Tomography, X-Ray Computed. Machine Learning. Deep Learning. Neural Networks, Computer.*

RESUMO

Justificativa e Objetivos: A COVID-19 foi declarada uma pandemia pela Organização Mundial da Saúde, representando um grande desafio em todo o mundo. Um método de diagnóstico precoce da COVID-19 é baseado em tomografias computadorizadas, que podem ser analisadas usando inteligência artificial para economizar recursos médicos, logísticos e humanos. Portanto, o objetivo deste estudo foi apresentar o atual estado da arte na aplicação do aprendizado de máquina para classificar imagens de tomografia computadorizada na pandemia de COVID-19. **Conteúdo:** A revisão descreve brevemente os tipos de métodos de aprendizado de máquina para detecção de COVID-19, os estágios de construção do modelo de aprendizagem profunda (segmentação, aumento) e aspectos selecionados da inteligência artificial explicável.

Finalmente, os resultados da aplicação são discutidos e os indicadores de desempenho mais comuns para modelos individuais são dados. **Conclusão:** Modelos e algoritmos desenvolvidos durante o pico da pandemia de COVID-19 podem ser reusados no caso de futuros surtos desta ou doenças infecciosas semelhantes.

Descritores: *COVID-19. Tomografia Computadorizada, Raios X. Aprendizado de Máquina. Aprendizado Profundo. Redes Neurais de Computação.*

RESUMEN

Justificación y Objetivos: La Organización Mundial de la Salud ha declarado que la COVID-19 es una pandemia, lo que ha planteó un gran desafío a nivel mundial. Un método de diagnóstico precoz para COVID-19 se basa en tomografías computarizadas, que pueden analizarse mediante inteligencia artificial para ahorrar recursos médicos, logísticos y humanos. Por lo tanto, el objetivo de este estudio fue presentar el estado actual del arte en la aplicación del aprendizaje automático para clasificar imágenes de tomografía computarizada en la pandemia de COVID-19. **Contenido:** La revisión describe brevemente los tipos de métodos de aprendizaje automático para la detección de COVID-19, las etapas de construcción del modelo de aprendizaje profundo (segmentación, aumento) y aspectos seleccionados de la inteligencia artificial explicable. Finalmente, se discuten los resultados de la aplicación y se presentan los indicadores de rendimiento más comunes para modelos individuales. **Conclusión:** Los modelos y algoritmos desarrollados durante el pico de la pandemia de COVID-19 pueden reutilizarse en caso de futuros brotes de esta o de enfermedades infecciosas similares.

Palabras Clave: *COVID-19. Tomografía Computarizada, Rayos X. Aprendizaje Automático. Aprendizaje Profundo. Redes Neurales de la Computación.*

INTRODUCTION

The first human cases of coronavirus disease 19 (COVID-19) were reported in Wuhan City, China, in December 2019.[1-3] The COVID-19 pandemic was declared on March 11, 2020, by the World Health Organization.[4,5] As of November 1, 2023, 771,548,954 cases and 6,974,460 deaths have been confirmed, ranking COVID-19 fifth among the deadliest epidemics and pandemics in history.[4]

Widely accepted management strategies to restrict the spread of COVID-19 have included lockdowns, travel restrictions, quarantines, social distancing, isolation, infection control measures, and vaccination.[5-7] Different drug types have also been developed and many substances with other indications have been "repurposed" to treat patients with COVID-19.[4] However, the emergence of new worrying variants has become a major problem in the efficient prevention and treatment of the infection.[8] SARS-CoV-2 may cause no symptoms, only mild symptoms such as cramps and fever, or serious complications such as shortness of breath and kidney failure.[3] The risk of severe disease is also higher for older people and for those with underlying conditions, such as diabetes and cancer.[2]

Real-time reverse transcription-polymerase chain reaction (rRT-PCR) is currently the diagnostic gold standard used to confirm COVID-19 infection.[8,9] However, the method is expensive, laborious, time-consuming, requires well-trained personnel to perform sophisticated procedures, and has a relatively low positive detection rate in the early stage.[1,10-15] Furthermore, new genetic variants of SARS-CoV-2 may lead to false-negative results.[16] An early diagnostic method for COVID-19 is based on computed tomography (CT) scans,[1,5,6,10,11,13,17,18] which provide a higher sensitivity rate (88-98%) than RT-PCR (59-71%).[19] Compared with X rays, CT generates more detailed cross-sectional images without tissue overlap, has higher sensitivity and specificity, and can distinguish between COVID-19 and other conditions, such as pneumonia.[2,8,9,12,16] Indeed, CT provides 3D examinations of organs from multiple angles and allows the severity of the infection to be assessed.[6] Three main types of COVID-19-related irregularities have been identified on lung CT images: ground-glass opacification, consolidation, and pleural effusion.[1,9,11,12] To further improve CT analysis, artificial intelligence (AI) can be used,[1,12,20] saving time as well as medical, logistical and human resources,[2,3,8,11] facilitating the detection, classification, diagnosis, segmentation, prediction, and improvement of image quality.[5,20,21]

Therefore, our study aimed to present the current state of the art in the application of machine learning to classify computed tomography images in the COVID-19 period.

METHODS

This narrative review was conducted to assess the literature with a focus on machine learning methods and their use to classify CT images during the COVID-19 pandemic, not to answer a specific research question. This review gathered a group of literature articles on the above-mentioned topic in a qualitative manner. In addition, a quantitative analysis of the literature or its quality was not the main aim of this study. The selection of articles was based on the following inclusion and exclusion criteria.

Eligibility criteria

Only full-text articles on applying machine learning methods to COVID-19 detection based on CT scans were included. The selected articles were published in English between January 1, 2021, and December 31, 2023.

Exclusion criteria

Preprints, conference abstracts, books, book chapters, notes, technical reports, as well as studies not addressing the scientific knowledge about applying machine learning methods to detect COVID-19 based on CT scans were excluded.

Information source and search strategy

The following query was used for searching PubMed (November 24th, 2023): machine learning AND computed tomography AND image classification AND COVID-19.

Selection of studies

Articles that appeared to meet the inclusion criteria were selected for full reading to determine their eligibility. Supplementary articles were included after checking their reference lists.

Data collection

The initial number of articles was 213 but it was reduced to 60 after applying exclusion criteria. The thorough reading and critical evaluation of article content resulted in the selection of the 40 most relevant articles (Figure 1).
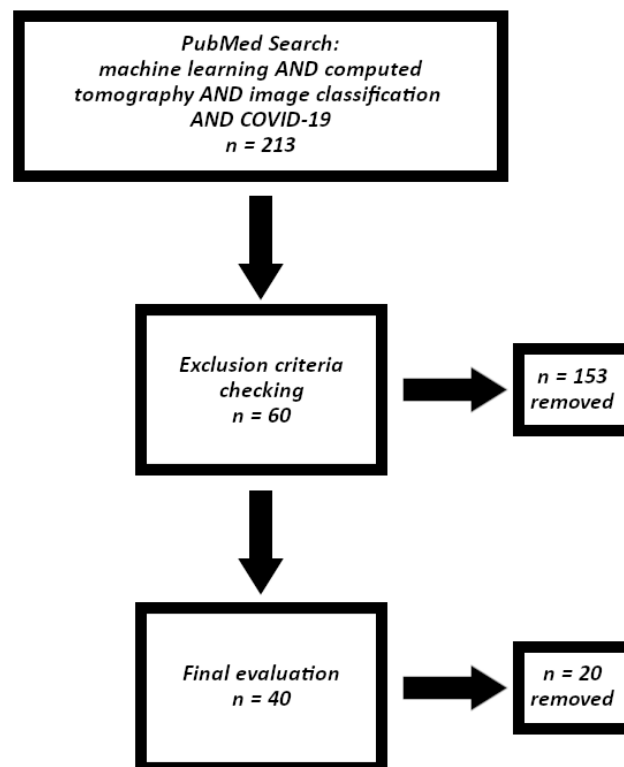
Figure 1. The procedure of article selection (studies from around the world (2019-2023).

RESULTS

Segmentation and augmentation

Among the five models (U-Net, LinkNet, R2U-Net, Attention U-Net, and U-Net[++]),[12] the highest values were achieved by LinkNet for Dice coefficient (DC) and intersection over union (IoU) for lung segmentation (0.980 and 0.967, respectively), whereas R2U-Net showed the lowest values (0.962 and 0.928, respectively).[9] The lung area was also segmented from the small cohort of CT images with BCDU-Net,[22] which was inspired by U-Net[23] and involved bi-directional convolutional long short-term memory (ConvLSTM) with densely connected convolutions. In other studies, candidate infected regions were segmented from pulmonary CT images, using a 3D deep learning (DL) model (region proposal network)[14] or Visual Basic NET (VB-Net), followed by various classification methods [convolutional neural networks (CNN) and inception network or random forest (RF)].[13] The authors developed a VB-Net algorithm, which combined the V-Net model with the bottleneck layer, thus integrating the fine-grained COVID-19 image features, reducing the number of feature mapping channels, and effectively increasing the convolution speed. Dynamic fusion segmentation network (DFSN) is another image segmentation method,[18] whose IoU and DC values were 0.800 and 0.530, respectively. The first component of this system automatically segmented infection-related pixels and served as the backbone to extract dynamically selected pixel-level information, which was used to make a final diagnosis. Other authors[24] used a semi-supervised lung infection segmentation deep network (Inf-Net) for chest CT images, including a parallel partial decoder to aggregate high-level features.[10] They obtained a slightly lower accuracy for non-infected CT regions and applied an additional classifier to improve the overall model performance.

Lung-lesion maps were obtained from input images processed by different segmentation networks (U-net, DRUNET, FCN, SegNet, and DeepLabv3).[5,25] Pre-trained 2D UNet,[26] unsupervised lung segmentation (Shift3D),[27] entire-lung segmentation (followed by resizing, bin discretization, and radiomic feature extraction),[28] k-means clustering with gray level co-occurrence matrices (for extracting regions of interest and textural features),[29] and a segmentation network within the DL framework (for segmenting lung and lesion areas, thus extracting spatiotemporal information from multiple CT scans to perform auxiliary diagnosis)[30] were also used for image segmentation. In another study,[31] over-segmentation mean shift was followed by a superpixel-simple linear iterative clustering algorithm for pulmonary parenchyma segmentation. Each superpixel cluster was described according to its position, grey intensity, second-order texture, and spatial-context-saliency features. Subsequently, the watershed segmentation was applied to the mean-shift clusters to identify ground-glass opacity and pulmonary infiltrates only in the pulmonary parenchyma segmentation-indicated zones.

Application of the EfficientNet and EfficientDet networks[19] yielded DC values of 0.980 and 0.730 for lung and COVID-19 segmentation, respectively, whereas a DC of 0.590 was reported for a Unet-like architecture with backbone residual network (ResNet-34).[32] Finally, a DC of 0.575 was obtained using a weakly-supervised method based on a generative adversarial network (GAN),[33] whereas a multitask model outperformed individual segmentation models for the joint segmentation of pulmonary lesions.[34]

To prevent overfitting, data augmentation and transfer learning (TL) can be used. The former includes translation, horizontal (and vertical) flipping, and random rotation to enhance the accuracy of model prediction.[5] Augmentation may reduce class imbalance or data scarcity problems.[5,10] Some authors[3,9] applied facile image transformation (scaling, rotation, and flipping) resources to increase the number of records, whereas others[35] improved the representational learning capability by distortion, painting, and perspective transformation. Finally, GAN was used in two studies on data augmentation.[33,36] The first one involved GAN hyperparameter tuning with the whale optimization algorithm to avoid overfitting and instability, whereas the second one used image-level labels to generate normal-looking CT slices (from those with COVID-19 lesions), whose reality was improved with a feature match strategy.

Classification

An open-source framework consisting of several DL algorithms differentiated COVID-19 from community-acquired pneumonia and other lung diseases.[22] It could deal with heterogeneous data and small sample sizes irrespective of the CT image source. To increase accuracy and decrease logarithmic loss and testing time, another study used augmented data to train CNN and ConvLSTM-based DL models. They were compared with traditional machine learning (ML) models [support vector machines (SVM) and k-nearest neighbors (k-NN)], and their performance was lower. [3] COVID-19 probability was also predicted using a weakly supervised DL model based on 3D CT volumes from the segmented 3D lung regions.[26] Lung lesions were determined from activation regions in a classification network and unsupervised connected components.

An infection size-aware RF automatically rated patients into classes with the different lesion ranges using the thin-section CT image records for COVID-19 and community-acquired pneumonia.[13] Model performance was further increased by including radiomic features. Another method distinguished COVID-19 from common pneumonia based on lung vessel morphology.[9] It used maximum intensity projection to indicate small-density changes in CT

scans, thus accurately reflecting blood vessel condition and calcification of their walls. The applied capsule network used the DenseNet-121 feature extractor and outperformed ResNet-50 and Inception-V3. Community-acquired pneumonia and other non-pneumonic images were also analyzed with a 2D CNN (COVNet), which extracted visual features from volumetric chest CT scans.[23] Input CT slices were fed to a pre-trained ResNet50 to obtain features, which were then combined and processed by a fully connected layer. To increase the contrast between the local lesion regions and the abdominal cavity, another deep CNN-based classification algorithm performed convolution and deconvolution operations.[11] Moreover, discrimination between image types was improved with middle-level features, and they were classified in each channel using a modified open-source COVID-CT dataset.

One of the DL architectures (ResNet-18) distinguished among COVID-19, influenza, and normal subjects.[14] Segmented images were categorized with their corresponding confidence scores using a location-attention classification model. Another ResNet-18 architecture was trained on a large CT dataset for differentiating COVID-19 and other types of viral pneumonia.[25] This system involved segmentation, classification, and quantitative measurements. However, it required manually segmented images and multi-modal data that were difficult to obtain. COVID-19 was also differentiated from common pneumonia and healthy subjects by using a dynamic transfer-learning classification network in which dynamically selected pixel-level information was used for the final diagnosis.[18]

Features extracted by several CNN models (AlexNet, ResNet18, ResNet50, Inceptionv3, Densenet201, Inceptionresnetv2, MobileNetv2, GoogleNet) from the images stored in the COVID-19 Radiography Database were fed to the traditional ML models [SVM, k-NN, naïve Bayes (NB) and decision trees (DT)]. Their hyperparameters were determined with Bayesian optimization.[5] A pretrained InceptionV3 model was also developed for feature extraction and classification using the SARS-CoV-2 CT-Scan dataset.[36] Four different data [University of Texas (Southwestern Medical Center), China Consortium of Chest CT Image Investigation (CC-CCII), COVID-CT set, and MosMedData] sources were used for training DL models. Their best performance was obtained with multiple 3D CT datasets whose classification accuracy decreased when evaluated on an external set without lung field segmentation.[35] In another study,[12] datasets of COVID-19 were distinguished from those of community-acquired pneumonia with a pipeline (including a capsule network with the DenseNet121 block) consisting of four connected modules for lesion slice selection and slice- and patient-level prediction.

A multitask learning framework (involving task prioritization, convergence acceleration, and joint learning performance improvement) automatically classified CT images into COVID-19 positive or negative cases using a random-weighted loss function.[27] COVID-19 was detected with 3D CNN and an auxiliary feed-forward ANN based on chest CT scans and RT-PCR results. Clinical metadata also helped with distinguishing between COVID-19 and other viral pneumonia in a patient-level method (including InceptionResnetV2), which aggregated chest CT volumes into 2D representations.[34] A combination of features from chest CT volumes improved model performance compared with clinical data alone. Other DL models (AlexNet, ResNet50, and SqueezeNet) were also compared with the traditional ML ones (NB, bagging, and Reptree). They classified CT images into two categories (COVID and non-COVID),[29] whereas a custom 3D CNN trained on the CT scans from patients with suspected or known COVID-19 assigned images to three groups (COVID-19, other type of pulmonary infection or lack of infection signs).[32] More classes (severe-, moderate-, mild-, and non-pneumonic patients) were included in a multinomial logistic regression model, which was trained on the CT radiomic features selected by two feature selection algorithms (RF and multivariate adaptive regression splines).[28]

Automatic systems trained on multiple COVID-19 CT images were developed for COVID-19 detection (using spatiotemporal information fusion)[30] or identification of ground-glass opacity, and pulmonary infiltrates to assess disease progression during the patient's follow-up assessment and evaluation.[31] Differently, thousands of labeled CT images were used for a COVID-19 decision support and segmentation system (involving the EfficientNet and EfficientDet networks), which rejected non-related images using a header analysis and classifiers.[19]

Performance indicators for the models included in this review presented different values (for studies with two or more models, only that with the maximum sensitivity is mentioned) (Table 1).

Table 1. Performance indicators for COVID-19 detection models (studies from around the world, 2019-2023).

Author	Country	Objectives	Methods	Main results	Se (Re)	Sp	PPV (Pr)	NPV	Acc	F1	MCC	AUC
Sedik et al. (2020)[3]	Egypt	To improve the learnability of CNN and the convolutional long short-term memory-based DL models and increase the accuracy of COVID-19 detection.	Two data-augmentation techniques based on simple image transformations and generative adversarial networks.	Acc, logarithmic loss, and testing time were improved relative to DL models without data augmentation; an increased Acc (4-11%) was observed between data-augmented DL models and other investigated ML techniques.	0.997	0.987	0.987	0.997	1.000	0.990	0.984	0.990
Aslan et al. (2022)[5]	Turkey	To classify CT chest images from the COVID-19 Radiography Database and determine hyperparameters of ML algorithms.	Automatic lung segmentation with ANN; data augmentation, feature extraction with CNN, classification with support vector machines, k-nearest neighbors, naive Bayes, and decision trees; hyperparameter determination with Bayesian optimization.	DenseNet201 model and support vector machines showed the best predictive performance.	0.964	0.981	0.964	-	0.963	0.945	0.964	-
Wu et al. (2023)[9]	China	To accurately and automatically distinguish between COVID-19 and CAP using DL.	A DL method was based on maximum-intensity projection images (obtained from CT scans); they served as inputs into a capsule network trained and validated on 333 and 3581 CT scans, respectively.	LinkNet achieved the highest DC; the capsule network with the DenseNet-121 feature extractor outperformed ResNet-50 and Inception-V3; Acc decreased to 0.857 and 0.818 without maximum-intensity projection or capsule	0.971	0.968	0.971	-	0.970	-	-	0.986

network, respectively; Acc of 0.961, 0.997, and 0.949 were achieved on the external validation datasets; Se was higher than or comparable to other state-of-the-art methods.

Author	Country	Aim	Method	Results								
Qi et al. (2022)[12]	China	To improve existing ML methods for distinguishing between COVID-19 and CAP based on CT images.	A fully automatic DL pipeline comprising four connected modules (for lung segmentation, slice selection, and slice- and patient-level prediction) was trained and tested on 326 CT scans; its generalization capability was evaluated on a public dataset of 110 patients.	LinkNet exhibited the largest IoU and DC; the capsule network with ResNet50 achieved an Acc of 0.925 and AUC of 0.933 in the selection of slices with lesions; the capsule network with DenseNet121 showed an Acc of 0.971 and AUC of 0.992 for slice-level prediction; Acc of 1.000 was obtained for patient-level prediction.	0.997	0.966	0.965	-	0.981	-	-	0.983
Shi et al. (2021)[13]	China	To rapidly and accurately screen patients with COVID-19 and CAP using ML.	COVID-19 (1658) and CAP (1027) patients underwent thin-section CT; segmentation of infection and lung fields were used to extract location-specific features; a random forest categorized patients with different ranges of infected lesion sizes and classified them within each group.	Large performance margins were achieved against comparison methods, especially for medium infection size (0.01% to 10%); the inclusion of radiomic features slightly improved classification results.	0.907	0.833	-	-	0.879	-	-	0.942

Xu *et al.* (2020)[14]	China	To establish an early screening model for distinguishing COVID-19 from influenza-A viral pneumonia and healthy cases based on pulmonary CT images and DL.	Different numbers of samples were used for COVID-19 (219), influenza-A viral pneumonia (224), and healthy subjects (175); infection regions were determined using a 3D DL model; separated images were categorized with the corresponding confidence scores using a location-attention classification model; infection types and confidence scores were calculated using the noisy-OR Bayesian function.	The overall Acc on the benchmark dataset was 86.7% for all CT cases taken together.	0.900	-	0.931	-	0.867	0.915	-	-
Zhang *et al.* (2022)[18]	China, UK, Belgium	To automatically segment lesions in CT images and distinguish COVID-19 in common pneumonia patients and healthy subjects.	A dynamic fusion segmentation network segmented infection-related pixels and aggregated low-level features that were fused to model multi-scale semantic information; COVID-19 patients were identified with a dynamic transfer-learning classification network.	Two models achieved state-of-the-art performance in segmentation and classification tasks.	0.980	-	0.820	-	0.770	-	-	-
Carmo *et al.* (2021)[19]	Brazil	To develop and deploy a COVID-19 decision support and segmentation system	EfficientNet and EfficientDet segmented and classified images in a real-time scalable	Acc values of 0.94 and 0.98 were achieved for CT and X-ray classification, respectively, whereas	0.953	-	0.905	-	0.944	0.928	-	0.979

		based on CT and X-ray images.	manner in communication with a Picture Archiving and Communication System; non-related images were rejected using header analysis and classifiers.	those DC for lung and COVID-19 segmentation were 0.98 and 0.73, respectively; the median response times were 7 s for X-ray and 4 min for CT.							
Shiri et al. (2021)[20]	Iran, Switzerland, Canada, The Netherlands, Denmark	To develop prognostic survival models for COVID-19 patients using clinical data and lung and/or lesion radiomic features extracted from chest CT images.	Survival modeling was based on radiomic features and clinical data (separately or in combination); the maximum-relevance minimum-redundancy method and XGBoost were used for feature selection and classification.	Cancer comorbidity, consciousness level, and radiological score were highly correlated with survival; oxygen saturation and blood urea nitrogen were important clinical features; small-area high-gray-level emphasis and high-gray level-zone emphasis from gray-level size-zone matrix, run-length non-uniformity from gray-level run-length matrix, and high-gray-level-zone emphasis from gray-level size-zone matrix yielded the highest predictive performance; the most accurate prognostic model included combined lung, lesion, and clinical features.	0.880	0.890	-	-	0.880	-	-	0.950
Javaheri et al. (2021)[22]	Iran, USA, Canada, Vietnam	To enhance the Acc of CT image-based COVID-19 recognition.	CovidCTNet (an open-source framework) differentiated COVID-19 from CAP	CovidCTNet increased the Acc of CT image-based COVID-19 detection to 95% compared with	0.909	1.000	-	-	0.933	-	-	0.940

Study	Country	Objective	Method	Findings								
			and other lung diseases.	radiologist evaluation (70%) and was independent of the CT imaging hardware.								
Li et al. (2020)[23]	China	To develop a fully automatic framework for COVID-19 detection using chest CT scans.	COVID-19 detection neural network (COVNet) extracted visual features from 4352 volumetric chest CT scans obtained from 3322 patients.	COVNet showed a high predictive performance on the independent test set.	0.900	0.960	-	-	-	-	-	0.960
Zhang et al. (2020)[25]	China	To develop an AI system for diagnosing COVID-19 pneumonia and differentiating it from other common types of pneumonia and normal controls.	AI system identified clinical markers correlated with COVID-19 lesion properties and provided accurate clinical prognosis together with clinical data.	Globally available AI systems showed high predictive performance.	0.949	0.911	-	-	0.925	-	-	0.980
Wang et al. (2020)[26]	China	To develop a DL-based model for automatic COVID-19 diagnosis based on chest CT images.	A weakly-supervised DL framework used 3D CT volumes for COVID-19 classification and lesion localization; the UNet-segmented 3D lung regions were fed into a 3D DL network; CT volumes were used for training (499) and testing (131).	The algorithm took only 1.93 s to process a single patient's CT volume; the weakly-supervised DL model could accurately predict COVID-19 without lesion annotation.	0.907	0.911	0.840	0.982	0.901	-	-	0.959
Bao et al. (2022)[27]	China, Australia	To develop an end-to-end multitask learning framework (COVID-MTL) capable of automated and simultaneous detection and severity	COVID-MTL learned different COVID-19 tasks in parallel through the random-weighted loss function; the 3D real-time augmentation	COVID-MTL achieved high performance in the detection of COVID-19 against radiology and nucleic acid tests, outperforming other state-of-the-art models;	0.902	-	0.912	-	0.902	0.905	-	0.939

			algorithm (Shift3D) introduced space variances for 3D CNN components; MTL accelerated convergence and improved joint learning performance compared to single-task models; COVID-MTL was trained on 930 CT scans and tested on 399 cases.	COVID-MTL yielded AUC of 0.800 and 0.813 for classifying control and/or suspected, mild and/or regular, and severe and/or critically-ill cases.							
Guhan *et al.* (2022)[29]	India, Saudi Arabia	To segment the CT images using k-means clustering, extract textural features using gray level co-occurrence matrix, and classify CT scans using ML.	One hundred COVID-19 and non-COVID-19 images were segmented and classified with naive Bayes, bagging, and REPTree; pre-trained AlexNet, ResNet50, and SqueezeNet were used for predictive performance comparison.	Naive Bayes and ResNet50 achieved the highest Acc (97.0% and 99.0%, respectively).	0.990	-	0.980	-	0.991	-	-	0.990
Li *et al.* (2021)[30]	China	To automatically detect COVID-19 based on spatiotemporal information fusion.	The spatiotemporal information features of multiple CT scans were extracted using a segmentation network to perform auxiliary diagnosis.	High predictive performance was achieved in the classification of COVID-19 and non-COVID-19 CT scans; each scan took about 30 s for detection.	0.953	-	0.967	-	0.944	0.960	-	0.946
Tello-Mijares *et al.* (2021)[31]	Mexico	To automatically identify ground-glass opacity and pulmonary infiltrates in CT images from COVID-19 patients	Oversegmentation mean-shift followed by superpixel-simple linear iterative clustering was applied to COVID-19 CT	Pulmonary parenchyma identification had a precision and recall of over 92.0% on twofold cross-validation; pulmonary infiltrate	0.968	-	0.967	-	-	0.967	-	0.983

		and assess disease progression during the patient's follow-up evaluation.	images for pulmonary parenchyma segmentation.	identification for ground-glass opacity showed a precision and recall of 96.0%.									
Topff *et al.* (2023)[32]	The Netherlands, Spain, Belgium	To develop a DL-based clinical decision support system for the automatic diagnosis of COVID-19 on chest CT scans and construct a complementary segmentation tool for assessing the extent of lung involvement and measuring disease severity.	Data annotation was performed by 34 radiologists and/or radiology residents including quality control measures; 2,802 CT scans were ranked with a multi-class classification model created using a 3D CNN; an UNET-like architecture with a backbone Residual Network (ResNet-34) was selected for image segmentation.	The diagnostic multiclassification model yielded high micro-average and macro-average values for AUC (0.93 and 0.91, respectively) on the external test dataset; the segmentation performance was moderate (DC=0.59).	0.870	0.940	0.950	0.830	0.900	-	-	0.830
Yang *et al.* (2021)[33]	China, Taiwan	To localize COVID-19 lesions with a weakly-supervised method based on a generative adversarial network using only image-level labels.	A generative adversarial network-based framework generated normal-looking CT slices from CT slices with COVID-19 lesions; a feature match strategy improved the quality of generated images; the localization map of lesions was obtained by subtracting the output image from its corresponding input image; a diagnostic system with improved classification Acc was	The weakly-supervised learning method obtained a DC of 0.575 and exceeded other widely used weakly-supervised object localization approaches; its performance was similar to that of fully supervised learning methods in the COVID-19 lesion segmentation task (DC of 0.575); the common severity cohort had the largest sample size as well as the highest visual score.	0.647	0.929	-	-	0.884	0.640	-	0.883

Author	Country	Objective	Method	Results							
Ortiz *et al.* (2022)[34]	USA	To assess the value of aggregated chest CT data for COVID-19 prognosis compared to clinical metadata alone.	obtained by adding a classifier branch to the generative adversarial network-based framework. A patient-level algorithm aggregated chest CT volumes into 2D representations that were integrated with clinical metadata to distinguish COVID-19 patients from healthy participants and patients with other viral pneumonia; the multitask segmentation approach was compared to combining feature-agnostic volumetric CT classification feature maps with clinical metadata for predicting mortality.	A multitask model for joint segmentation of different classes of pulmonary lesions present in COVID-19-infected lungs outperformed individual segmentation models for each task; a combination of features derived from chest CT volumes improved AUC values to 0.80 from 0.52 obtained by using only patients' clinical data.	0.590	-	0.690	-	0.920	0.750	-	0.810
Goel *et al.* (2021)[36]	India, Australia, Korea	To generate CT images using a generative adversarial network and optimize its hyperparameters using the whale optimization algorithm.	The method was tested with different classification and meta-heuristic algorithms using the SARS-CoV-2 CT-Scan dataset, consisting of COVID-19 and non-COVID-19 images.	The performance of the optimized model was better than that of other state-of-the-art methods.	0.998	0.978	0.978	0.998	0.992	0.988	-	-

Abbreviations: **Acc**: accuracy; **ANN**: artificial neural network; **AUC**: area under the curve; **CAP**: community-acquired pneumonia; **CNN**: convolutional neural network; **CT**: computed tomography; **DC**: Dice coefficient; **DL**: deep learning; **F1**: F_1-score; **IoU**: intersection over union; **MCC**: Matthew correlation coefficient; **ML**: machine learning; **NPV**: negative predictive value; **PPV** (Pr): positive predictive value (precision); **Se** (Re): sensitivity (recall); **Sp**: specificity.

DISCUSSION

Types of methods

Machine learning, which belongs to the AI domain, can generally be divided into "traditional methods" and deep learning (both of which can be applied for pattern recognition, regression, or classification).[18] The difference lies in the way images are pre-processed, among other things. Whereas the first group relies on expert-derived inputs (such as the average greyscale) that require human involvement, the second uses the whole images as inputs and extracts the features by itself.[5-7,10,11,16,21] It can be successfully used for medical-related imaging tasks, such as image preprocessing, registration, detection, and segmentation.[6] In the context of COVID-19, DL has been applied at the molecular (*e.g.*, protein structure prediction), patient (*e.g.*, medical imaging for diagnosis), and population (*e.g.*, epidemiology) scales.[18] Deep learning, as a data-driven approach, performs classification based on the image features learned by a model during the training stage.[6,8]

It usually involves the type of artificial neural network (ANN), also called convolutional neural network (CNN). They have gained much popularity due to their higher performance in automatic disease detection tasks.[5,6,11,16] Other DL methods include recurrent neural networks, deep belief networks, and reinforcement learning.[10] One of the CNN architectures (named AlexNet, with fully supervised learning) achieved excellent performance on highly challenging datasets. It was the winner of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012.[6,16] A wide range of ANN settings and training skills (ReLU, dropout, pooling, and local response normalization) enabled more effective CNN training and better performance.[6,37] It has been used in many studies on COVID-19 detection that mainly differed in the feature selection method and training of multiple classifiers. Since AlexNet was created, more advanced pre-trained networks based on this architecture (VGG, GoogLeNet, ResNet, DenseNet, MobileNet, SqueezeNet, and Network in Network) have been applied to COVID-19 detection.[5,7,16] Visual Geometry Group, which is simple in architecture but effective in performance, was the winner of the ILSVRC challenge in 2014.[6] ResNet and DenseNet both use residual blocks and skip connections to make image-level classification. They also employ attention mechanisms, multi-view presentation learning, and semi-supervision because high-level features tend to lose details of the input image and the above-mentioned methods may fail in complex imaging data.[18]

Pretrained networks can be reused in the process called transfer learning (TL).[10] The trained model can be transferred to a new one, for which additional training data may be

provided and in which modified neural layers can be incorporated.[16] After automatic feature extraction (using TL with pre-trained models or custom CNN developed from scratch), ML methods (such as k-NN, SVM, DT, or NB) can be used to classify these features as COVID-19 or non-COVID-19 (*e.g.*, normal or viral pneumonia).[5,6]

Deep learning stages

The DL algorithm may include several steps, such as pre-processing, segmentation, feature extraction, classification, performance evaluation, and explainable model prediction.[6,10] Preprocessing [2] is the first stage in CT image analysis, for which different techniques are used. In preprocessing, raw images are converted into an appropriate format for further analysis. Medical images collected from different devices can vary in size, slice thickness, and the number of scans (*e.g.*, 60-70 in CT).[6] During preprocessing, resizing, normalization, and sometimes conversion from RGB to grayscale are performed.[16] In addition, the voxel dimension is resampled to account for the variation across datasets (resampling to an isomorphic resolution). Images are also improved with smoothing to increase the signal-to-noise ratio.

Segmentation is the next step of image preprocessing, for which a full CNN and its variants have been used.[1,6] An image that shows only the lungs is more appropriate for infection detection. This is probably because it prevents the model from focusing on unwanted targets like bone and soft tissue.[8] To achieve this, the lung region must be segmented from the raw image, which enables a more successful diagnosis. The lung area of the original image is cut by the segmentation process.[5] Sometimes, pixel values are also limited to obtain a proper range of Hounsfield units in the lung image.[6] In segmentation, underused multi-scale context information, high variance in texture, size and position of infected regions, and small inter-class variance of lesions are potential challenges.[18] Manual lung segmentation is laborious, tedious, time-consuming, and heavily depends on the radiologists' knowledge and experience.[6] However, DL-based segmentation techniques can automatically identify infected regions, thus allowing rapid screening of COVID-19 images. Classic U-Net, UNet[++], and VB-Net are the popular segmentation methods.[2,6,10]

Of all DL models, U-Net is the most famous architecture for segmentation, whose results may also be affected by image type. For example, two different segmentation approaches were used for the NIFTI and DICOM CT lung images as no method works for all image formats.[8]

Dice coefficient (DC) and intersection over union (IoU) are the two common measures for evaluating segmentation effectiveness.[18] The first one is defined as:[38]

$$DC = \frac{2|A \cap B|}{|A| + |B|},$$

where *A* is a set that represents the ground truth and *B* represents the computed segmentation.

IoU, also known as the Jaccard index, is the most commonly used metric for comparing the similarity between two arbitrary shapes.[39] It encodes the shape properties of the objects under comparison into the region property and calculates a normalized measure with a focus on their areas (or volumes). It is given by the following formula:

$$IoU = \frac{|A \cap B|}{|A \cup B|}.$$

After segmentation, augmentation is employed to increase the segmented image count, thus providing data diversity.[5,16] Rotation, shifting in the width and height dimensions, shearing, zooming, flipping in the horizontal and vertical axes, and brightness changing can be used for this purpose.[10]

Explainable artificial intelligence

Deep learning black-box models provide no evidence of correctly extracted features. On the other hand, explainable AI is an emerging field that assigns certain values to image regions leading to the predicted outcome. Thus, radiologists can locate abnormalities in the lungs and have an insight into the important areas responsible for image classification.[6] According to some authors,[21] CT was the second most common (20.0%) image modality coupled with explainable AI, although other studies reported a combined application to CT and X rays. It should be noted that the performance of COVID-19 detection models can be further improved by incorporating both kinds of images (chest X-ray or CT).[6] Explainable AI has most often been applied to lung examination and used different publicly available data repositories of CT images for COVID-19 diagnosis (Kaggle, Signal Processing Grand Challenge on COVID-19 dataset, COVIDx CT, COVIDx CT-2A & COVIDx CT-2B, CC-CCII, MosMedData, COVID-Ctset, LTRC dataset, CT Chest Images Dataset from Mendeley, COVID pandemic, iRoads, Caltech-256, and Caltech-101). The availability of such repositories was the main reason for the advancement of COVID-19 studies among those using explainable AI.

Supervised *vs*. unsupervised learning

Further division of ML is based on the role of a "teacher" or "trainer": in supervised learning, a loss function is optimized considering predicted labels and ground truth requiring manual annotation; in unsupervised learning, data patterns are found automatically using

clustering.[2] To achieve the best performance, all ML methods must be configured before the training process using hyperparameter optimization.[5,16] Hyperparameters differ from model parameters: the former (such as the number of ANN layers, size, shape, type, number of neurons, intermediate processing elements, etc.) are calculated before the training phase, whereas the latter (such as weights) are optimized during learning. There are several ways to set the hyperparameters and different strategies can be adopted (including a manual one). Many algorithms, such as Bayesian optimization, grid search, swarm optimization (*e.g.*, Sparrow optimization algorithm), etc., can be used to search the optimal hyperparameter.[16]

Performance indicators

The most frequently reported model performance indicators are as follows: sensitivity (or recall; *Se*), specificity (*Sp*), accuracy (*Acc*), positive predictive value (or precision; *PPV*), negative predictive value (*NPV*), F-measure (*F₁*), Matthews correlation coefficient (*MCC*), and area under the curve (*AUC*).[2,8,16] They are expressed by the following equations:[6,9-12,40]

$$Se = \frac{TP}{TP+FN}, \; Sp = \frac{TN}{TN+FP}, \; PPV = \frac{TP}{TP+FP}, \; NPV = \frac{TN}{TN+FN}, \; Acc = \frac{TP+TN}{TP+TN+FP+FN},$$

$$F_1 = 2\frac{PPV \cdot Se}{PPV+Se}, \; MCC = \frac{TN \cdot TP - FN \cdot FP}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}},$$

where *TP*, *TN*, *FP*, and *FN* are the numbers of true positives, true negatives, false positives, and false negatives, respectively. Area under the curve (AUC) is the area under the receiver operating characteristic curve (Figure 2).

To evaluate the performance of a model, the dataset is usually divided into a training, validation, and test set. Training data are used to develop a model, whereas the learning process and model quality are assessed by monitoring overfitting or underfitting on the validation set. The model is finally evaluated on an independent test set, assuming that the input features are similar to those learned in the training set.[6] *K*-fold cross-validation is an alternative approach to model testing.[10]
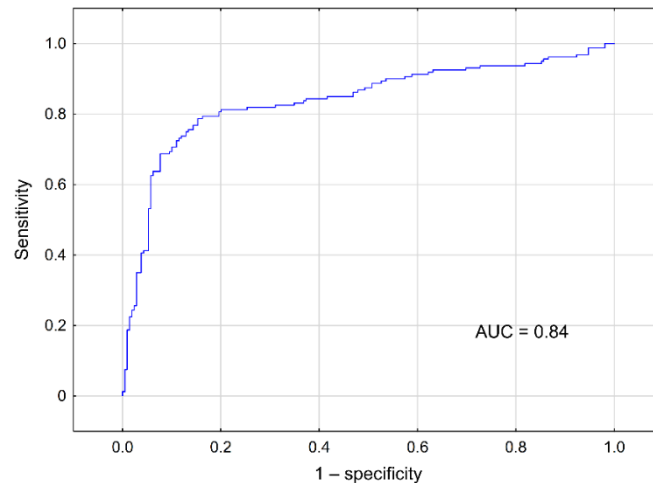
Figure 2. An example of a receiver operating characteristic (ROC) curve (studies from around the world, 2019-2023); AUC: area under the curve.

Finally, some limitations of the present study must be mentioned. The first limitation of this review is the total number of references (40) that were finally included in the text. The second limitation, which is also a drawback, is the use of only one database (PubMed) for article search. However, the inclusion of additional literature sources would have increased the number of references even further. Therefore, a final representative subset of original studies and review articles was selected from the largest biomedical bibliographic database in the world.

CONCLUSION

Most studies on the use of artificial intelligence for COVID-19 diagnosis involved deep learning and feature extraction methods. Segmentation and augmentation were also frequently applied to improve model performance and overcome data scarcity. More extensive data sets and standardized modeling procedures, including an objective evaluation of model predictive capabilities, will be required in the future to introduce these methods into the common clinical practice. Models developed during the peak of the COVID-19 pandemic can be reused in future outbreaks of other similar diseases.

REFERENCES

1. Abdel-Basset M, Chang V, Hawash H, et al. FSS-2019-nCov: A deep learning architecture for semi-supervised few-shot segmentation of COVID-19 infection. Knowl-Based Syst. 2021;212:106647. https://doi.org/10.1016/j.knosys.2020.106647.

2. Mondal MRH, Bharati S, Podder P. Diagnosis of COVID-19 using machine learning and deep learning: A review. Curr Med Imaging. 2021;17(12):1403–18. https://doi.org/10.2174/1573405617666210713113439.

3. Sedik A, Iliyasu AM, Abd El-Rahiem B, et al. Deploying machine and deep learning models for efficient data-augmented detection of COVID-19 infections. Viruses. 2020;12(7):769. https://doi.org/10.3390/v12070769.

4. Aboul-Fotouh S, Mahmoud AN, Elnahas EM, et al. What are the current anti-COVID-19 drugs? From traditional to smart molecular mechanisms. Virol J. 2023;20(1):241. https://doi.org/10.1186/s12985-023-02210-z.

5. Aslan MF, Sabanci K, Durdu A, et al. COVID-19 diagnosis using state-of-the-art CNN architecture features and Bayesian Optimization. Comput Biol Med. 2022;142:105244. https://doi.org/10.1016/j.compbiomed.2022.105244.

6. Aggarwal P, Mishra NK, Fatimah B, et al. COVID-19 image classification using deep learning: Advances, challenges and opportunities. Comput Biol Med. 2022;144:105350. https://doi.org/10.1016/j.compbiomed.2022.105350.

7. Jia G, Lam H-K, Xu Y. Classification of COVID-19 chest X-Ray and CT images using a type of dynamic CNN modification method. Comput Biol Med. 2021;134:104425. https://doi.org/10.1016/j.compbiomed.2021.104425.

8. Fallahpoor M, Chakraborty S, Heshejin MT, et al. Generalizability assessment of COVID-19 3D CT data for deep learning-based disease detection. Comput Biol Med. 2022;145:105464. https://doi.org/10.1016/j.compbiomed.2022.105464.

9. Wu Y, Qi Q, Qi S, et al. Classification of COVID-19 from community-acquired pneumonia: Boosting the performance with capsule network and maximum intensity projection image of CT scans. Comput Biol Med. 2023;154:106567. https://doi.org/10.1016/j.compbiomed.2023.106567.

10. Awassa L, Jdey I, Dhahri H, et al. Study of different deep learning methods for coronavirus (COVID-19) pandemic: taxonomy, survey and insights. Sensors (Basel). 2022;22(5):1890. https://doi.org/10.3390/s22051890.

11. Fang L, Wang X. COVID-19 deep classification network based on convolution and deconvolution local enhancement. Comput Biol Med. 2021;135:104588. https://doi.org/10.1016/j.compbiomed.2021.104588.

12. Qi Q, Qi S, Wu Y, et al. Fully automatic pipeline of convolutional neural networks and capsule networks to distinguish COVID-19 from community-acquired pneumonia via CT images. Comput Biol Med. 2022;141:105182. https://doi.org/10.1016/j.compbiomed.2021.105182.

13. Shi F, Xia L, Shan F, et al. Large-scale screening to distinguish between COVID-19 and community-acquired pneumonia using infection size-aware classification. Phys Med Biol. 2021;66(6):065031. https://doi.org/10.1088/1361-6560/abe838.

14. Xu X, Jiang X, Ma C, et al. A deep learning system to screen novel coronavirus disease 2019 pneumonia. Engineering. 2020;6(10):1122–9. https://doi.org/10.1016/j.eng.2020.04.010.

15. Kuo K-M, Talley PC, Chang C-S. The accuracy of machine learning approaches using non-image data for the prediction of COVID-19: A meta-analysis. Int J Med Inf. 2022;164:104791. https://doi.org/10.1016/j.ijmedinf.2022.104791.

16. Baghdadi NA, Malki A, Abdelaliem SF, et al. An automated diagnosis and classification of COVID-19 from chest CT images using a transfer learning-based convolutional neural network. Comput Biol Med. 2022;144:105383. https://doi.org/10.1016/j.compbiomed.2022.105383.

17. Dey A, Chattopadhyay S, Singh PK, et al. MRFGRO: a hybrid meta-heuristic feature selection method for screening COVID-19 using deep features. Sci Rep. 2021;11(1):24065. https://doi.org/10.1038/s41598-021-02731-z.

18. Zhang X, Jiang R, Huang P, et al. Dynamic feature learning for COVID-19 segmentation and classification. Comput Biol Med. 2022;150:106136. https://doi.org/10.1016/j.compbiomed.2022.106136.

19. Carmo D, Campiotti I, Rodrigues L, et al. Rapidly deploying a COVID-19 decision support system in one of the largest Brazilian hospitals. Health Informatics J. 2021;27(3):14604582211033017. https://doi.org/10.1177/14604582211033017.

20. Shiri I, Sorouri M, Geramifar P, et al. Machine learning-based prognostic modeling using clinical data and quantitative radiomic features from chest CT images in COVID-19 patients. Comput Biol Med. 2021;132:104304. https://doi.org/10.1016/j.compbiomed.2021.104304.

21. Champendal M, Müller H, Prior JO, et al. A scoping review of interpretability and explainability concerning artificial intelligence methods in medical imaging. Eur J Radiol. 2023;169:111159. https://doi.org/10.1016/j.ejrad.2023.111159.

22. Javaheri T, Homayounfar M, Amoozgar Z, et al. CovidCTNet: an open-source deep learning approach to diagnose covid-19 using small cohort of CT images. NPJ Digit Med. 2021;4(1):29. https://doi.org/10.1038/s41746-021-00399-3.

23. Li L, Qin L, Xu Z, et al. Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: Evaluation of the diagnostic accuracy. Radiology. 2020;296(2):E65–71. https://doi.org/10.1148/radiol.2020200905.

24. Fan D-P, Zhou T, Ji G-P, et al. Inf-Net: Automatic COVID-19 lung infection segmentation from CT images. IEEE Trans Med Imaging. 2020;39(8):2626–37. https://doi.org/10.1109/TMI.2020.2996645.

25. Zhang K, Liu X, Shen J, et al. Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. Cell. 2020;181(6):1423–33. https://doi.org/10.1016/j.cell.2020.04.045.

26. Wang X, Deng X, Fu Q, et al. A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT. IEEE Trans Med Imaging. 2020;39(8):2615–25. https://doi.org/10.1109/TMI.2020.2995965.

27. Bao G, Chen H, Liu T, et al. COVID-MTL: Multitask learning with Shift3D and random-weighted loss for COVID-19 diagnosis and severity assessment. Pattern Recognit. 2022;124:108499. https://doi.org/10.1016/j.patcog.2021.108499.

28. Shiri I, Mostafaei S, Haddadi Avval A, et al. High-dimensional multinomial multiclass severity scoring of COVID-19 pneumonia using CT radiomics features and machine learning algorithms. Sci Rep. 2022;12(1):14817. https://doi.org/10.1038/s41598-022-18994-z.

29. Guhan B, Almutairi L, Sowmiya S, et al. Automated system for classification of COVID-19 infection from lung CT images based on machine learning and deep learning techniques. Sci Rep. 2022;12(1):17417. https://doi.org/10.1038/s41598-022-20804-5.

30. Li T, Wei W, Cheng L, et al. Computer-aided diagnosis of COVID-19 CT scans based on spatiotemporal information fusion. J Healthc Eng. 2021;2021:6649591. https://doi.org/10.1155/2021/6649591.

31. Tello-Mijares S, Woo L. Computed tomography image processing analysis in COVID-19 patient follow-up assessment. J Healthc Eng. 2021;2021:8869372. https://doi.org/10.1155/2021/8869372.

32. Topff L, Sánchez-García J, López-González R, et al. A deep learning-based application for COVID-19 diagnosis on CT: The Imaging COVID-19 AI initiative. PLoS One. 2023;18(5):e0285121. https://doi.org/10.1371/journal.pone.0285121.

33. Yang Z, Zhao L, Wu S, et al. Lung lesion localization of COVID-19 from chest CT image: A novel weakly supervised learning method. IEEE J Biomed Health Inform. 2021;25(6):1864–72. https://doi.org/10.1109/JBHI.2021.3067465.

34. Ortiz A, Trivedi A, Desbiens J, et al. Effective deep learning approaches for predicting COVID-19 outcomes from chest computed tomography volumes. Sci Rep. 2022;12(1):1716. https://doi.org/10.1038/s41598-022-05532-0.

35. Nguyen D, Kay F, Tan J, et al. Deep learning–based COVID-19 pneumonia classification using chest CT images: model generalizability. Front Artif Intell. 2021;4:694875. https://doi.org/10.3389/frai.2021.694875.

36. Goel T, Murugan R, Mirjalili S, et al. Automatic screening of COVID-19 using an optimized generative adversarial network. Cogn Comput. 2021. https://doi.org/10.1007/s12559-020-09785-7.

37. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. 2012;25. https://doi.org/10.1145/3065386

38. Kang D, Park S, Paik J. SdBAN: Salient object detection using bilateral attention network with dice coefficient loss. IEEE Access. 2020;8:104357–70. https://doi.org/10.1109/ACCESS.2020.2999627.

39. Rezatofighi H, Tsoi N, Gwak J, et al. Generalized intersection over union: A metric and a loss for bounding box regression. Proc. IEEECVF Conf. Comput. Vis. Pattern Recognit. 2019. p. 658–66. https://doi.org/10.48550/arXiv.1902.09630.

40. Zaborski D, Proskura WS, Grzesiak W, et al. The comparison between random forest and boosted trees for dystocia detection in dairy cows. Comput Electron Agric. 2019;163:104856. https://doi.org/10.1016/j.compag.2019.104856.

Author's contributions:

Daniel Zaborski and Jacek Sieredziński contributed to the literature search, writing of the abstract, introduction, methodology, discussion, interpretation and description of results, preparation of tables, conclusions, review, and statistics.

All authors have approved the final version to be published, being responsible for all aspects of the work and ensuring its accuracy and integrity.