6

Limitations and challenges on the diagnosis of COVID-19 using radiology images and deep learning

Merve Ayyuce Kızrak[1], Zümrüt Müftüoğlu[1], Tülay Yıldırım[2]

[1]THE PRESIDENCY OF REPUBLIC OF TURKEY, THE DIGITAL TRANSFORMATION OFFICE, ANKARA, TURKEY; [2]DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING, YILDIZ TECHNICAL UNIVERSITY, ESENLER-ISTANBUL, TURKEY

1. Introduction

The first coronavirus case, recorded in 1960, was thought to be a cold. The virus, first detected in Wuhan, China, in December 2019, was declared as a new coronavirus by the World Health Organization (WHO) on February 11, 2020. It was renamed COVID-19 (2019-nCov) by WHO. The epidemic, which reached a life-threatening dimension, announced in March 2020 that the WHO case report covers 58 countries worldwide [1]. In the middle of April 2020, the number of patients worldwide has exceeded 2.3 million. The number of those who overcome the disease reaches 600 thousand, and more than 161 thousand deaths have been recorded [2].

Governments have taken a series of measures to isolate the social distance by closing the borders to reduce the extent of the outbreak. However, the number of individuals affected by the virus has increased exponentially in most countries. For the struggle to be medically efficient, significant clinical trials, medicines, and vaccines need to be developed and implemented. For this, sufficient scientific numbers and resources are needed. During this crisis, solidarity was initiated to produce medical equipment with three-dimensional printers and to produce masks for textile companies to support healthcare personnel [2,3]. Based on the data published in the process, while waiting for the Polymerase Chain Reaction (PCR) test time required for diagnosis, pioneering estimates are made, data science and artificial intelligence (AI) studies are carried out and competitions are organized to assist healthcare professionals. Table 6.1 and Fig. 6.1 show the outbreak trend based on the data published by the WHO.

Data Science for COVID-19. https://doi.org/10.1016/B978-0-12-824536-1.00007-1

Table 6.1 COVID-19 Outbreak statistics worldwide.[1]

Country	Health score	Total active	Total test	Total case	Total recovered	Total deaths
USA	6	763.616	5.015.602	919.414	103.609	52.189
UK	6	123.234	612.031	143.464	724	19.506
Italy	7	106.527	1.642.356	192.994	60.498	25.969
Spain	7	104.885	930.230	219.764	92.355	22.524
France	5	94.195	595.154	159.952	43.512	22.245
Turkey	6	80.575	830.257	104.912	21.737	2.600
Russia	4	62.439	2.552.000	68.622	5.568	615
Germany	7	39.487	2.072.669	155.054	109.800	5.767
Netherland	7	31.924	187.667	36.535	322	4.289
Belgium	6	27.492	189.067	44.293	10.122	6.679

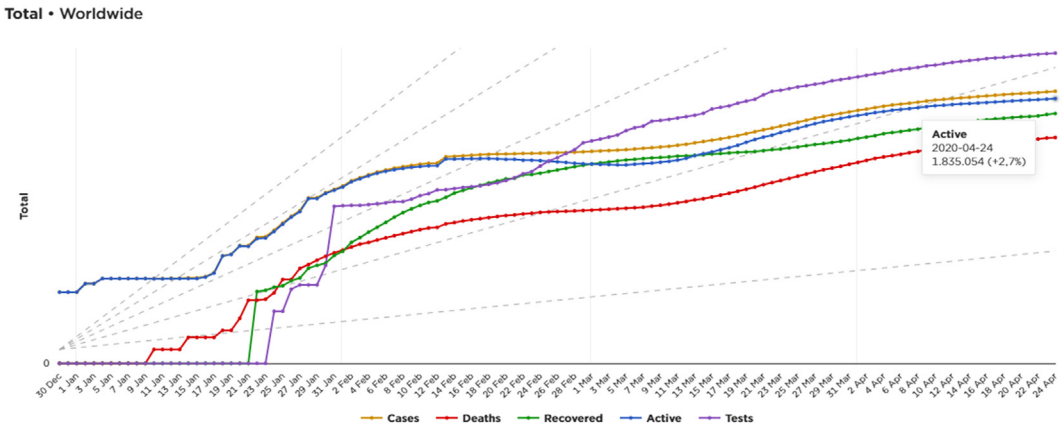


FIGURE 6.1 Graph for COVID-19 outbreak statistics worldwide[1].

The health score in this table is calculated as follows:

- Case of growth rate (50%)—Rate of new cases. Combination of daily % of cases asses as well as double derivative of cases.
- Death growth rate (30%)—Rate of new deaths. Combination of daily % of deaths as well as double derivative of cases.
- Recovery (20%)—The progress toward recovery.

This chapter points to AI for a successful and rapid diagnostic recommendation using chest radiography to reduce waiting time for tests because of the intensity in hospitals as part of efforts to prevent this deadly epidemic disease that appeared in late 2019 and

[1]Governments and research institutes publish similar statistics publicly. This table has utilized from the Presidency of the Republic of Turkey https://corona.cbddo.gov.tr/ prepared by the Digital Transformation Office address. In addition, https://gisanddata.maps.arcgis.com/ website prepared by John Hopkins University WHO and System Science and Engineering Center was used. Updated: 25 April 2020.

spread around the world in the first half of 2020. However, some limitations and privacy approaches should be taken into consideration while producing AI solutions by making use of health data. This article also covers an application to draw attention to data privacy while reaching fast solutions. As an example, a dataset consisting of 373 CXR images collected from open sources, of which 139 were COVID-19 infected, was used for the diagnosis with deep learning approaches of COVID-19 to show the limitations. For training, using EfficientNet, an up-to-date and robust deep learning model, offers the possibility of infected with an accuracy of 94.7%.

Using these results, this chapter details the limitations of deep learning models used for COVID-19 diagnostics from radiology images with the aim of drawing attention to important issues, such as data quality, data amount, explainability, and data privacy while achieving fast solutions. The last section of the chapter aims to provide a perspective about fast, robust, unbiased, and human-centered, AI-powered health applications.

2. COVID-19 radiology imaging dataset

Computing tomography (CT) is an imaging procedure with a cross-section as a result of computer processing of the signals obtained by rotating rapidly around the patient's body with an X-ray. It is a method used by physicians to diagnose many diseases because it contains more detailed information than traditional X-rays. Chest X-ray (CXR) is frequently used with CT to analyze bone tumors, lesions in the abdomen, heart disease anomalies, injury to the brain, stroke, tumor, and bleeding [4]. Chest radiography, such as this, but with fewer details, is used in the CXR. It is also known that CXR is more useful in disease progression monitoring because of its applicability per patient and low X-ray exposure [5].

One of the important symptoms of the COVID-19 epidemic disease, which appeared by the WHO at the end of 2019, is intense cough and breathing difficulties. CXR and CT imaging are frequently used for such symptoms. The pattern of MERS and SARS and COVID-19 pneumonia are similar. However, the finding of COVID-19 becomes clear when discrete nodules and a reverse halo are present [6]. In studies conducted, the relationship between chest images and PCR tests is also examined and it is generally found to be correlated. Besides, CT images have typical findings, and cases with a negative initial PCR result but positive 2−8 days later have been described [7]. Although the clinical picture of COVID-19 is still not very clear, CXR and CT can be used in addition to other test procedures known and developed in the diagnosis of the disease. The rapid evaluation of these images by experts is important in the days when the epidemic is struggling [8]. While diagnostic suggestion systems from AI-powered applications are already in use, CXR and CT images from COVID-19 infected patients are also a data type that paves the way for data scientists to work with healthcare professionals. During this struggle, scientists agree that it is important to publish informative and reliable sources both at an academic level.

In this research, a dataset consisting of CXR images is used by making use of deep learning approaches. A dataset containing 234 healthy and 139 infected totals of 373 CXR images were collected by the T-Covid group from publicly published data [9].

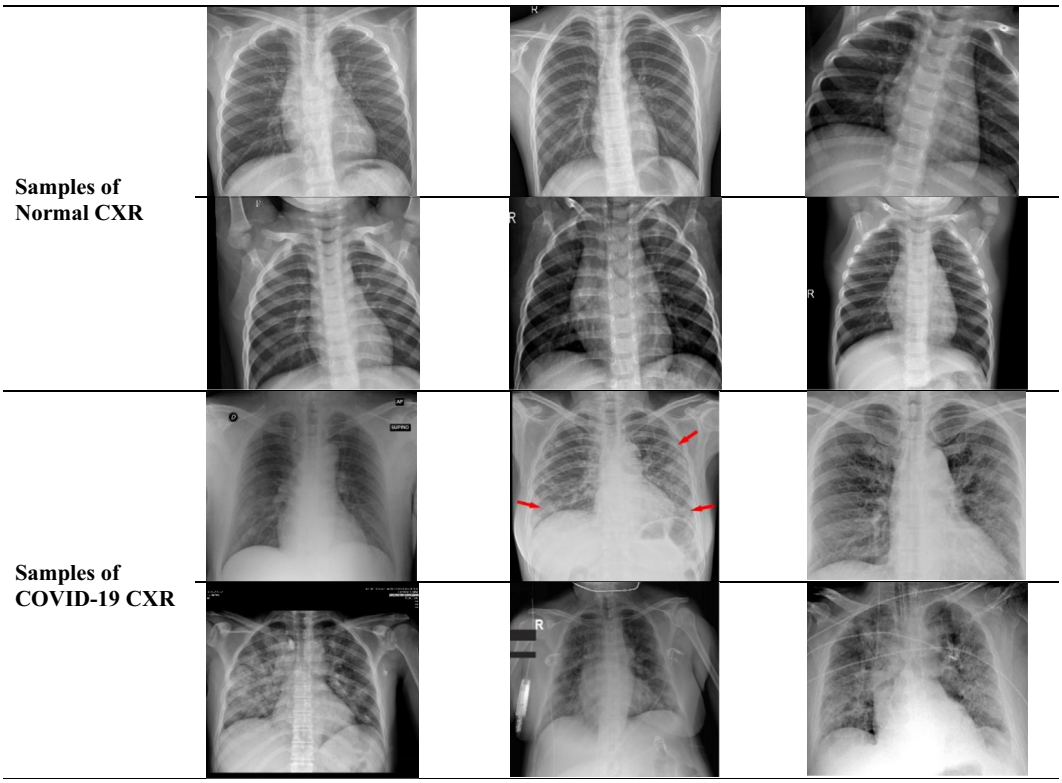


FIGURE 6.2 Samples of the dataset [9].

The images in this new dataset have different sizes. The dataset is divided into 80% for the train, 10% for validation and test. Fig. 6.2 shows details about the dataset.

3. Recent works using radiology images for COVID-19

The epidemic that is still happening pushes scientists to publish new research and light on the slowing down of the pandemic. Imaging-based approaches since the outbreak were announced in this section of the chapter. However, under the frame of the data published for COVID-19, a differential privacy application is not included in the literature yet.

The study examines the clinical features of pneumonia patients infected with coronavirus and influenza virus and emphasizes that it is possible to record the stage of the disease based on radiology images of chests and other test results. CT and CXR shots are used as an auxiliary diagnostic parameter recommended by lung surgery doctors [10,11]. Wang et al. deep learning algorithm use 1119 CT images in their studies. However, these images only include images of patients diagnosed with viral pneumonia, not

just COVID-19. The total accuracy rate is calculated as 79.3%, specificity for test dataset 83%, sensitivity 67%. In the study, where modified-Inception was used as the deep learning model, the accuracy of giving correct diagnosis to COVID-19 positive patients was 85.2% [12]. Zhao et al. use a convolutional neural network (CNN)—based deep learning model, utilizing 275 CT images. In this binary classification study, while the general accuracy rate is 84.7%, it achieves to reach 85.3% in F1-score [13]. In addition to academic resources, many platforms published as open-source and online tools are instance datasets and AI-powered studies of researchers [14,15].

4. Deep learning basics

This section will describe the deep learning approach to diagnosing COVID-19 using radiology images. So, it is useful to make brief deep learning basic. An entire model size needs hardware that requires high-processor capacity, depending on the mathematical processing complexity and the size of the input data. To meet this need, the central processing unit, graphic processing unit, or sometimes more powerful tensor processing units[2] are used. It is also possible to access this hardware using cloud services. However, depending on the conditions under which the memory and AI model will work, advanced batteries are also needed if, for example, it is used on an autonomous robot. A simplified fully connected deep neural network structure is shown in Fig. 6.3.

Although deep learning is popular since 2006 and has taken this name, it is a research area that has been studied under different definitions throughout the history of AI [16].

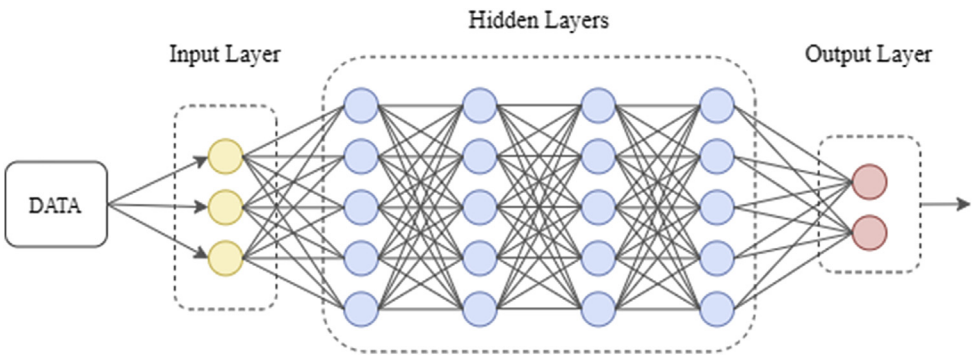


FIGURE 6.3 Simplified deep neural network model.

[2]The tensor is called arrays placed in a grid structure in a regular x, y, z plane. The name of the hardware developed by Google that allows processing on these three-dimensional arrays is Tensor Processing Unit (TPU).

Especially for the training of Artificial Neural Networks (ANNs), the increase of the data and model dimensions and the ability of the equipment to meet this processing power allow projects to be projected in real life by leaving the university laboratories. It is understood in the light of numerous researches that it gives more successful results than classical approaches in solving complex problems.

Deep learning does not have a basic purpose to simulate a brain. However, neuroscience or computational neuroscience sources may also be the focus of interest for some deep learning researchers. Contemporary deep learning is fed from many fields such as information theory, probability-statistics, and linear algebra [17].

In summary, deep learning, which is one of the leading application methods today, is an interdisciplinary field of study on the creation of systems that require data, algorithm, model, and hardware knowledge that aims to solve complex problems requiring intelligence that people can solve. Deep learning has several subtopics. These can be classified according to the way they extract features from the data they are applied to. For example, frequently used structures to extract patterns over images are CNN. Long-term short-term memory models, which are generalized as repetitive and recursive neural networks, are used for data where time series and memory information is required [18]. Despite the fact that the generative adversarial networks (GANs) developed in 2014 have become very talked-about, especially when it comes to style transfer,[3] by producing works such as pictures, music, and poetry, it has been successful in the field of recognition, classification, when it is necessary to produce synthetic data[4] or in low-resolution medical images [19]. Since the deep learning approach used to diagnose COVID-19 in radiology images is CNN models, this section will summarize the CNN structure.

4.1 Convolutional neural networks

Although there is no brain simulation goal for deep learning, there are some approaches inspired by neuroscience or based on neuroscience. Although some of these have failed, CNNs are the deep neural network method that regained confidence in ANN methods and found many applications. The results published by the neurophysiologist Hubel and Wiesel with their experiments on the visual cortex of mammals as of the end of the 1950s were crowned with the Nobel Prize [20]. Fig. 6.3 shows the visual cortex hierarchy features of mammals. The most important findings of experiments[5] on cats belong to the V1 region of the brain, which is defined as the primary visual cortex. Spatial mapping of the image perceived from the retina of the V1 region, detection of small and complex layers, and light and spatial transitions are important sources of information in the

[3]Style transfer is to produce a new result by extracting patterns in two different data and transferring the pattern structure in one. For example, transferring the styles of the artists to the current photographs taken.

[4]Synthetic data is the production of new and unreal data using the patterns of existing data. The production of human faces that do not actually exist in recent years can be given as an example.

[5]It is possible to find the records of the experiment that Hubel and Wiesel cats measured from the visual cortex on YouTube. See https://www.youtube.com/watch?v=8vdff3egwfg.

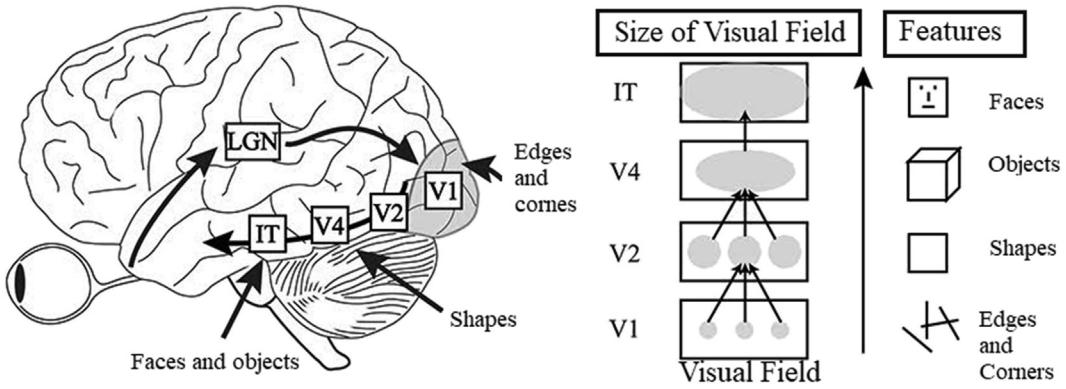


FIGURE 6.4 Visual cortex hierarchy features of mammals [20].

CNN structure. However, the CNN structure also has very different aspects of the mammalian visual perception system. For example, most image detection sensors use high-resolution images as input information. However, in mammals, the image is quite low resolution outside the fovea region. A small focused area of the image defined as the input has high-resolution detection (Fig. 6.4).

Krizhevsky and Sutskever have achieved great success, almost doubling the success of the previous year under the advisor of Hinton in the ImageNet object recognition contest in 2012 with the grid-topology and the ability to capture patterns in high-dimensional data [21]. In an interview, Ng[6] mentions that he suggested using parallel graphic processing units in this study. Thus, it was understood that the CNNs and the image data combined with the developing hardware technology gave more successful and faster results. CNNs are, in their simplest definition, a neural network structure with layers that contain multiple processes based on the convolution process rather than matrix multiplications along with their layers. Convolution, which is the basis of this structure, is a mathematical operation that contains temporal (t) and spatial (x) information. When measuring an object displaced over time, noises are also recorded in the measuring channel, and the importance of the noises in the function increases as time progresses, which is undesirable. To reduce this noise effect, a function $w(a)$ that averages the measurements is included in the convolution process. It is defined as the probability density function, known as the weight function. In machine learning approaches, it is possible to define different functions instead of w. When this process is applied to spatial information at all times, Eq. (6.1) is obtained. The representation of the convolution process is indicated by the $*$ sign. Here, the x entries provide the kernel function defined as w, and the s attribute map calculated as output.

[6]Andrew Ng is the founder of many artificial intelligence companies in the 2000s and a scientist with the title of professor of Stanford University in Electrical and Computer Science. See interview: https://www.youtube.com/watch?v=-eyhCTvrEtE.

$$s(t) = \int x(a)w(t-a)da \equiv s\,(t) = \int x * w(t)da \tag{6.1}$$

Periodic and discrete measurements of time are the most used by researchers in problem assessment. Therefore, the convolution process Eq. (6.2) is shown discretely in time. It is suitable for definition within different dimensions. For example, a standard image is defined in two dimensions. This necessitates defining the convolution process for two axes. Since convolution is an unordered function, the order of the kernel function and the input matrix in the process does not affect the result. If the input information is defined as I and core function O, output S is expressed as Eq. (6.3) in i and j axes for discrete samples at a time, and m and n show samples as a result of the convolution process.

$$s(t) = \int x * w(t)da = \sum_{a=-\infty}^{\infty} x(a)w(t-a) \tag{6.2}$$

$$S(i,j) = (I * O)(i,j) = \sum_{m}\sum_{n} I(m,n)O(i-m, j-n) \tag{6.3}$$

Based on this equation structure, instead of transforming the core function on the axis, machine learning models adopt another very similar mathematical representation, cross-correlation expression. This notation is shown in Eq. (6.4).

$$S(i,j) = (O * I)(i,j) = \sum_{m}\sum_{n} I(i+m, j+n)O(m,n) \tag{6.4}$$

Cross-correlation and reverse correlation are similar to neuroscience-related Gabor functions in the structure of weights defined as the V1 cell in the visual cortex of mammals. When this mathematical approach and the attribute maps obtained with the CNN are visualized, it extracts high-frequency spatial information such as edge, corner, brightness, and color transitions, which are defined as simple attributes in the visual input.

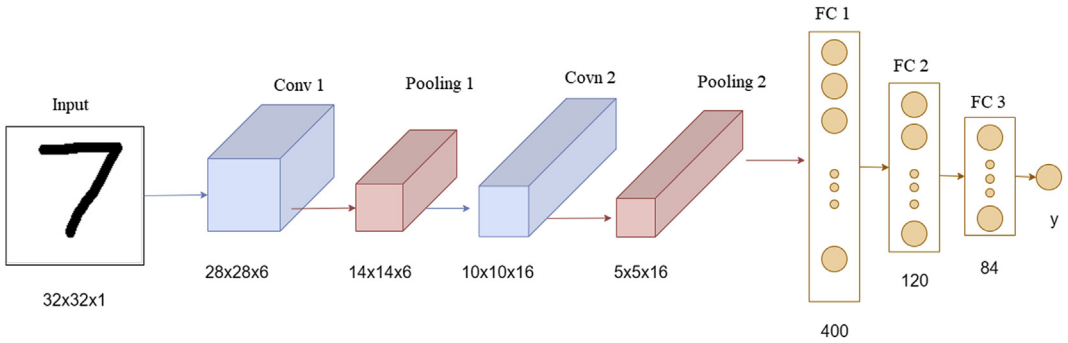


FIGURE 6.5 LeNet convolutional neural network model [22].

As components of a CNN, the convolution process and a nonlinear activation function applied to the output of this process are used. Here, the pooling operation is applied for size reduction, which causes a loss of information.

Fig. 6.5 shows a simple representation of the CNN LeNet model. Both height and width sizes are given as 32 [22]. Due to the nature of the convolution process in the first layer, the height-width decreases as the channels are formed. If it is necessary to keep the height-width size constant, it is necessary to padding the input data before applying the convolution process or if the smaller size needs to be obtained, the stride should be selected large. In the pooling layer, while the number of channels remains constant, the size of height and width are reduced in size. This is a mathematical approach that eases the computational complexity of the process and does not cause much loss from the input information. Situations performed by taking the average of the pooled values, choosing the largest, or choosing the median value can be encountered.

There are parameters to be considered for regularization such as number of filters, activation function selection, optimization algorithm selection, number of layers, quality and bias of data, pretraining. When working with limited and small data, it is advantageous in terms of speed, computational complexity, and accuracy to start a new problem with weights that have learned the basic features of previously trained images. The name of this method is transfer learning.

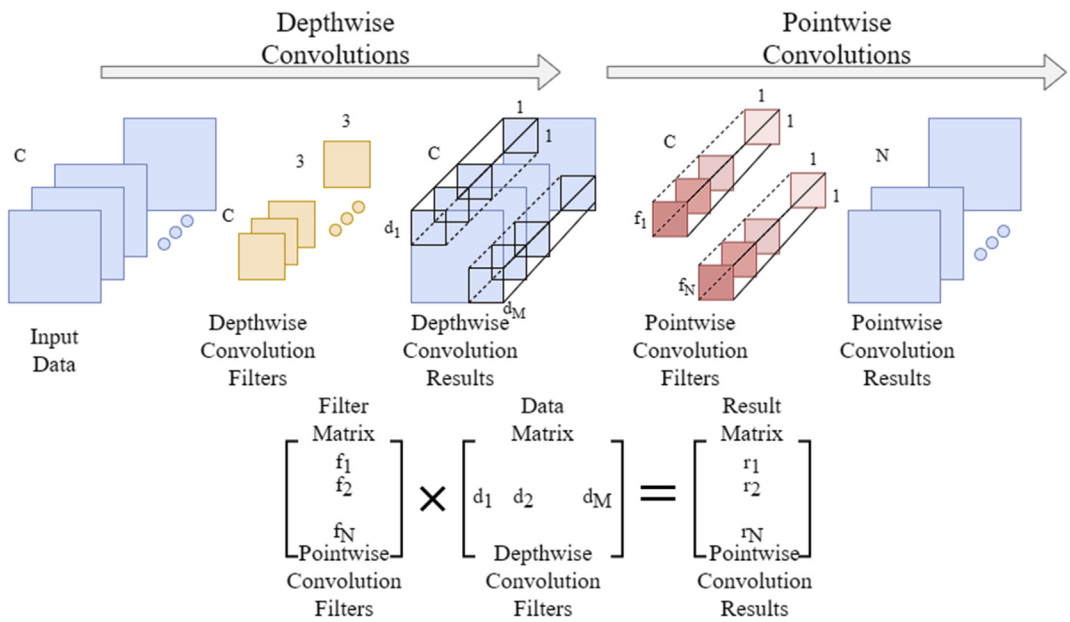


FIGURE 6.6 A basic representation of depthwise and pointwise convolutions.

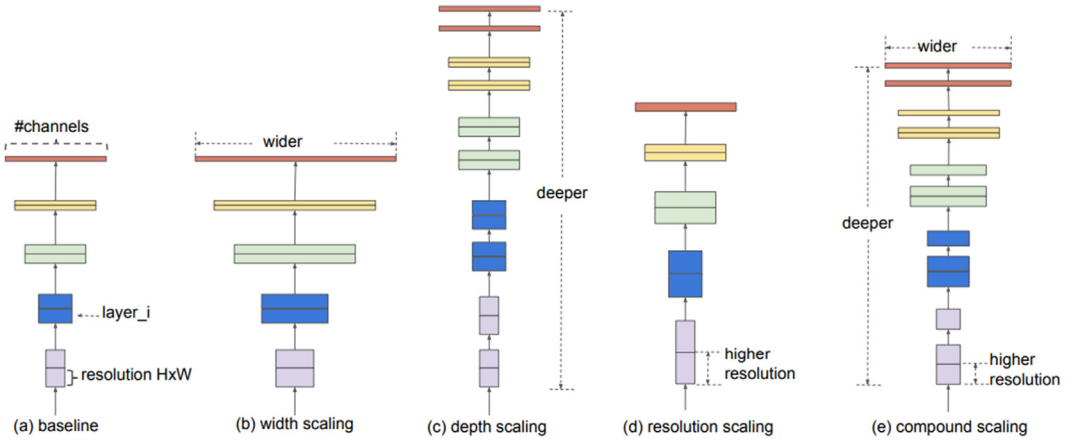


FIGURE 6.7 Model Scaling. (A) is a baseline network; (B) scaling dimension of network width; (C) scaling dimension of network depth; (D) scaling of network resolution; (E) scales three dimensions with a fixed ratio [23].

4.2 EfficientNet as the deep learning model

EfficientNet is a robust deep learning model based on a CNN. EfficientNet B0−B7 is a group of eight models. The model uses depthwise and pointwise convolution, which divides the original convolution into two stages to significantly reduce the cost of computing with optimal accuracy. The input size of the EfficientNet-B0 model is set to 224×224. This approach is among the most successful and fastest models of 2019 by scaling according to depth, width, and input resolution. Figs. 6.6 and 6.7 show the scaling structure that summarizes EfficientNet. Depth, width, and resolution notations are shown in Eqs. (6.5)−(6.7).

$$\text{depth} \rightarrow d \; = \; \alpha \, \phi \rightarrow \alpha \geq 1 \tag{6.5}$$

$$\text{width} \rightarrow w = \beta \, \phi \rightarrow \beta \geq 1 \tag{6.6}$$

$$\text{resolution} \rightarrow r = \gamma \, \phi \; s.t. \; \alpha \cdot \beta \, 2 \cdot \gamma \, 2 \approx 2 \rightarrow \gamma \geq \; 1 \tag{6.7}$$

$\varphi \; = \; 1$ and grid search for $\alpha, \; \beta$, and to scale from B0 to B1. α, β, γ set: Thus, for scaling from B2 to B7, φ is selected between 2−7. The main building block for EfficientNet is the inverted bottleneck MBConv of MobileNetV2, another deep learning model. Using this bottleneck approach, it manages to reduce the computational complexity k^2 rather than traditional methods. In the case where k expresses the kernel size, it indicates the height and width of the two-dimensional convolution window.

The B7 model from the EfficientNet group maintains 84.4% top-1 and 97.1% top-5 classification accuracy in the ImageNet dataset. A major characteristic of this model is its scaling, which is 8.4 times smaller and 6.1 times faster than the model with the closest classification accuracy [23].

4.3 Using EfficientNet as classification method for radiology images

The EfficientNet implementation using radiology images for the diagnosis of COVID-19 is described in this section of the chapter. One of the important differences that distinguish this application from others is that it has been carried out only to work with images from the COVID-19 structure rather than the coronavirus CT or CXR images of previous years. EfficientNet-B0 model, one of the state-of-the-art deep learning models, was applied. The training was conducted by applying data augmentation and parameter optimization. The equation of Adam optimization update is shown in step η for w weights in Eqs. (6.8)−(6.12), where m_t and v_t initial moments vector for the first and second-order, $\widehat{m}_t$ and $\widehat{v}_t$ bias-corrected estimators for the first and second moments. The Adam optimization method is more accurate for machine learning and deep learning models by combining AdaGrad's ability to deal with sparse gradients and RMSProp's ability to deal with nonfixed targets. It also has a low memory requirement. It is a recommended approach for nonconvex optimization problems [24]. Tables 6.2 and 6.3 shows the parameters selected for training.

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t \tag{6.8}$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \tag{6.9}$$

$$\widehat{m}_t = \frac{m_t}{1-\beta_1^t} \tag{6.10}$$

$$\widehat{v}_t = \frac{v_t}{1-\beta_2^t} \tag{6.11}$$

$$w_t = w_{t-1} - \eta \frac{\widehat{m}_t}{\sqrt{\widehat{v}_t + \varepsilon}} \tag{6.12}$$

Evaluation of the results: Classification results can be evaluated in different ways. However, just looking at the accuracy rate can be misleading, especially on critical subjects. For this reason, the achievements in different outcome evaluation criteria should also be calculated when working on medical data. Hence, it is necessary to create a confusion matrix first. This indicates the state of the four possible outcomes.

Table 6.2 Model parameters.

Optimizer	Adam (β_1= 0.9, β_2= 0.999)
Learning rate	0.001
Batch size	32
Epoch	10/120

Table 6.3 Confusion matrix.

		Predicted	
Actual		Positive	Negative
	Positive	True positive	False negative (type 2 error)
	Negative	False positive (type 1 error)	True negative

These;

- True positive: Deciding the positive result correctly,
- True negative: Deciding the negative result correctly,
- False positive (type 1 error): Making a positive decision to an actual negative class,
- False negative (type 2 error): Negative decision for a positive class.

Based on these four metrics, accuracy, recall, precision, and F1-score achievements should be calculated. The correlation of these metrics to the confusion matrix is expressed by the following Eqs. (6.13)−(6.16) [25].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{6.13}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6.14}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6.15}$$

$$F1 - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6.16}$$

Our results: Accordingly, Table 6.4 shows confusion matrix values. Table 6.5 shows the precision, recall, and $F1$-score metrics as well as total validation and test accuracy. Fig. 6.8 shows the accuracy and loss graphs for the EfficientNet-B0 model. These results show the best accuracy among recent similar studies. Examples of the classification result in the images in the test group obtained are shown in Table 6.6.

Table 6.4 Confusion matrix of the EfficientNet-B0 model.

		Predicted	
Actual		COVID-19	Normal
	COVID-19	0.941	0.058
	Normal	0.047	0.952

Table 6.5 Performance of the EfficientNet-B0 model.

Model	Label	Test precision	Test recall	Test F1-Score	Validation accuracy	Test accuracy
EfficientNet-B0	Normal	0.950	0.950	0.950	0.970	0.947
	COVID-19	0.940	0.940	0.940		

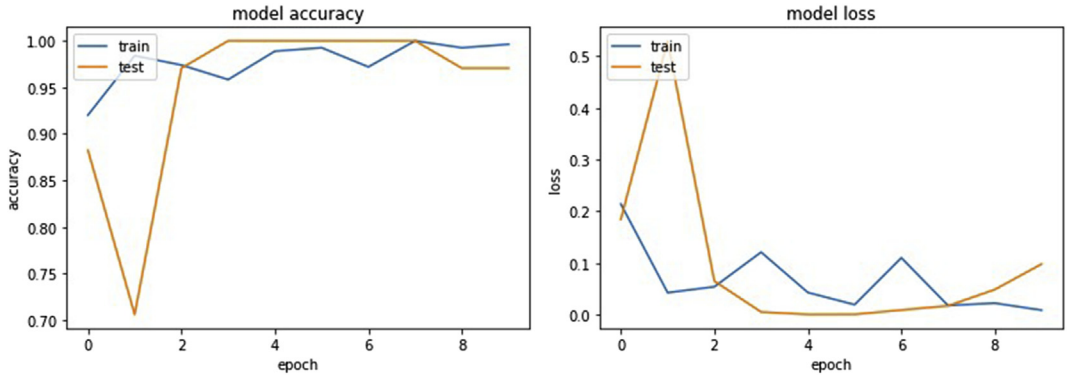


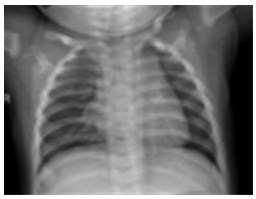
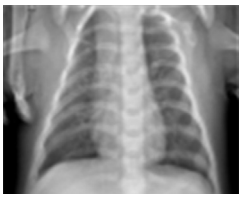
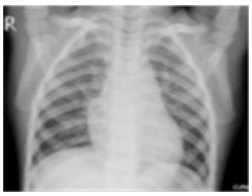
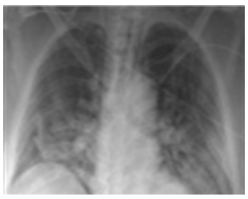
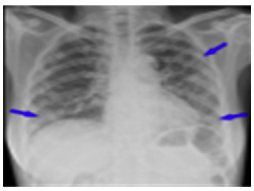
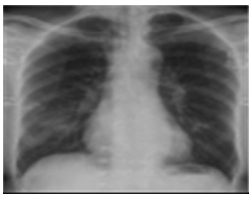
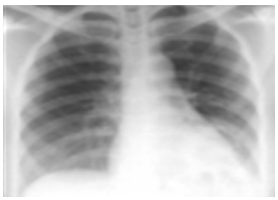
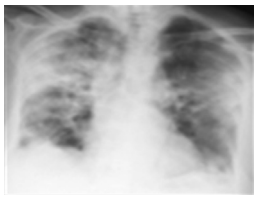
FIGURE 6.8 Result of EfficientNet-B0 model accuracy and loss.

At this stage of the study, better accuracy has been obtained from recent studies in the literature, despite the limited data on detection of an advanced deep learning method and COVID-19 infected images. However, all these researches still have many limitations and challenges. The amount of data, data quality, and reliability of the data are concepts that directly affect the impact of research. Besides, explainability, accessibility, and privacy are also subject matter to be addressed.

5. Limitations and challenges

Despite all its promises, AI brings difficulties in many stages from its development to its use. The main reason is that, in most health-related systems, it starts with the need for real data rather than synthetic data, and as the AI models become more complex, algorithms lose their explainability. Legal approval of the process is based on the understandability of the systems in some cases. Much is written about black-box algorithms; there are cases where deep neural networks, in particular, are not possible to understand the result produced [26]. This blur caused the European Union's General Data Protection Regulation clarification requests for transparency before an algorithm was used for patient care. These discussions on whether it is acceptable to use

Table 6.6 Samples of after classification to Normal and COVID-19 CXR.

Samples of normal result	Samples of COVID-19 result	Samples of misclassification result
		
		
		

nontransparent algorithms for patient care are up to date. Prescribing a drug without a known mechanism of action, it is noteworthy that many aspects of drug administration are not explained, and it is another research side of the subject with social content.

An important issue for AI applications, as in other sectors, is based on how well data privacy and security can be ensured. Given the most common attack and data protection infringement problems, it is predicted that it is not possible to use algorithms that carry the risk of revealing details of the patient's medical history. It may also be possible to identify an individual by facial recognition or genomic sequences from mass databases, thereby making it difficult to protect privacy. On the other hand, GANs have achieved the success that can even deceive people in manipulating content.[7] These types of studies are researches that require a balance between transparency and turbidity, which must be studied carefully in the field of health. To keep the use of AI applications for health

[7]The synthetic data produced by these GANs have reached an indistinguishable point from their realities. Synthetic human faces produced by Mike Tyka in 2017 using GAN. See. http://www.miketyka.com/.

safety, the use of high-security data platforms, and the establishment of state legislation are increasingly important, as in Estonia[8] [26].

5.1 Small data

Academic studies by AI-based medical applications, especially those related to imaging, are above human performance. However, it is not possible to see that it has started to be used in hospitals for now. Because there are some bottlenecks for this, and the most important thing is the low/small data. To deep learning to the efficiency of successful results, the distribution of the data must be of similar density for each class and the data must be large amounts. The apparent contradiction between the population's big data focus and personalized medical practice contributes to relatively little and slow applications of big data in medicine compared to other areas of information. If you do not have 10000 labeled CXR imaging data, your AI application cannot be expected to exceed the performance of a radiologist. Several ways to overcome this bottleneck are suggested:

- Transfer learning
- Few-shot/one-shot learning
- Self-supervised learning
- Data augmentation

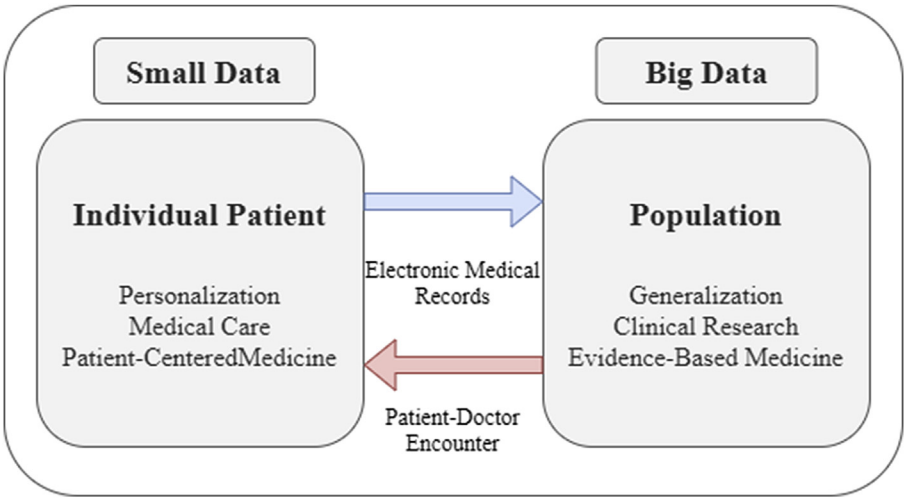


FIGURE 6.9 Learning health care system. Each medical act is the intersection between small and big data [27].

[8]Estonia is among the countries that have taken important steps to use high-security data platforms in the field of health and to establish state legislation. See https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/2019/07/26/Estonian+Central+Health+Information+System+and+Patient+Portal.

These data augmentation methods also increase the performance of models in terms of generalizability and robustness.

On the other hand, while researching AI for medical applications, there are studies on the use of small data. Small data is needed to produce valuable information, and without small data, there is no big data. The fact that AI can turn into a high performance and real medical applications can only be achieved by using small and big data together [27]. Fig. 6.9 shows the relationship between small data and big data in medical applications.

Studies are proposing to use small and distributed data for medical applications instead of centralized big data, with the view that it is possible to develop a real AI-powered health and patient care application, through the continuous and effective interaction between big data and small data. However, at this stage, because of the lack of sufficient and reliable data about COVID-19, it is not possible to use in hospitals the deep learning studies carried out so far.

5.2 Explainability

In the past 10 years, deep learning models have gained popularity with their use in almost every industry and their high accuracy. However, one of the important deficiencies of AI applications is transparency, interpretability, and explainability which have been frequently discussed in recent years. The explainability of the results becomes more important when these models, which have hundreds of layers and millions of artificial neurons, are used in critical sectors such as health. So much so that in some areas such as object recognition, there are issues that deep learning models overcome human accuracy. However, with some simple attacks,[9] models can make wrong decisions. This causes questioning of reliability.

Along with the accuracy of advanced applications, the complexity increases and its explainability becomes difficult. Sates that it faces difficulties in the autonomous and symbiotic systems developed by the US Department of Defense, as in the health sector [28]:

"Explainable AI—especially explainable machine learning—will be essential if future warfighters are to understand, appropriately trust, and effectively manage an emerging generation of artificially intelligent machine partners."

It is aimed to have the ability to explain the reasons for AI systems, identify their ability and disability, and understand how they will behave in the future. The strategy advanced to achieve this goal is to develop novel or modified AI techniques that will produce more explicable models. These models are present to be combined with state-of-the-art human-computer interface techniques that can be translated into understandable and useful explanation dialogs for the end-user. Fig. 6.10 shows the relation between explainability and learning performance.

[9]Understanding the one-pixel attack: Propagation Maps and Locality Analysis, See https://arxiv.org/pdf/1902.02947.pdf.

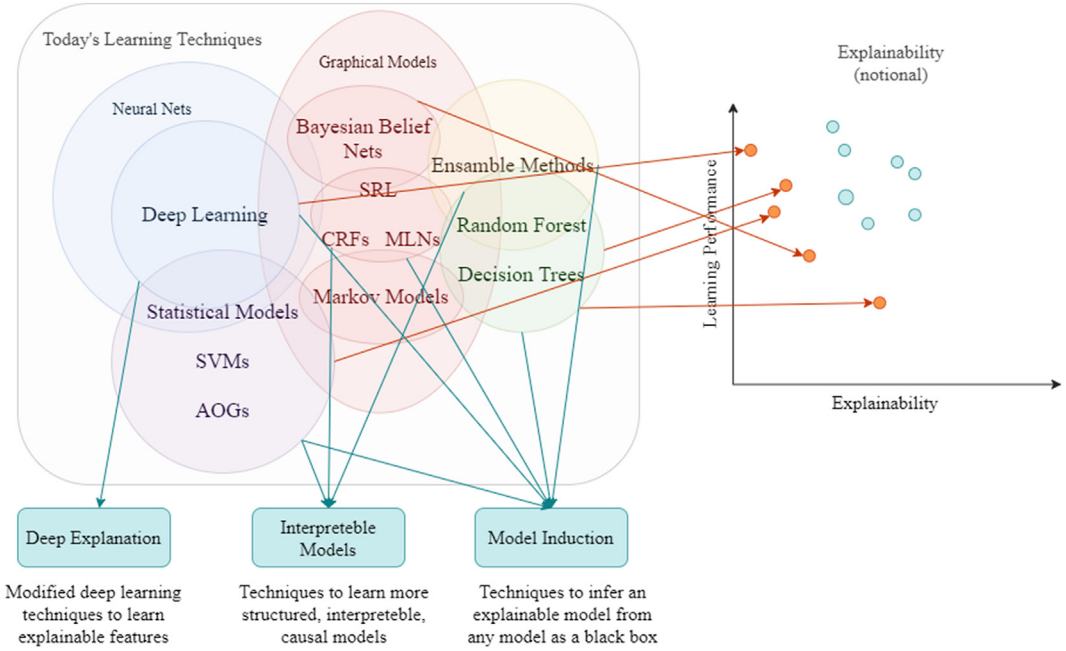


FIGURE 6.10 Artificial intelligence performance versus explainability [28].

It is desired to approach the system with three basic expectations [28]:

- How the sides that design and use the system are affected,
- Data sources used and how they affect the results,
- To explain what kind of a result is obtained by starting from inputs in an AI model.

If it is handled specifically for health practices, when the AI-assisted diagnosis system makes the diagnosis of COVID-19 with high probability for the patient, it should be understood by the physician and the physician should be able to explain to the patient. Direct and relative (indirect) data used to train the system before coming to this phase are also important criteria. It should also be explained what data is needed and why. In analytics, interpretability, transparency, and explainability are kept at the point where security is provided at the end of the analyzes. More clearly, in the best conditions, it means that the best explanation is expected from a system that produces the best performance. This ultimately becomes an optimization problem. It is essential to balance high performance or explainability [29].

There are studies on the explainability of models using visual data and the classification of the models by looking at which regions of the images. The reliability of the deep learning classifier used with this model named Grad-CAM can be tested. It can also be used to identify possible bias in datasets. In the medical, this might guide the physician's information about the patient with natural language processing methods, writing abstracts from the radiology image, or answering visual questions. Gradients take the value

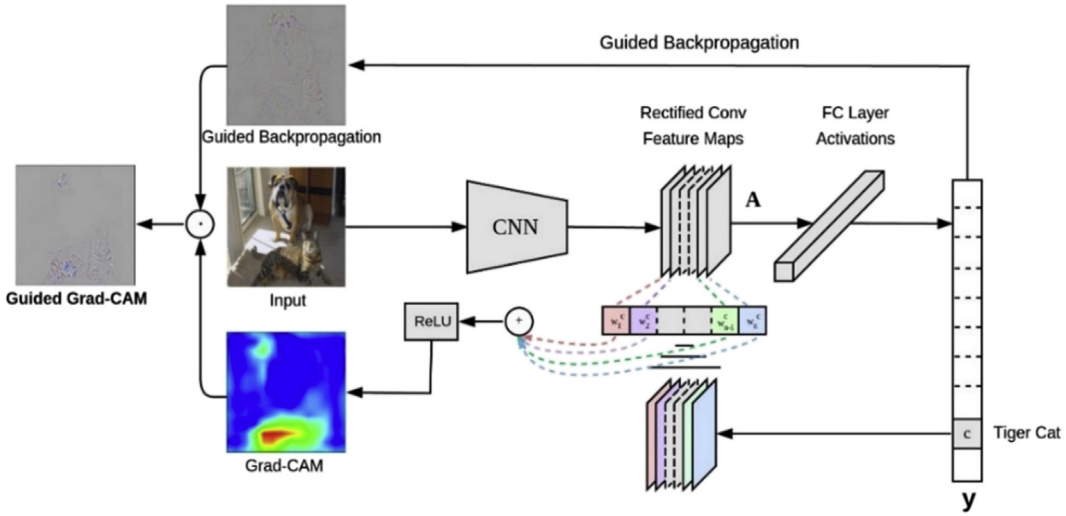


FIGURE 6.11 Grad-CAM model [30].

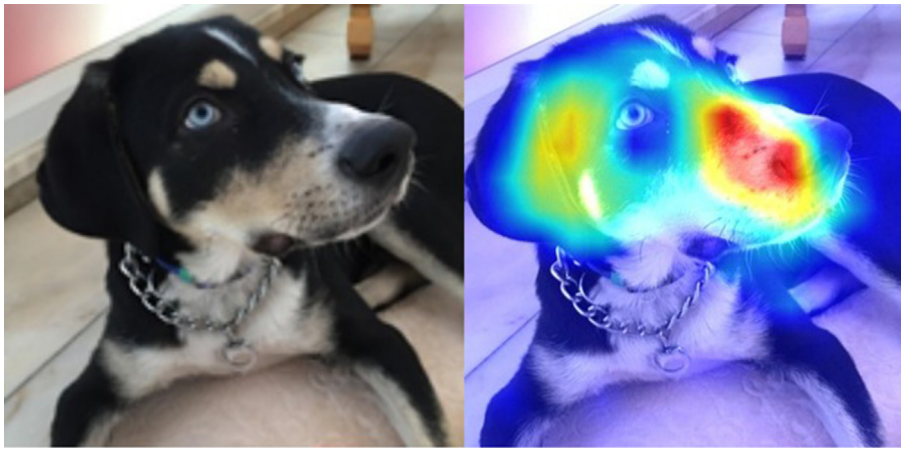


FIGURE 6.12 Original image and after VGGNet classification feature visualization.

1 for the relevant class and other possible classes are shown in blue (gray in printed version) on the heat map. To obtain Guided Grad-CAM visualizations, the dot product process is performed with the backpropagation of the heat map directed [30]. The model that summarizes the study is shown in Fig. 6.11.

In a classification problem, based on the density of the attribute in the layers of the convolutional neural network -based model, it visualizes in which parts of the image the information is used. It is understood that in Fig. 6.12, it is concentrated in the ear and nose region. Fig. 6.13 shows the EfficientNet model used for the diagnosis of COVID-19 from CXR images. However, the amount of data is an important limitation in these studies.

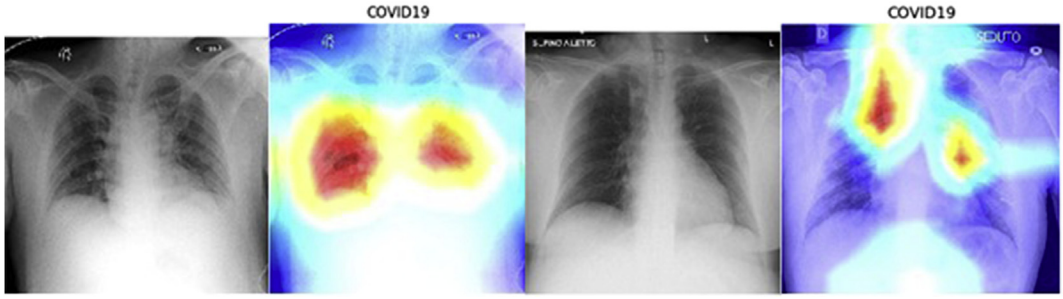


FIGURE 6.13 Original chest X-ray and after EfficientNet classification feature visualization.

Open AI has annexed a different dimension to the explainability and interpretability perspective in April 2020 by inviting neuroscience researchers. The published Microscope Tool offers layer-level and neuron-level visualization of state-of-the-art deep learning models trained with known large visual datasets such as ImageNet. The pattern recognition relationship between the first layers and deep layers helps to understand and investigate complex nervous systems [31]. Microscope visualization is shown in Fig. 6.14 for the first convolution layer and Fig. 6.15 for the fifth convolution layer of the AlexNet model trained on the ImageNet dataset.

For the diagnosis of COVID-19, which is the subject of this section, explainability for artificially assisted applications and other health applications are expected both in terms of the appropriate development of the systems and patient safety. AI applications from an explainability perspective are an important and current field of research for scientists.

5.3 The perspective of differential privacy

Deep learning models are sensitive to several types of attacks as other machine learning models can disclose sensitive information. In literature, there are studies on the model-inversion attack that recovers images from a facial recognition system [32], access to the training mechanism and the model parameters [33], and general adversarial setting in which potential privacy leaks can root in malicious inference with the model's inputs and outputs [34]. Differentially private deep learning methods are shown in Table 6.7.

The methods to guarantee differential privacy can be classified into two types. The first type adds noise to the running process of the optimization algorithm. The second puzzles the objective by adding differentially private noise to the objective functions before the performing learning procedure.

Differential privacy is a probabilistic privacy mechanism that ensures an information-theoretic security warranty. The definition of differential privacy given by Dwork as follows [37]:

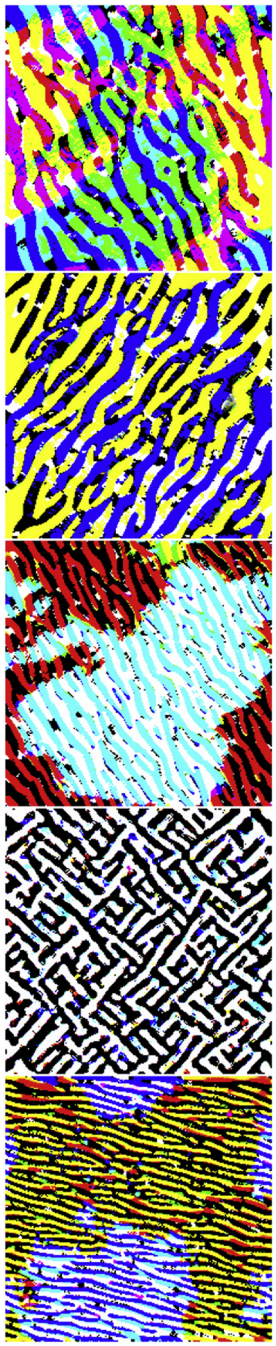


FIGURE 6.14 Sample Microscope results of the first convolutional layer of AlexNet [31].

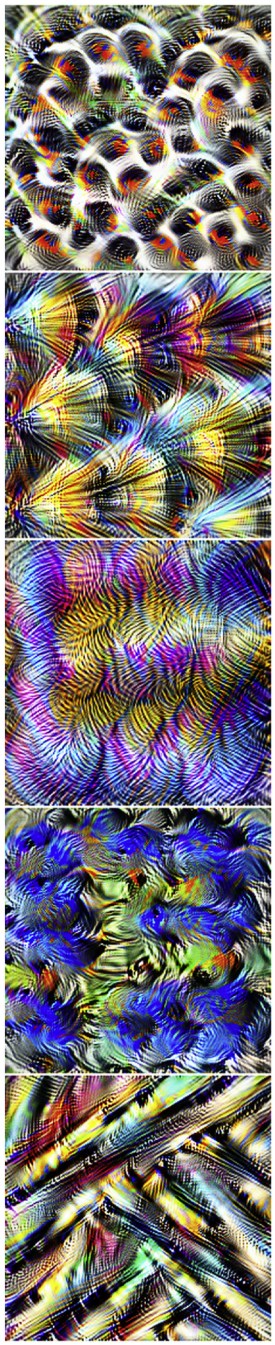


FIGURE 6.15 Sample Microscope results of the fifth convolutional layer of AlexNet [31].

Table 6.7 Differentially private deep learning methods [36].

Related work	Adversarial setting	System setting	Privacy guarantee method
Shokri and Shmatikov [35]	Additional capabilities	Distributed system	Differentially private stochastic gradient descent (SGD) algorithm with convex objective functions
Adabi et al. [33]	Additional capabilities	Centralized system	Differentially private SGD algorithm with nonconvex objective functions
Phah et al. [34]	General capabilities	Centralized system	Objective function perturbation of deep auto-encoder

Differential Privacy $(\varepsilon, \boldsymbol{\delta})$: A randomized mechanism M preserves ($\mathcal{E}$,δ) differential privacy for each set of outputs S, and any neighboring datasets of D and D' differing by at most one record, if M satisfies [38]:

$$Pr[M(D) \in S] \leq \exp(\varepsilon).Pr[M(D') \in S] + \delta \tag{6.17}$$

where ε is the privacy budget and δ is failure probability. The ratio on two probabilities is restrained by e^{ε} for a certain output. The randomized mechanism M grants $\varepsilon-$ differential privacy by its strictest definition if $\delta = 0$. When $\delta = 0$, a strictly stronger notion of $\varepsilon-$ differential privacy is achieved. (ε, δ) -differential privacy maintains latitude to break rigid $\mathcal{E}$-differential privacy for some low probability events [39].

The following definition is called privacy loss:

$$\ln \frac{Pr[M(D) \in S]}{Pr[M(D') \in S]} \tag{6.18}$$

One of the ways to achieve $\varepsilon-$ differential privacy and $(\varepsilon, \delta)-$ differential privacy is adding noise sampled from Laplace and Gaussian distributions, respectively. Here, It should be noted that the noise is proportional to the sensitivity of the mechanism M [40].

The privacy budget (ε): The parameter ε is described as a *privacy budget* that enables to control of the privacy assurance grade of mechanism M. A smaller $\mathcal{E}$ stands for more powerful privacy [41].

Sensitivity (Δ): Sensitivity states how much complexity is required in the mechanism. For example, when we publish a specified query f of dataset D, the sensitivity will calibrate the required volume of noise for $f(D)$. There are two types of sensitivity in the literature of differential privacy: the global sensitivity and the local sensitivity.

6. Summary and future perspective

Digitalization in health has great advantages. Significant breakthroughs can be made in health services when health personnel has access to the information they need. However, the danger of misuse of sensitive information is always a natural problem.[10] European Data Protection Board points out that there may be difficulties in protecting privacy even when using the information ourselves. Generally, major problems arise from interrelated data in terms of privacy. Such correlated data catalyze personal identification. Health data is among the most sensitive data group [42]. The problem is that even if the data is encrypted, it can be recognized on a personal level. It is necessary to have confident and fair information technology routines just before the legislative discussions so that personal privacy violations cannot be made easily. Building a robust and reliable information system should be accepted as the basis of digitalization in health.

This chapter focuses on the use of CXR images, which is an additional procedure in which the healthcare personnel can obtain efficient and rapid results while performing the PCR test in the COVID-19 outbreak. With the recent deep learning model EfficientNet, technical details about the successful diagnosis from other studies in the literature have been given. Important limitations such as small data, explainability, and privacy are emphasized in detail. Furthermore, in AI studies administer in the domain of health and medicine, it is underlined that it requires fastidiousness in terms of data and model privacy and robustness.

As further studies, it will be useful for researchers to try several different privacy practices and compare different mechanisms in deep learning-based medical image classification models. Thus, a fair and secure network can be achieved while providing AI and digitalization in health.

Acknowledgments

We wish to acknowledge Yavuz Kömeçoğlu, a machine learning engineer who worked with us for modeling. We are grateful to Radiologist Dr. Nevit Dilmen who shared information about the use of CT and CXR images for the diagnosis of COVID-19. We would also like to thank T-Covid powered by Turkish AI start-up T-Fashion, who created the dataset and shared it for this study.

References

[1] S.N. Mali, A.P. Pratap, B.R. Thorat, The rise of new coronavirus infection-(COVID-19): a recent update, Eurasian J. Med. & Oncol. (EJMO) 4 (1) (2020) 35–41.

[2] Collective Production Movement in against COVID-19, 2020. Available: https://www.3boyutludestek.org/.

[10]"Big Data and privacy: A technological perspective," President's Council of Advisors on Science and Technology, Reported: May 2014, See https://obamawhitehouse.archives.gov/the-press-office/2015/11/16/fact-sheet-pcast-report-big-data-and-privacy-technological-perspective.

[3] COVID19 Global Online Hackathon, 2020. Available: https://covid-global-hackathon.devpost.com/.

[4] M.Y. Ng, et al., Imaging profile of the COVID 19 infection: radiologic findings and literature review, Radiol. Cardiothoracic Imaging 2 (1) (2020) e200034.

[5] T. Ai, Z. Yang, et al., Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases, 2019.

[6] S. Salehi, A. Abedi, et al., Coronavirus disease 2019 (COVID-19): a systematic review of imaging findings in 919 patients, Amer. J. Roentgenol. Diagn. Imaging & Relat. Sci. (2020) 1–7, https://doi.org/10.2214/AJR.20.23034.

[7] Y.H. Jin, L. Cai, et al., A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-nCoV) infected pneumonia, Mil. Med. Res. 7 (2020) 4.

[8] A. Bernheim, X. Mei, et al., Chest CT Findings in Coronavirus Disease-19 (COVID-19): Relationship to Duration of Infection, 2020, https://doi.org/10.1148/radiol.2020200463. Available:.

[9] T-Covid, A Fast COVID-19 Diagnosis Tool Powered by AI, 2020. Available: https://covid.tfashion.ai/.

[10] Q. Ding, P. Lu, et al., The clinical characteristics of pneumonia patients co-infected with 2019 novel coronavirus and influenza virus in Wuhan China, J. Med. Virol. (2020). https://doi.org/10.1002/jmv.25781. Available:.

[11] X. Li, M. Liu, et al., Preliminary recommendations for lung surgery during the 2019 novel coronavirus disease (COVID-19) epidemic period, 20, Zhongguo Fei Ai Za Zhi 23 (3) (2020) 133–135. https://doi.org/10.3779/j.issn.1009-3419.2020.03.01.

[12] S. Wang, B. Kang, et al., A Deep Learning Algorithm Using CT Images to Screen for Corona Virus Disease (COVID-19), 2020. https://doi.org/10.1101/2020.02.14.20023028. Available:.

[13] J. Zhao, Y. Zhang, et al., COVID-CT-Dataset: A CT Scan Dataset about COVID-19, 2020. Available: https://arxiv.org/abs/2003.13865.

[14] Detecting COVID-19 in X-Ray Images with Keras, TensorFlow, and Deep Learning, 2020. Available: https://www.pyimagesearch.com/2020/03/16/detecting-covid-19-in-x-ray-images-with-keras-tensorflow-and-deep-learning/.

[15] COVID-19 Task Force, 2020. Available: https://curae.ai/covid19-task-force/.

[16] G.E. Hinton, S. Osindero, Y. Teh, A fast learning algorithm for deep belief nets, Neural Comput. (18) (2006) 1527–1554.

[17] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning Book, 2018. Available: http://www.deeplearningbook.org/.

[18] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.

[19] I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al., Generative adversarial nets, Adv. Neural Inf. Process. Syst. (2014) 2672–2680.

[20] D.H. Hubel, T.N. Wiesel, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, J. Physiol. 160 (1) (1962) 106–154.

[21] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25 (1) (2012) 1097–1105.

[22] Y. LeCun, B. Boser, et al., Backpropagation applied to handwritten zip code recognition, Neural Comput. 1 (4) (1989) 541–551.

[23] M. Tan, Q.V. Le, EfficientNet: rethinking model scaling for convolutional neural networks, in: Thirty-sixth International Conference on Machine Learning (ICML), CA, USA, 2019.

[24] P.K. Diederik, B. Jimmy, ADAM: a method for stochastic optimization, in: 3rd International Conference for Learning Representations, San Diego, 2014.

[25] A.K. Gopalakrishna, T. Ozcelebi, et al., Relevance as a Metric for Evaluating Machine Learning Algorithms, International Workshop on Machine Learning and Data Mining in Pattern Recognition, Machine Learning and Data Mining in Pattern Recognition, 2013, pp. 195–208.

[26] E. Topol, Deep Medicine, Basic Books, 2019.

[27] J.A. Sacristán, T.D. Pharm, No big data without small data: learning health care systems begin and end with the individual patient, J. Eval. Clin. Pract. (2015). ISSN 1365-2753, Available: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jep.12350.

[28] M. Turak, Explainable Artificial Intelligence (XAI), Defense Advanced Research Projects Agency, Program Information, 2016. Available: https://www.darpa.mil/program/explainable-artificial-intelligence.

[29] F.K. Došilović, M. Brčić, N. Hlupić, Explainable artificial intelligence: a survey, 41st international convention on information and communication technology, Electron. & Microelectron. (MIPRO) (2018). https://doi.org/10.23919/MIPRO.2018.8400040.

[30] R.R. Selvaraju, M. Cogswell, et al., Grad-CAM: visual explanations from deep networks via gradient-based localization, Int. J. Comput. Vis. 128 (2020) 336–359.

[31] A.I. Open, Microscope Tool, 2020. Available: https://openai.com/blog/microscope/.

[32] M. Fredrikson, S. Jha, T. Ristenpart, Model inversion attacks that exploit confidence information and basic countermeasures, in: Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security, ACM, New York, NY, USA, 2015, pp. 1322–1333.

[33] M. Abadi, A. Chu, I.J. Goodfellow, et al., Deep learning with differential privacy, in: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, 2016, pp. 308–318.

[34] N. Phan, Y. Wang, et al., Differential privacy preservation for deep autoencoders: an application of human behavior prediction, in: AAAI, 2016, pp. 1309–1316.

[35] R. Shokri, V. Shmatikov, Privacy-preserving deep learning, in: SIGSAC, 2015, pp. 1310–1321.

[36] T. Zhu, G. Li, et al., Differential Privacy and Applications, Springer, 2017.

[37] C. Dwork, Differential privacy: a survey of results, in: International Conference on Theory and Applications of Models of Computation, 2008.

[38] C. Dwork, A firm foundation for private data analysis, Commun. ACM 54 (1) (2011) 86–95.

[39] A. Beimel, K. Nissim, S. Stemmer, Private Learning and Sanitization: Pure vs. Approximate Differential Privacy, CoRR, abs/1407.2674, 2014.

[40] B. Jayaraman, D. Evans, Evaluating differentially private machine learning in practice, in: 28th USENIX Security Symposium, Santa Clara, CA, USA, 2019.

[41] A. Haeberlen, B.C. Pierce, A. Narayan, Differential Privacy under Fire, 2011.

[42] E. Løkke, Privacy: Private Life in a Digital Society, Translated by Başaran, D, Koc University Press, 2018.

Further reading

[1] The Presidency of the Republic of Turkey, the Digital Transformation Office, Coronavirus, COVID-19 Outbreak Map, 2020. Available: https://corona.cbddo.gov.tr/.