Facilitating COVID recognition from X-rays with computer vision models and transfer learning

Aparna S. Varde[1,2] · Divydharshini Karthikeyan[1] · Weitian Wang[1]

Abstract

Multimedia data plays an important role in medicine and healthcare since EHR (Electronic Health Records) entail complex images and videos for analyzing patient data. In this article, we hypothesize that transfer learning with computer vision can be adequately harnessed on such data, more specifically chest X-rays, to learn from a few images for assisting accurate, efficient recognition of COVID. While researchers have analyzed medical data (including COVID data) using computer vision models, the main contributions of our study entail the following. Firstly, we conduct transfer learning using a *few images* from publicly available big data on chest X-rays, suitably adapting computer vision models with *data augmentation*. Secondly, we aim to find the *best fit models* to solve this problem, adjusting the number of samples for training and validation to obtain the *minimum number* of samples with *maximum accuracy*. Thirdly, our results indicate that combining chest radiography with transfer learning has the potential to improve the *accuracy* and *timeliness* of radiological interpretations of COVID in a *cost-effective manner*. Finally, we outline applications of this work during COVID and its recovery phases with future issues for research and development. This research exemplifies the use of multimedia technology and machine learning in healthcare.

Keywords AI in medicine · Electronic health records · Big data mining · Computer vision · Image recognition · Transfer learning

✉ Aparna S. Varde
vardea@montclair.edu

Divydharshini Karthikeyan
divyakarthi@gmail.com

Weitian Wang
wangw@montclair.edu

[1] School of Computing, Montclair State University, Montclair, NJ, USA

[2] Max Planck Institute for Informatics (Visiting Researcher), Saarbrucken, Germany

1 Introduction

1.1 Background and motivation

The COVID pandemic has spread exponentially across the globe, becoming a huge concern [44]. It has been highly stressful mentally, in addition to causing serious physical damage that has been fatal in some cases. Intensive efforts have been made worldwide since 2020 to seek fast, reliable solutions to help diagnose patients with COVID and triage patients for proper allocation of rather limited resources in combating this disease. In line with these efforts, there have been initiatives to conduct accurate testing of COVID. Machine learning plays a vital role here, e.g. in classification of relevant multimedia data such as images [39]. The task of image classification has long been a part of our daily lives, e.g. classification of X-ray images to diagnose illnesses. In the realm of machine learning, deep learning with Convolutional Neural Networks (CNNs) is among the popular approaches today. CNNs are based on biological processes such that their connectivity patterns mimic the organization of the visual cortex in our brain. They are deployed for image recognition in computer vision tasks. Accordingly, computer vision models [4] have been derived based on the fundamental concept of CNNs. These can be useful in image classification tasks to enhance accuracy and efficiency. A related paradigm is transfer learning [26] which involves storing knowledge acquired while solving a particular problem and then applying it to a different yet pertinent problem, e.g. the knowledge gained from recognizing airplanes could further be applied to recognize spacecraft. As per recent research on transfer learning, anomaly detection in small-sized medical image data is an achievable target, often with remarkable results [2]. Such work provides the motivation for our research. We aim to explore a solution based on computer vision models and transfer learning to address the problem of COVID detection from chest X-rays in an accurate and efficient manner, thriving on earlier research and making further contributions. Based on this background, we proceed to define our specific problem in this study.

1.2 Problem definition

We address the automated classification of COVID from image data in medicine, i.e. in the realm of E-health applications. Some COVID symptoms can be similar to those of another infectious disease, namely, pneumonia, yet there is a clear difference between these two diseases as well known to medical experts. Therefore, it is imperative to distinguish COVID cases from pneumonia cases while conducting the automated classification. Hence, we have the following goals in our problem:

- Classify multimedia data on chest X-ray images, as COVID-positive, pneumonia-positive or normal / healthy
- Conduct classification such that the images do not overlap and each image fits exactly into one class
- Find the best fit models such that high classification accuracy is achieved from few images efficiently

The data source for our work is ImageNet with its publicly available chest X-ray images. These contain image data on COVID and pneumonia, in addition to normal X-rays, i.e.

for people who tested negative in both these diseases. This pre-labeled data with appropriate diagnosis serves as the notion of correctness and hence forms the training data in our problem. Thus, the labeled chest X-ray images for COVID and pneumonia allow us to distinguish between the two and thereby aim to conduct the classification for COVID recognition by learning from the existing data.

1.3 Proposed solution approach

We propose a solution approach based on *transfer learning* using *computer vision models* over the multimedia data on chest X-rays images along with *data augmentation* in order to identify COVID symptoms by learning from a few images in existing labeled data. The computer vision models employed in this approach are VGG16, VGG19 and RestNet101 [4] since the architectures of these models are designed such that they give high accuracy along with optimization in learning, as evident from the fundamental research on them. VGG16 and VGG19 belong to the same family of models, incorporating convolutional layers in neural networks with kernel-sized filters lined up to mimic the effect of large filters for efficient and accurate learning. ResNet101 is noted for its skip connections aiding optimization analogous to pyramidal cells in the prefrontal cortex of our brain. Details on these will be elaborated later in the article as we explain our "Methods and Models".

It is to be noted that the multimedia data on chest X-ray images in public repositories is huge (Gigabytes and higher orders). This pertains to the *volume* aspect of big data. It is complex with considerable variations thereby addressing the *variety* aspect of big data. The inferences drawn from the data are crucial and need to be validated with adequate testing for accurate diagnosis, referring to its *veracity* aspect. In other words, this work addresses at least three Vs of big data, i.e. volume, variety and velocity. In order to deal with this multimedia big data efficiently and accurately we deploy the technique of transfer learning [26] as opposed to directly using standard deep learning models or other approaches. We focus on transfer learning in this study since we deal with a few images from big data. In our work we investigate transfer learning with computer vision models, focusing on accuracy as well as optimization in order to facilitate learning.

In addition to computer vision models and transfer learning, we deploy the paradigm of data augmentation [37] in this work. The fundamental concept of data augmentation involves effectively enhancing the training set using some methods without acquiring additional data samples. To some extent, this can be considered analogous to the principle of heteroscedasticity derived from the ancient Greek words *hetero* (meaning "different") and *skedasis* (meaning "dispersion"). As per this concept, the data augmentation techniques adapted in our study include flipping, rotation, shifting, shearing and zooming, all of which are explained with illustrations in our section on methods and models. These techniques are found to be useful because they serve to increase the training set size and variety such that learning from a few data samples is facilitated. This is an important focus of our work since we deal with multimedia big data freely available from a public website. It does not incur the overhead cost of acquiring greater data samples, nor does it need additional storage and high processing complexity. Note that data augmentation also helps to avoid overfitting, thereby gearing towards better accuracy in learning.

1.4 Contributions of study

As per this discussion and prequel, the main contributions of our work are as follows.

1. We perform transfer learning using a *few images* from free open source publicly available data on chest X-rays by suitably adapting selected computer vision models along with *data augmentation*.
2. We intend to find the *best fit models* in order to solve this problem by altering the number of samples for training and validation in order to obtain the *minimum number* of samples that provide *maximum accuracy*.
3. Our results reveal that a combination of chest radiography with transfer learning has the potential to improve the *accuracy* and *timeliness* of the radiological interpretation of COVID in a *cost-effective manner*.
4. We outline specific applications where this work would be useful, in the context of COVID and its aftermath, pointing out open issues for further research and development.

1.5 Manuscript organization

The rest of this manuscript is organized as follows. Section 2 overviews related work in the area. Section 3 explains our methods and models based on computer vision and transfer learning for COVID and pneumonia detection on multimedia data, i.e. chest X-ray images. Section 4 outlines our experimental evaluation and discussion. Section 5 states the potential applications of this work. Section 6 gives the conclusions with a roadmap for future work.

2 Related work

Paradigms within machine learning and computer vision have often been the subject of research in the context of medicine and healthcare, especially for multimedia data such as images and videos. The realm of this work is typically referred to as Computer-Aided Detection (CAD). In this overall realm, we focus on related work pertaining to image data in medicine, since our study deals with images of chest X-rays.

Some interesting research by Apostolopoulos et al. [2] conducts a study on a dataset entailing numerous X-ray images obtained from patients suffering from pneumonia, COVID as well as normal healthy cases. These images are collected to assist automatic detection of the coronavirus. In this study, multiple CNN architectures are deployed in order to classify medical images. These leverage computer vision models, namely, VGG19, Mobile Net, Inception, Xception, and Inception ResNetv2. Among all the models investigated in this study, VGG19 and MobileNet provide the best performance in terms of the resulting classification accuracy, in the 90% range. This study is often cited in the literature and is one of the inspiring works in the area providing a baseline for several other works. Our research in this article complements such work and is focused on finding the best fit in classification with few images and high accuracy.

Researchers Hemdan et al. [15] explore COVID diagnosis via deep learning CNN based approaches. The authors of this work conduct binary classification in their research, i.e. they aim to classify COVID versus non-COVID cases. The approaches they explore thrive on VGG19 and DenseNet201 computer vision models. This research yields good results with high accuracy in the 80% to 90% ranges. However, this study is only binary whereas we consider three classes. This is because some COVID symptoms can be analogous to pneumonia symptoms though not identical, hence we claim that it is important to consider those cases as well.

A Bayesian CNN model with a drop-weights approach is propounded in the study conducted by Ghoshal and Tucker [12] with execution on bacterial pneumonia, non-COVID viral pneumonia, COVID and normal cases. This is a very interesting piece of research that addresses bacterial diseases in addition to viral diseases. The results obtained using deep learning based on CNN are found to be satisfactory with respect to the concerned goals. While this study is promising with high accuracy, we intend to fine-tune the objectives by focusing on viral diseases only, hence we consider only those cases of pneumonia that are viral since these are more challenging to distinguish from COVID. This is because the ultimate goal of our work is to provide assistance in COVID detection, and it is clearly known that COVID is a viral illness, since it is caused by the coronavirus.

In other related work, X-ray images are subjected to deep feature extraction and SVM (support vector machine) classification by the research team of Sethy and Behera [36] to enable assistance in COVID detection. They extract data from the open source datasets provided by Kaggle, Open-I, and GitHub. In their approach, they first conduct deep feature extraction based on multiple CNN models used in computer vision, among which it is observed that ResNet50 outperforms the others. Hence, the ResNet50 model is used for further work in their research. After feature extraction, they execute a classification approach of SVM instead of using deep learning models as classifiers. Their argument is that deep learning for classification requires very large datasets for training the models while SVM can function with relatively smaller datasets (after feature extraction is already performed). This approach is found achieve acceptable accuracy, sensitivity and F1 scores, all in the 90% range. Motivated by the success of this work, we aim to further investigate the ResNet model; in our work, we consider RestNet101 instead to explore whether it can provide enhanced performance. In addition, we aim to investigate whether deep learning based approaches can achieve even better performance, and consider data augmentation to combat the issue of very large data required for training.

Researchers Zhang et al. [55] propound a model basically for assisting pneumonia diagnosis that is based on anomaly detection through the adaptation of deep learning from chest X-rays. Their procedure entails first distinguishing between viral pneumonia, non-viral pneumonia, and healthy controls as a single-class classification-based anomaly detection problem, based on which they propose their confidence-aware anomaly detection (CAAD) model. This comprises a shared feature extractor, an anomaly detection module, and a confidence prediction module. This approach helps in COVID diagnosis as well as indicated in their experiments with relevant datasets. It is observed in their study that this model achieves fairly good accuracy, sensitivity and specificity, all in the 80% range. However, it is found that this model has some threshold-based limitations. Our study in this paper is not limited based on thresholds, and thus has a wider scope. In addition, we aim to improve the performance of COVID detection beyond these accuracy values. Moreover, their study has its roots fundamentally in pneumonia detection which is thereafter adapted to COVID through a separate procedure. Our work considers COVID cases, pneumonia cases and healthy cases, all in one, and proposes an efficient straightforward approach to distinguish these from one another.

Recent work [32] highlights that technology is available to detect diseases from chest X-rays which could be an important tool in assisting clinicians during the triage process of screening COVID patients. While the PCR (Polymerase Chain Reaction) test results may take hours, this X-ray imaging method would give doctors instant results to make possible life-saving decisions. Novel deep learning based models have been built in this regard, hence the authors aim to explore the fine tuning of pre-trained CNNs for the classification of COVID using chest X-rays. They indicate that if fine-tuned pre-trained CNNs can offer

good classification results than other more sophisticated CNNs, then AI-based tools for diagnosing COVID using chest X-ray data can be cost-effective as well as efficient. While this research is promising, more study should be conducted analyzing X-rays at different stages of illness. Moreover, other studies also need to be conducted in this area to reinforce such claims so that they become widespread and can indeed be accepted for rapid detection of illnesses from chest X-rays. Pneumonia and COVID are both respiratory illnesses with similar symptoms.

A related [35] study aims to assist medical experts to better treat critically affected patients. X-rayed images of patients with pneumonia, COVID and normal patients are studied. A simple deep learning convolutional neural network is used to analyze these images. They produce good results which prompt more contemporary pre-trained models to better understand the patterns and behavior of each image. This study occurs at a later stage of the pandemic with a large and wide variation of data. One pre-trained model was able to give accuracy in the 90% range for identifying patients with COVID and pneumonia. However, this study does not address transfer learning to cater to larger datasets and harness the knowledge acquired from smaller datasets to deal with the issues of data availability during this pandemic. Nor does it address the efficiency aspect.

In other related work, edge computing has attracted attention due to its ability to instantly analyze and process data at the edge of the network where the data is collected [16]. As described by the authors, edge computing works by distributing the work of the device to the edge node near the physical server as well as the cloud. It enhances data security as sensitive information is processed at the edge nodes, making it suitable for analyzing data as sensitive as medical records. Hence, a hybrid model computing method that combines both edge and cloud computing can potentially be beneficial based on the level of the operation, and can be usable for medical data. However, in this work the authors do not elaborate on how such work can actually be harnessed for X-ray analysis.

Generative Adversarial Network (GAN) is an important turning point in generative modeling since introduced by Google Brain [28]. Which is also used to generate non-image data, such as voice and natural language. GAN uses two networks to conduct training: generator and discriminator. The generator converts random noise into a true-to-life image, whereas the discriminator distinguishes whether the input image is real or synthetic. GAN has made some major updates to the overall stability of the model throughout the recent years; however, some challenges still remain. Some researchers suggest that the mode collapse mitigation method is to change the distance measurement algorithm used for statistical distribution comparison, others suggest that the training instability nonlinear manifold theorem makes it difficult to reach the target, and that the initial algorithm should be randomly selected. Developers will continue to develop this technology with the goal that this stabilization will mature, and we will be able to train the model without any problems in the near future.

A related model [27] uses two neural networks: a generator that creates a realistic image, and a discriminator that distinguishes whether the image is fake or genuine. In this work, GAN has evolved to be able to generate non-image data such as voice and natural language. Boundary Equilibrium Generative Adversarial Network (BEGAN), which outperforms an earlier version of GANs, learns the latent space of the image while balancing the generator and discriminator. However, in BEGAN there is a fundamental problem with GANs called mode collapse. Mode collapse happens when the generator produces only a few images or a single image and is divided into partial collapse and complete collapse. Numerous updates are conducted in an attempt to solve mode collapse, improvements are envisaged, yet experimental results prove otherwise. Hence, while such technologies seem

interesting, we have not explored them in our research. We have built upon the success of transfer learning with other computer vision models, and have added to the success stories with our own experimentation.

In connection with this, there is research [31] using three pre-trained CNNs: AlexNet, GoogleNet and SqueezeNet to analyze chest X-rays of COVID and normal chest X-rays. The researched models in this work prove to be effective. Transfer learning is used and proves useful to identify false positive and false negative on enhanced data sets. This research has high accuracy which would give healthcare professionals the computer-aided diagnosis technology to improve the diagnosis of COVID patients more efficiently. However, the limitations of this study are that "due to database updates over time and the public availability of other data collections, it is impossible to carry out exact comparisons of the results reported" [31]. Moreover, they do not use data augmentation which we use in our work. Our own research builds upon such work, while aiming to investigate other models such as VGG19 and ResNet101, and adding data augmentation.

Another piece of research introduces two methods for fast and automated classification of chest X-ray images [1]. These are based on AlexNet and VGGNet16. Experimentation with these methods on chest X-rays provide average area under the curve values of 98% and 97%, respectively which seems promising. However, in this study only 12 X-rays are used thereby presenting limitations. A larger data set would show whether the models are generalizable or not. Models should be tested on more than one dataset to evaluate them in terms of generalizability.

In this broad context, it is to be noted that RT-PCR (Reverse Transcription - Polymerase Chain Reaction) is the most widely used test to diagnose COVID. Research shows that RT-PCR has been only 30% to 70% accurate in China [21]. While chest CT (Computerized Tomography) is considered to be more accurate, both CT scans and X-rays have been proposed for COVID diagnosis by various researchers. It has been asserted that using AI-based models with data available on repositories such as GitHub and Kaggle for diagnosing COVID, supplemented by RT-PCR and Rapid test kits will reduce the possibility of misdiagnosis. In line with this hypothesis, more studies need to be conducted on multiple datasets that can visually make diagnosis simpler and make it easier to detect anomalies. Hence, additional experimentation in this overall technology could assist in a faster and more efficient diagnosis of COVID patients. This is precisely where our study in this paper aims to make contributions. Our work adds to the research that assets such hypotheses to make things more convincing.

Finally, we wish to mention that the realm of E-health on the whole presents several interesting studies relevant to AI in general. The usefulness of the Medical Markup Language (MML) is explored by Tancer et al. [41] with respect to Electronic Health Records (EHR) to facilitate storage, information retrieval and data mining of heterogeneous data seamlessly. They address cloud and server storage, security, privacy, accessibility and related issues critical in healthcare data records that are highly sensitive. However, though this study provides the ground for storage and processing of heterogeneous data, it does not actually conduct machine learning over these health records; we focus on the actual machine learning in our research.

Prediction of air quality considering fine particle pollutants is conducted by Du et al. [10], addressing structured data as well as social media. They deal with PM2.5 pollutants, i.e. particulate matter with a diameter less than 2.5 μm, and incorporate health standards set by the Environmental Protection Agency (EPA) of the USA. They deploy machine learning approaches such as association rules, clustering and classification on multicity road traffic data from global sources as stored in the databanks of the WHO (World Health

Organization). In addition, they conduct sentiment analysis via polarity classification of tweets posted by the common public depicting their reactions to the reform measures taken by government agencies to counterbalance air pollution in specific areas. Opinion mining provides useful observations here to gauge public reactions on health-related issues. Although this research addresses machine learning pertinent to healthcare data, the data types here are relatively simpler, i.e. mostly numbers and characters. It does not address complex data such as images which we address in our work here.

Considering such a plethora of research that overlaps healthcare and computational techniques, our research in this article contributes the two cents here. Our study is somewhat orthogonal to such works in the literature, deriving inspiration from their results. Our contributions focus mainly on deploying deep transfer learning on chest X-rays related to COVID, pneumonia, and normal / healthy cases, by using a few images from big data on benchmarked open datasets and making comparisons with well-known pre-trained models. We intend to find the best fit by trying to minimize the number of data samples and learning time while yet aiming to achieve as high a classification accuracy as possible, via different models and data augmentation through transfer learning on the big data. We survey related works in detail to derive inspiration from earlier studies, gather useful inputs, and comprehend the shortcomings of some studies, in order to progress with our own research. One of our core contributions entails a best fit approach with respect to low data set size and high accuracy in decision support for COVID detection from chest X-rays.

3 Methods and models

We propose an approach to solve the problem of classifying chest X-ray images as COVID-positive, pneumonia-positive, or normal, using few images from big data, based on transfer learning with data augmentation using computer vision models. The models selected in this work are VGG16, VGG19 and RestNet101. This approach proposed in our study is executed with the goal of facilitating decision support in accurate and timely COVID detection.

Transfer learning in field of machine learning is a paradigm in which the knowledge mined by CNNs is transferred to solve a different but related problem comprising new data [53]. In the literature, it is found that the new data is usually larger in size. Transfer learning is often employed over data consisting of images. This entire process encompasses the training of a selected CNN for a particular task such as classification utilizing large-scale datasets. It is to be noted that data accessibility is a crucial aspect for good training in the transfer learning process because the CNN learns to extract important features of the images based on the initial training data provided. As per the capability of the given CNN to detect and harness the most significant image features, it is assessed whether the concerned model is acceptable for transfer learning, in order to proceed further.

Computer vision models such as VGG16 play an important role in this process. Since these models are designed based on the fundamental concept of a CNN, they are well-suited for transfer learning using feature extraction [6, 14, 38, 49–51]. Moreover, such models are found to be completion-wining models in the literature for image recognition tasks, considering accuracy and efficiency, hence they are considered a good choice in our approach for learning to detect COVID from chest X-ray images. We focus on VGG16 and VGG19 because they are developed with 16 and 19 layers respectively, and have multiple 3×3 kernel-sized filters such that these small filters lined in sequence,

effectively serve to simulate larger filters [38, 49, 50]. The design simplicity and generalization power make the VGG architecture a widely-used paradigm. In addition, we explore ResNet101 due to its residual learning framework that can ease the training of considerably deeper networks, because the ResNet model skips some connections using the hypothesis that deeper layers can learn as well as shallow ones [6, 14]. Since depth is crucial in many computer vision tasks, ResNet proffers a refined method of image recognition. On the whole, the computer vision models of VGG16, VGG19 and RestNet101, are observed to yield solutions with very high accuracy in state-of-the-art competitions involving images [14, 38, 51]. This is the reason for selecting them in our study. We make the disclaimer that these are not the only models for learning, since other researchers have experimented with different models and obtained promising results, as discussed in our related work. We choose to experiment with these computer vision models using transfer learning and data augmentation; the results obtained in our study can be useful for decision support, and our approach per se is potentially extendable to other models as needed.

Hence, we put forth a simple hypothesis that our proposed approach based on transfer learning with computer vision models helps address our goal of conducting classification to facilitate decision support in COVID detection by helping to finding the best fit for the minimum number of images that aim to achieve the maximum accuracy in learning, thereby also enabling efficiency that is imperative to ensure timeliness of detection. Our approach thus caters to the overall goal of decision support in COVID detection. The novelty of our approach entails the following aspects, making modest contributions.

- We perform transfer learning by utilizing a few images from publicly and freely accessible data on chest X-ray images, by appropriately adapting selected computer vision models along with data augmentation.
- We strive to find the best fit for the concerned models in order to solve this problem, by altering the number of samples used for training and validation so as to achieve the minimum number of samples with the maximum possible accuracy.
- We combine chest radiography with transfer learning with the claim that it has the potential to augment both the accuracy as well as the timeliness of radiological interpretations of COVID in a cost-effective manner.

We do not make the claim that ours is essentially the most optimal study in this area. Our work fits into the plethora of related research, and is orthogonal to the existing literature. We take the stand that: as more research is conducted in this field, the resulting findings are more likely to be accepted by the medical community, and more organizations are likely to deploy chest X-ray classification for decision support in COVID detection to enhance accuracy, efficiency and cost-effectiveness, and thereby make COVID detection more widely prevalent, even in areas that have a shortage of healthcare professionals and testing kits. More on this is discussed in our Applications section later in the paper. We now proceed to describe its overall framework of our approach and the details of its steps. Transfer learning as well as computer vision models are elaborated more in the subsections below. In those respective subsections, we provide justifications on why a specific type of transfer learning approach is selected, and why we choose the concerned computer vision models, explaining how that facilitates COVID detection with timeliness and accuracy.

Our proposed approach for decision support in automated classification of COVID, pneumonia and normal / healthy cases from chest X-ray image data is synopsized in the framework of Fig. 1. This has 3 main steps as follows.

Fig. 1 Proposed methodology:
overall framework
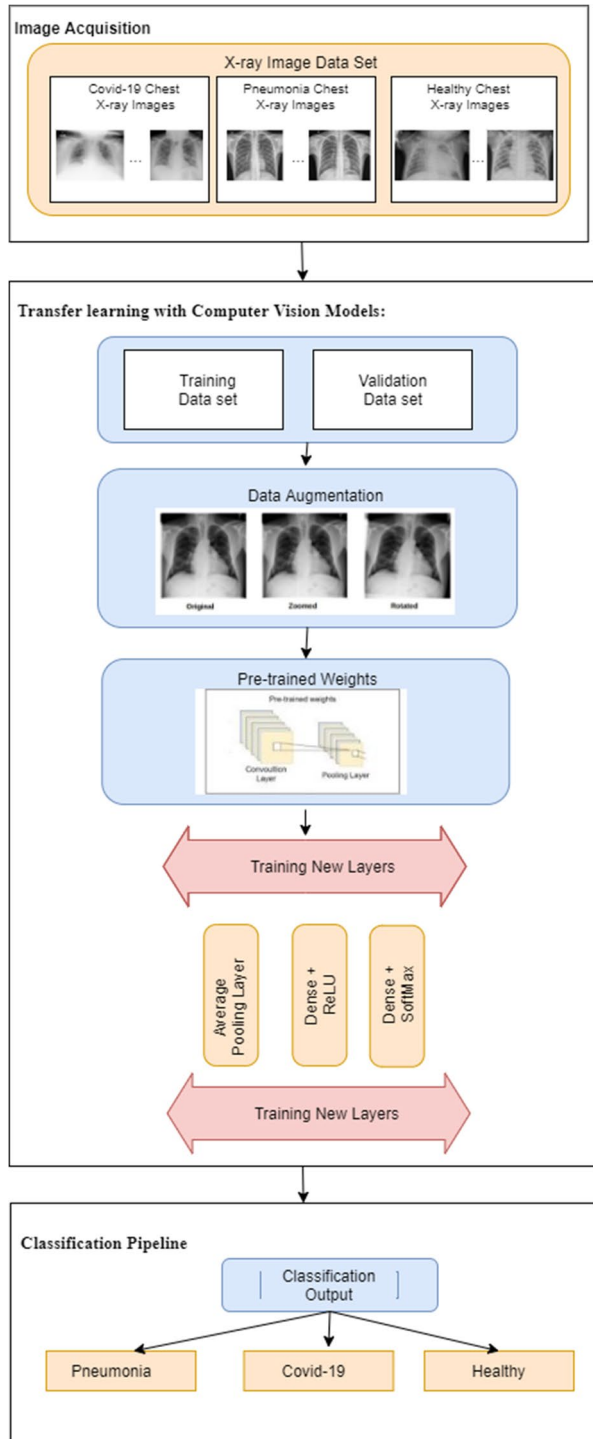
1. **Image Acquisition:** This involves collecting chest X-ray images from a benchmark open dataset and preprocessing them for suitable dataset generation to proceed with the learning.
2. **Transfer learning with Computer Vision Models:** This entails the deployment of computer vision models and transfer learning embodying data augmentation.
3. **Classification Pipeline:** This focuses on development of the actual classification approach to guide the diagnosis of new patient cases for decision support.

In the first step of our approach, a benchmark open dataset (in this case, ImageNet) is utilized as the source of big data on images needed for learning. These images are acquired, fundamental data preprocessing is conducted, and they are saved in 3 folders, namely, COVID, pneumonia, and normal / healthy cases. The second step deals with the crucial task of feature extraction which occurs via the utilization of pre-trained CNN architectures and newly trained layers, wherein some computer vision models are employed along with transfer learning. Data augmentation is performed in this step to effectively increase the number of data samples and to avoid overfitting during learning. In the third step, an actual classification pipeline is constructed that serves to guide the adequate diagnosis of new cases on COVID, pneumonia, and healthy ones. These are based on the 3 respective classes in the image acquisition step, utilizing the knowledge discovered via transfer learning. We explain details of our methods in the following subsections.

3.1 Image acquisition

The first step of our proposed methodology focuses on generating a suitable dataset for learning. Images in the data are acquired from the publicly available source ImageNet (large image database used for training highly efficient models including those in this study). We briefly describe our data below.

3.1.1 Data description

The ImageNet dataset has numerous chest X-ray images constituting those of COVID-positive, pneumonia positive and normal patients. Figures 2, 3 and 4 show examples of chest X-rays pertaining to a pneumonia positive, COVID-positive, and healthy case respectively. The model is expected to make a prediction on the probability of the 3 classes as seen in the labels above the respective images. This is useful in the image classification in our pipeline, to provide decision support for diagnosis.

Since these images are on a publicly available dataset that is freely accessible, the data has already been collected taking into account issues of medical privacy. This is the reason that even though the data is of a very sensitive nature, we can apply preprocessing operations on the data, and thereafter proceed with applying data augmentation techniques on it, e.g. rotation, flipping etc. (as elaborated in the next subsection 3.2 where we discuss data augmentation). It is to be noted that there are multiple images available, so the issue of data imbalance is not a major concern here, because the data has been sampled from a variety of patients. In fact, it constitutes big data in terms of volume, variety, and veracity (as explained in the Introduction section). However, in our work we aim to select smaller samples of the data in our experimentation in order to find the best fit between accuracy and efficiency, since our goal is to ensure timeliness of detection in addition to accurate

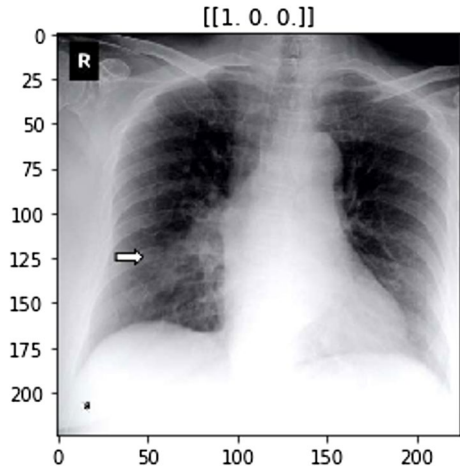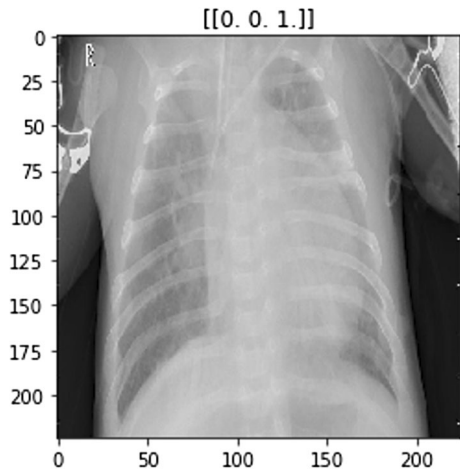**Fig. 2** Model image example for COVID-positive case

**Fig. 3** Model image example for pneumonia-positive case

diagnosis. In this context, data augmentation and transfer learning are helpful using the computer vision models over the data.

3.1.2 Data preprocessing

The data is first subjected to numerosity reduction using the selective sampling techniques in machine learning. The images resulting from this preprocessing are integrated to produce a sizeable dataset of 1GB, large enough for accurate learning and yet small enough for efficient learning (note that the original big data has several GB). Classes in this dataset constitute cases of confirmed COVID, confirmed pneumonia, and healthy chest X-ray images (see image thumbnails within Fig. 1).

Images in the dataset are mapped and stored in a comma-separated variables (csv) file with appropriate labeling and metadata: these are labeled as 0 for COVID-positive cases, 1 for pneumonia positive cases, and 2 for normal cases. Using these labels as guidelines,

Fig. 4 Model image example for normal / healthy case
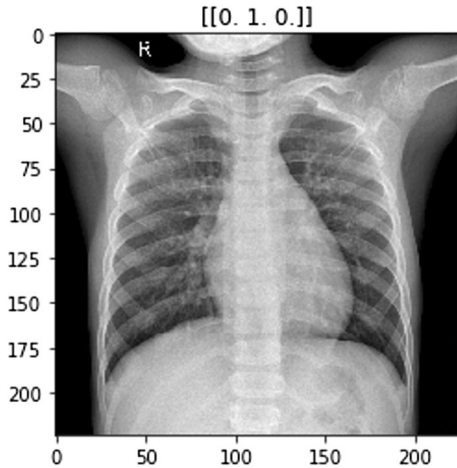
a program is coded (using Python in our implementation) for mapping and moving the images of each class into a separate folder. We create a separate folder for each class in order to distinguish between the classes and the corresponding input. Note that combining, labeling, and separating the images are crucial activities in this step of "Image Acquisition", allowing us to obtain a suitable number of images.

The programming language Python plays a vital role in our execution of this study. We utilize this software to code all the programs in this work. (Note that relevant parts of this code can be made available to some interested users upon request). The powerful libraries of Python and its framework for machine learning coupled with the simplicity of its repertoire make Python an excellent choice for coding within the implementation of our research.

3.2 Transfer learning with computer vision models and data augmentation

The fundamental concept of transfer learning has been previewed in the Introduction section of this article. This paradigm is well-known to many readers, yet in this section we dwell upon it with specific reference to our study. We first explain the transfer learning paradigm and then focus on the computer vision models used for transfer learning in our work. Thereafter, we explain the data augmentation concept with particular reference to our research.

3.2.1 Transfer learning

Transfer learning is a paradigm in the realm of machine learning where the knowledge mined in a given task is transferred to solve a different but related task involving new data. The new data is typically integrated within an existing model [53]. In deep learning, this process often entails the initial training of a CNN for a specific task such as classification on large datasets. Availability of data for initial training is a significant factor for adequate training because the CNN learns to extract the important features of the image accordingly. Based on the capacity of the given CNN model to extract the most noteworthy image features, it is judged whether the respective model is suitable for transfer learning. Thereafter,

the CNN is used to process a new dataset of images of a different type, thereby extracting features obtained using the knowledge from the initial training. There are 2 common strategies to deploy the capabilities of a pre-trained CNN.

Strategy 1: In this strategy, feature extraction via transfer learning [53] occurs such that the original pre-trained model retains its initial architecture as well as all its learned weights. Therefore, the pre-trained model in this strategy is used as a feature extractor and the extracted features therein are inserted into a new network that conducts the actual classification task. This strategy is typically used to circumvent the computational costs via training a very deep network from scratch, or for retaining useful feature extractors trained at the initial stage.

Strategy 2: This strategy involves specific variations applied to pre-trained models for obtaining optimal results. The variations typically consist of architectural adjustments and parameter tuning. Hence, only certain knowledge mined from the previous task is retained, while some new trainable parameters are also incorporated within the network. Note that this requires a considerable amount of training on huge datasets in order to be truly beneficial.

Since the focus of our research is to learn from a few images in the big data and find the best fit from an accuracy and efficiency standpoint, we prefer to avoid such exhaustive training. Hence, we focus on the first strategy. In the method we propose to use, transfer learning through computer vision models is adapted by incorporating Strategy 1 as described here, along with data augmentation as explained later. Transfer learning in computer vision allows us to take away the last few layers from a given model by integrating the pre-trained model (trained on thousands of images) and used on our pertinent image data.

3.2.2 Computer vision models

Based on the background presented here and our literature study, the computer vision models selected for this study are VGG16, VGG19, and ResNet 101 described as follows.

VGG16: This is a CNN-based model proposed by Simonyan and Zisserman from University of Oxford [38]. It is typically found to achieve an accuracy of 92.7% and attains the top-5 test accuracy in ImageNet. This model is portrayed in Fig. 5 as observed from sources in the literature [49].

Note that the abbreviation VGG stands for Visual Geometry Group, i.e. the name of the group at Oxford where this model has been developed while the number 16 in VGG16 refers to its 16 layers in which the weights and bias parameters are learned. It has 138 million parameters thus constitutes a very large network. VGG16 can be simplistic because it uses only 3×3 convolutional layers stacked over each other; these keep increasing in depth yet keep decreasing in the size handled by maximum pooling. The main intuition in VGG architecture, with reference to 16 or 19 layers is the multiple 3×3 kernel-sized filters. These small filters lined up in a sequence can mimic the effect of large filters [49].

The simplicity in design and the generalization power of the VGG architecture make it a suitable choice in many applications. Multiple scenarios are tested within the VGG16
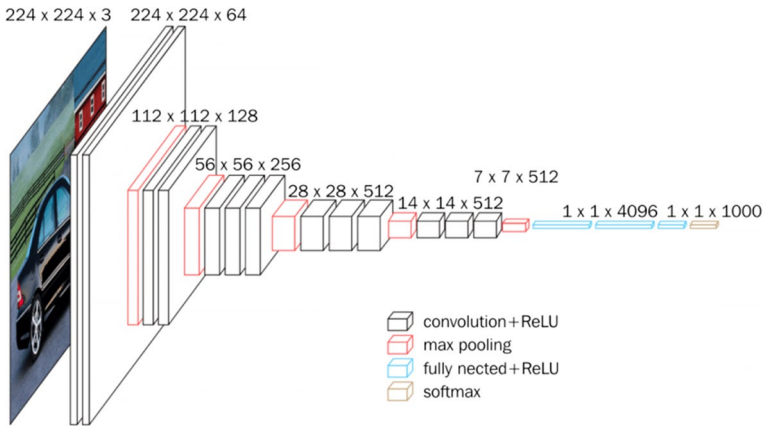
Fig. 5 VGG16 architecture [49]

model and they demonstrate high accuracy. On the basis of these factors, we consider this model as a viable choice in our research.

VGG19: This is another CNN-based model, its architecture being fundamentally similar to that of VGG16, the notable and obvious difference being that VGG19 has 19 convolutional layers instead of 16 thereby making it a larger network. A significant aspect of VGG19 is that it is a competition-winning model since it can classify images into 1000 object categories: including stationery items such as pencils, and living creatures such as various animals [50]. Given its training exposure, VGG19 learns rich feature representations for a broad spectrum of images. The network has an image input size of 224×224. Figure 6, as found in the literature [49], portrays the VGG19 architecture.
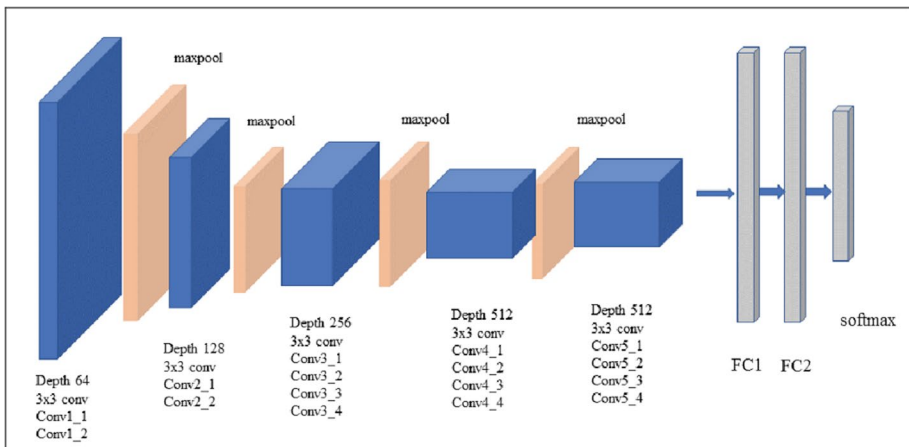
Fig. 6 VGG19 architecture [49]

It is important to note that in our research, we use two variations of VGG models to ascertain whether the extra layers make any difference to the classification accuracy. VGG models are probable choices for our work, yet they have some drawbacks [49]. It is observed in the literature [23–26] that they are very slow to train, their weights are very large and they also consume a substantial amount of bandwidth. Thus, it may be feasible to use smaller network architectures based on the classification tasks. On the other hand, VGG models have demonstrated very high accuracy as compared to several highly acclaimed models. The aspect of increasing the depth in the VGG architectures contribute to them being suitable options as high performance models.

Res101: This model has a CNN that is 101 layers deep. The term ResNet stands for Residual Network and 101 refers to its number of layers. This model builds upon simulating the function of the pyramidal cells in the prefrontal cortex of our brain. It is based on the concept that deeper neural networks are more difficult to train and hence a residual learning framework eases the training of networks that are substantially deeper [14]. To address issues with very deep neural networks, the ResNet101 model proposes to skip some connections, i.e. it takes shortcuts to jump over some layers hypothesizing that deeper layers should be able to learn as well as shallow layers [6] analogous to the learning in our prefrontal cortex. The founders of this model, He et al. [6], claim that they propose to copy activations from shallow layers and set additional layers to identify the mapping. Figure 7, observed in the literature [6], displays the ResNet101 model.

The ImageNet dataset has evaluated networks with a depth of up to 152 layers, i.e. 8 times deeper than VGG models and with lower complexity. An ensemble ResNet incurs an error rate as low as 3.57% on the ImageNet test set [14]. Since depth is of great significance in many computer vision tasks, ResNet provides a refined method of object recognition.

As in the VGG models, ResNet models provide solutions with very high accuracy in ImageNet competitions. It is found that ResNet offers reformulated layers for learning residual functions with reference to layer inputs rather than learning unreferenced functions [14]. ResNet frameworks are easier to optimize and they gain accuracy from considerably increased depth as noted in the literature. Hence, we explore ResNet101 in our research due to its ability for optimization and accuracy. This is in addition to exploring
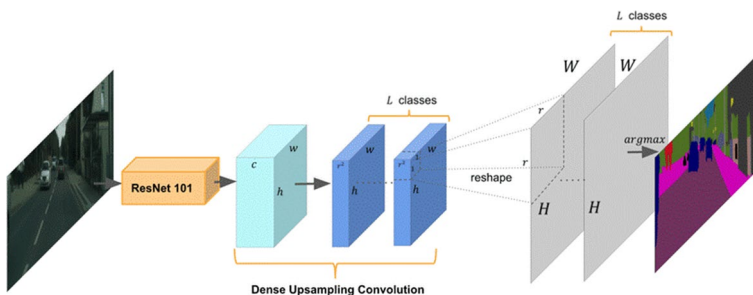

Fig. 7 ResNet101 architecture [51]

VGG16 and VGG19, in order to present a comparative study on the performance of the models.

3.2.3 Data augmentation

It is important to dwell on the concept of data augmentation as relevant to computer vision, transfer learning and heterogeneous data. Data augmentation is a concept that effectively enhances the training set of a network and is useful mainly when the training dataset contains only a few samples; being especially beneficial to avoid overfitting [13]. Oftentimes geometric distortions or deformations are applied either for increasing the number of samples to enable deep network training or for balancing out the size of datasets. If there are microscopic images, the techniques of shift and rotation invariance, as well as robustness for deformations and grey value, are some of the required modifications applied to each image of the training set [20]. These techniques prove to be rapid, reproducible, and consistent. Seemingly increasing the number of data samples via data augmentation may efficiently improve CNN's training and testing accuracy, reduce the loss, and improve the network's robustness. This makes it a suitable choice for computer vision models, especially when deployed with transfer learning. However, it is to be noted that heavy data augmentation requires careful thought and vigilance, as this may produce unrealistic images and confuse CNN.

In our work, we sample the original multimedia big data on images from a free publicly available benchmark open dataset to a smaller dataset of size 1GB. Therefore, we find it advisable to deploy data augmentation within the overall transfer learning framework based on computer vision models since this helps to enhance the effective data available for learning and yet does not incur the overhead of additional storage or added computational complexity while also positively impacting the accuracy standpoint. It helps to reduce overfitting by providing greater size and variety to the training data without acquiring more samples.

Figure 8 demonstrates the data augmentation techniques such as rotate, shift, shear, zoom, rescale, flipping, and Gaussian performed on X-ray images. We describe these approaches.

Flipping: In this technique, we can flip images horizontally and vertically. Some frameworks do not provide a function for vertical flips. Note that a vertical flip is equivalent to rotating an image by 180 degrees and then performing a horizontal flip.

Rotation: Image rotation is one of the most commonly used augmentation techniques. It can help our model become robust to the changes in the orientation of objects. Although we rotate the image, the information within the image remains the same.

Shifting: By shifting the images, we can change the position of the objects in the image and hence offer more variety to the model. This eventually leads to more generalization.

Shearing: Shearing is a technique used to transform the orientation of the image. Hence, it appears as though the image is somewhat different, thereby providing robustness.

Zooming: The zooming operation allows us to either zoom in or zooms out of an image. This also seemingly enhances the dataset size since a zoomed image can look different from the original one with respect to training and testing datasets.
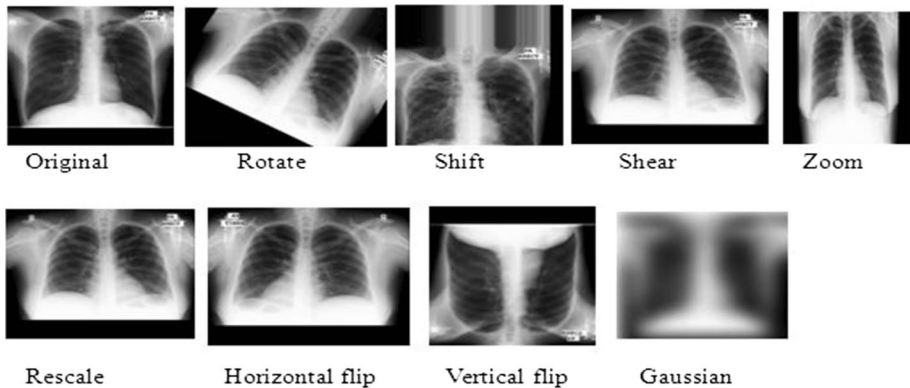
Fig. 8 Data Augmentation techniques performed on an X-ray image

We wish to reemphasize the fact that we collect the images used in our study from a publicly available dataset that has free and open access. Hence, this already circumvents the issues of medical privacy because the images entail the consent of the corresponding patients and get stored after incorporating privacy-preserving techniques. This is the reason that although we deal with highly sensitive data, we can apply various preprocessing operations, and then execute data augmentation techniques, e.g. rotation, flipping etc. Note that since there are numerous images available, the problem of data imbalance is not a major concern, because the data samples emanate from a variety of patients. We select a few images from this big data during the learning process aiming to find a best fit between number of data samples and accuracy in learning. However, the data per se is abundant due to which imbalance is no problem.

3.3 Classification pipeline

The classification pipeline in our proposed methodology is portrayed in Fig. 9 herein. This involves reading in the images sorted into subdirectories with the subdirectory name used as the corresponding image class. The images are then randomly split using a 75:25 ratio for training and validation data subsets respectively. Resizing of the images occurs into a 224×224 pixels array. The image data is augmented with the data augmentation methods of shear, zoom, horizontal flipping, and rescaling described herewith that are often used in scientific data mining applications [13]. The algorithm proposed in the classification pipeline in our work is outlined as Algorithm 1 herewith. This is coded into the corresponding Python program and is used in the automated detection of COVID, pneumonia and healthy cases thereby helping to provide decision support.

In this classification pipeline, the computer vision models VGG16, VGG19, and ResNet101 are used with transfer learning for feature extraction. These models learn from images in our training dataset. All convolutional layers in the models here get activated by the Softmax [18]. The CNNs are compiled by deploying an optimization method called Adam [18]. Training is conducted for multiple epochs with suitable batch sizes. After this
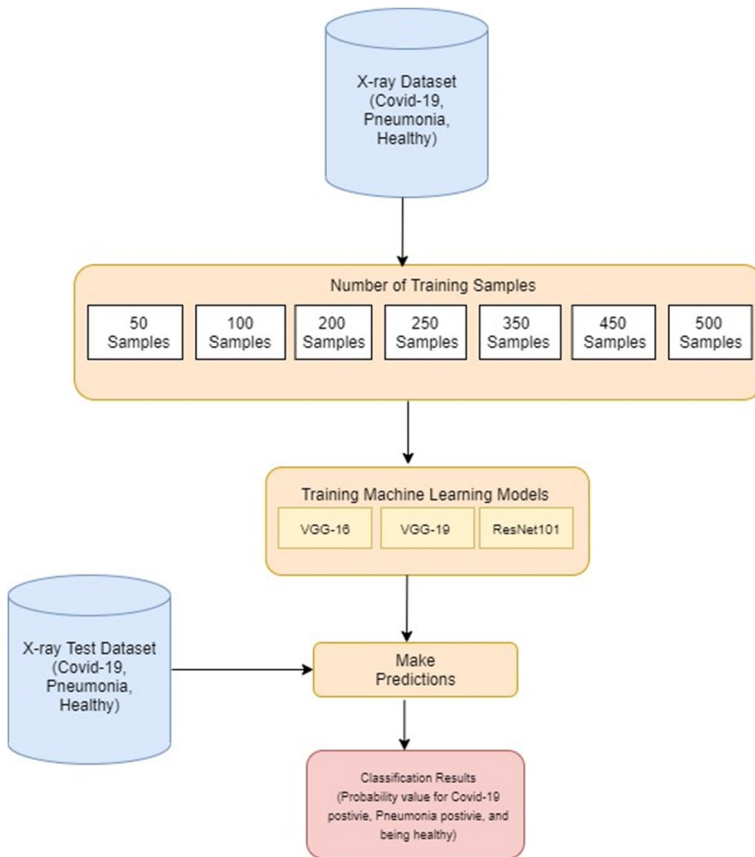
Fig. 9 Classification pipeline in the methodology

Input: X-ray images for COVID, pneumonia, normal healthy cases
1. Determine training set, test set and validation set
2. Perform data preprocessing on training set to enhance quality
3. Build a baseline with fully connected layers.
4. Use transfer learning to train classification models for 3 classes: COVID, pneumonia, healthy
5. Train VGG16, VGG19, ResNet101 models with training set
6. Return the trained classification models
7. Plot accuracy, loss and build the confusion matrix for the models
8. Use the models to make a prediction on unseen (new) images dataset
Output: New image predicted as COVID / pneumonia / healthy

Algorithm 1: Classification of Chest X-rays for COVID detection

procedure, we use unseen test data comprising chest X-ray images for classification to aid automated detection of new patient cases.

Considering the explanation of the above models, and examples of images used in the models, we now present a flowchart that illustrates the working of our proposed approach, as implemented, in order to produce the experimental results that we describe in our next

section on evaluation and discussion. This flowchart appears in Fig. 10. Our approach is implemented in Python.

4 Evaluation and discussion

We conduct the evaluation of our proposed methodology in this research. Independently, different versions of the pre-trained CNNs model are fine-tuned and the prediction is conducted on the test set based on the classification pipeline presented herewith. This evaluation occurs based on different criteria and the computer vision models VGG16, VGG19 and ResNet101 are compared. The resulting classification models to detect positive cases of COVID, positive cases of pneumonia, and healthy conditions based on popular pre-trained models are evaluated for performance.

4.1 Evaluation criteria

The commonly used criteria of accuracy, loss and confusion matrix are used for evaluation as follows.

Accuracy: Accuracy (A) is the fraction/percentage of correctly classified samples. It is calculated in Eq. (1) as the number of True Positives (TP) plus True Negatives (TN) divided by the total number of samples in the dataset (i.e. number of chest X-ray images). For example, a TP is an image detected as being COVID-positive when the corresponding patient actually has COVID, while a TN is an image detected as COVID-
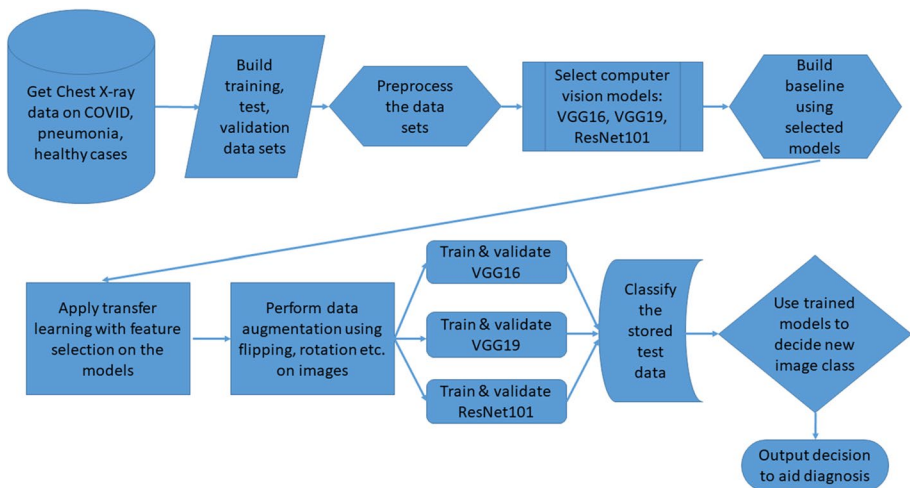
Fig. 10 Flowchart for execution of approach to illustrate implementation for experimentation

negative when the patient does not have this disease. Both these constitute correct classifications. Thus we have,

$$A = \frac{(TP + TN)}{D} \tag{1}$$

where TP is the number of True Positives, TN is the number of True Negatives, and D is the number of Data Samples.

Loss (L): The term Loss (L) pertains to the cross-entropy loss or the uncertainty of a classification based on how much the classification varies from the true value. It helps comprehend how close the predicted distribution is to the true or correct distribution. The loss, i.e. cross-entropy loss is calculated using Eq. (2) herewith.

$$L(p, q) = -\sum_{x=1}^{c} p(x) \log q(x) \tag{2}$$

where p(x) is the correct probability, q(x) is the predicted probability, and c is the number of classes (c = 3 here).

Therefore, the accuracy criterion helps to indicate whether the classification is equal to the true value or not (correct or wrong), while the term loss indicates the extent to which it is incorrect, i.e. how far it deviates from correctness.

Confusion matrix: A confusion matrix examines whether the prediction is consistent with the actual results. It is good for evaluating the classifier for multi-class objects. The focus is on the general properties of the learning algorithm to address the obstacles with multi-class classification and measure quality with the confusion matrix. The instances in a predicted class represent each row of the matrix, while each column represents the instances in an actual or correct class. The diagonal of the confusion matrix represents correctly classified data samples. We use this in our work to assess the images are being correctly or incorrectly detected as COVID-positive, pneumonia positive and normal / healthy cases.

4.2 Summary of experiments

In this work, for the classifiers trained from scratch, the Adam optimizer is used with an initial learning rate of 0.0005 and a batch size of 64. We train our models with only 50 epochs since we aim to maintain low training time as an important objective of this study. In all cases, the categorical cross-entropy is used as the loss function. The dataset used is randomly split into two independent datasets with 75% and 25% used for training and testing respectively. Figures 11, 12, 13, 14, 15, 16, 17, 18, 19 portray the evaluation with respect to accuracy, loss and confusion matrix for VGG16, VGG19 and ResNet101 as indicated in the captions alongside the figures. The number of samples used for evaluation is indicated in each figure. For example, VGG16 (Fig. 11) gives the best accuracy with 350 samples and VGG19 (Fig. 14) achieves that with 100 samples.

4.3 Discussion on results

As confirmed by the results of our experimentation, there is a consistency between the predicted and actual results, implying good performance of the models in the classification of multi-class objects. Table 1 herewith presents a summary of the accuracy

Fig. 11 Best accuracy achieved
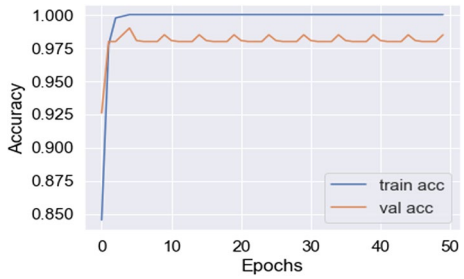in VGG16 (350 samples)

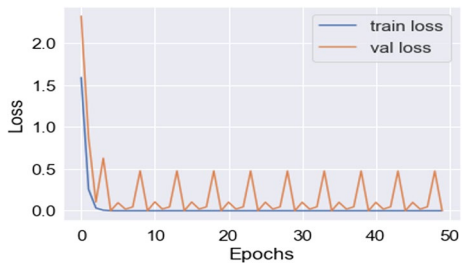**Fig. 12** Lowest loss achieved in
VGG16 (350 samples)

Fig. 13 Confusion matrix for
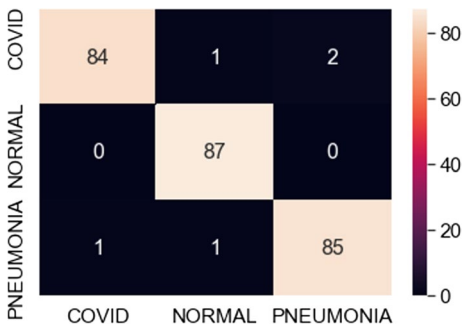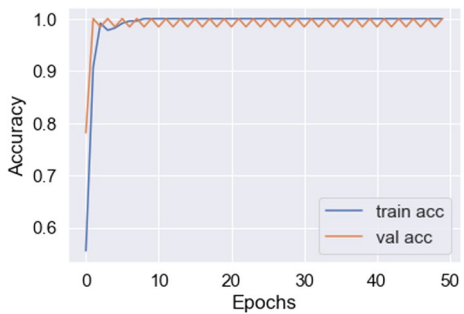best fit VGG16 model (350
samples)

**Fig. 14** Best accuracy achieved
in VGG19 (100 samples)

obtained with the different models being used in our classification pipeline, i.e. VGG16,
VGG19 and ResNet101 along with the number of data samples used per model. In our
experiments, we try to find the best fit model with constant epochs varying the number

Fig. 15 Lowest loss achieved in VGG19 (100 samples)
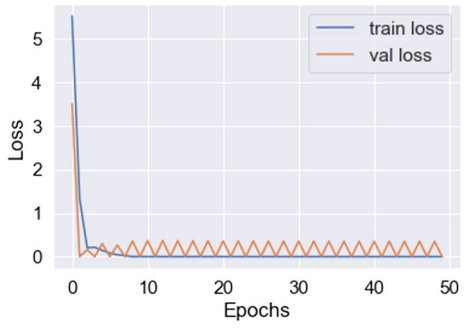
Fig. 16 Confusion matrix for best fit VGG19 model (100 samples)

Fig. 17 Best accuracy achieved in ResNet101 (500 samples)
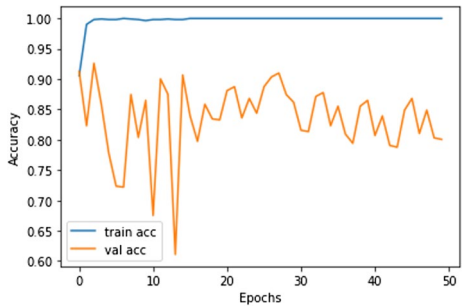
Fig. 18 Lowest loss achieved in ResNet101 (500 samples)
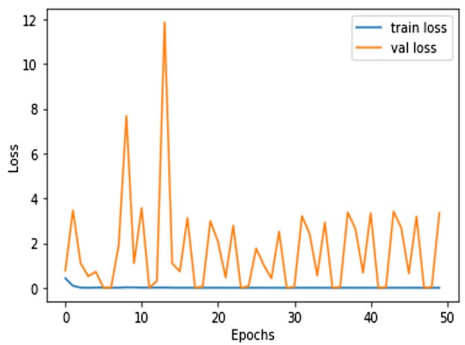
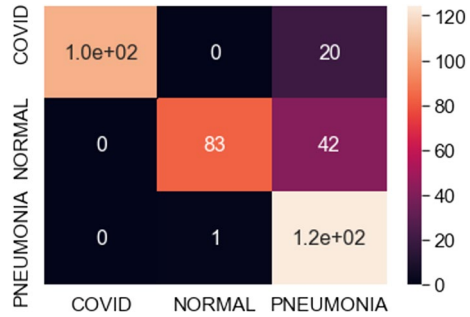**Fig. 19** Confusion matrix for best fit ResNet model (500 samples)

Table 1 Tabulation of accuracy of three models for varied sample size

Models	Number of Samples	Accuracy
VGG-16	50 Samples	97
	100 Samples	96
	200 Samples	97
	250 Samples	96
	350 Samples	98
	450 Samples	95
	500 Samples	97
VGG-19	50 Samples	97
	100 Samples	99
	200 Samples	97
	250 Samples	99
	350 Samples	98
	450 Samples	94
	500 Samples	97
ResNet101	50 Samples	58
	100 Samples	67
	200 Samples	72
	250 Samples	75
	350 Samples	76
	450 Samples	80
	500 Samples	83

of samples for training and validation. Note that in our learning process, we try to obtain high accuracy with as few images as possible. This is a striking contribution of our work, not found in any other studies heretofore, to the best of our knowledge. This is includes the studies surveyed in our related work section. Hence, this is a novel aspect of our study.

The observations in Table 1 indicate that VGG16 and VGG19 perform well compared to ResNet101. The best-case accuracy for VGG16 is 98% for 350 samples. The best-case accuracy for VGG19 is 99% for 100 samples. On the other hand, it is noticed that the accuracy for ResNet101 increases with the number of training samples whereas VGG16 and VGG19 perform well even with the smaller dataset. The best-case accuracy

in all our experiments is obtained for ResNet101 is 83% for 500 samples. This table thus provides the best fit for each model which can be observed with respect to accuracy and number of samples. Hence, our proposed solution approach in this work proves viable for the automated classification of COVID, pneumonia and healthy cases. This thereby paves the way for several potential applications as discussed next. As of now, the metrics used in this study are accuracy, loss and confusion metrics. This constitutes a limitation of this work. In the future, more metrics can potentially be developed to evaluate the proposed approach.

In terms of comparative evaluation, there are other studies conducted as mentioned in the Related Work section. We have conducted detailed investigation of these studies and summarized them in our discussions on these related works [1, 2, 12, 15, 16, 21, 27, 28, 31, 32, 35, 36, 55]. An important observation is that all these studies provide accuracy levels in approximately the 80% and 90% ranges. Each of these studies has their respective limitations as elaborated in that section while discussing the respective papers. For example, the study by Hemdan et al. [15] deals with binary classification only whereas we experiment with 3 classes: COVID-positive, pneumonia-positive and normal / healthy cases since the symptoms of pneumonia are somewhat analogous to those of COVID, yet there are differences that help in distinguishing these 2 diseases. On the other hand, the research of Ghoshal and Tucker [12] considers 4 classes and yields promising results but one of their classes pertains to bacterial pneumonia thus broadening the scope of the work. We prefer to be more focused with viral diseases only and hence consider 3 classes, as that has the added advantage of working with relatively less data and yet obtaining the required inferences, thus adding to efficiency and accuracy as well as being more specific and focused. Zhang et al. [55] achieve accuracy in the 80% range and consider other evaluation metrics, however, their study has some threshold-related limitations while our work is not limited based on thresholds. All these studies provide some benchmarks for our work. We intend to perform at least as well as or better than existing studies with the added contribution of learning from a few images in the big data on chest X-rays, and aiming to find the best fit in terms of number of data samples versus accuracy, as synopsized in Table 1. From our results here, it is evident that the VGG models used in our research yield high accuracy levels with as few as 50 samples and that additional samples marginally increase the accuracy. Hence, we can infer that using fewer samples and adapting the approach used in our research in this paper can be useful with respect to decision support in timely and efficient COVID diagnosis. We elaborate on this further in the section on Applications.

Furthermore, there are related studies conducted in other areas on different types of scientific datasets with which we can draw a parallel as per our work. An interesting study [11] focuses on using the Support Vector Regression (SVR) model with Empirical Mode Decomposition (EMD) to help predict electric load forecasts. When using the SVR model with decomposition to observe whether there is any improvement in forecasting, the authors find that training effects are 0.9912 and 0.9901 and forecasting effects are 0.9875 and 0.9887 over 7 and 23 days of recordings respectively. When using their EMD based SVR approach, they notice at the 0.025 and 0.05 significance levels that there is a significant increase in the accuracy compared to SVR alone. Another piece of research [5] proposes to use Least Squares Support Vector Machine with Fuzzy Time Series and Global Harmony Search Algorithm to help predict electric load forecasts. They obtain monthly electric loads on the 11th of each month and fuzz it. This yields a time series mostly less than 0.1 with a few numbers in the 0.8-0.9 range. After determining about 20 sets to use, they figure out the optimal parameter of GHSA (global search harmony) as $\gamma = 9746.7$ and $\sigma = 30.4$. Next, they used the LSSVM model to train the data then defuzzify it giving a

multiplier of each month right around the value of 1. Finally, after using various training methods, the GHSA-FSM-LSSVM provides an average accuracy rating of around 96%, which is very high. In another recent study [22], the authors investigate using quantum computing to help improve the "Bat Algorithms" global search rate. They hypothesize that this is possibly achievable by "increasing the diversity of the population and improving its ability to jump out of a local optimum, so that the population can maintain its ability to optimize continuously in the iterative process". After conducting multiple tests, they conclude that *QBA* and *CCQBA* are two approaches that offer the best improvements under the right circumstances. They believe to see the best improvement in the *QBA* using it along with the 3D cat mapping function and the X-condition cloud model. With *CCQBA,* they believe that when there is a multiple function problem, the values of certain parameters mixexe and dissca can be set to 0.9 and 0.35 respectively. When there are unimodal benchmark functions and simple multimodal functions, they can be set to 0.35 and 0.75 in solving the problem.

Hence, multiple such studies can serve as useful approaches in the literature to conduct some type of forecasting which falls in the general category of predictive analysis and decision support. Our work in this paper falls within this broad realm as well. The statistical tests conducted in the above studies are more applicable to electric load data used there and the concerned applications. We can potentially consider such tests in terms of future work as needed. In our current study in this paper, we have focused on the evaluation paradigms of accuracy, loss and confusion matrix. We have obtained accuracy levels acceptable for decision support, as compared to other studies in the literature on medical diagnosis; moreover, we have the added contribution of learning from few images in big data, incurring the advantages of accuracy, efficiency and timeliness. While we do not wish to make a very strong claim that our study is essentially superior to the work of others, we make a modest claim that it is orthogonal to the state-of-the-art and can be useful for decision support in COVID detection. We now proceed to describe its applications with examples.

5 Applications

Our proposed approach here can be a supplementary tool for services to screen COVID patients in emergency medical support. Although apt treatment is not ascertained only from automated diagnosis of X-ray images, the initial screening of cases can be advantageous, for example in the appropriate application of quarantine measures for positive samples. This is feasible until comprehensive examinations and specific treatments can be implemented. Our approach would be helpful in the areas where testing kits are unavailable or in short supply. Higher quality corpora of COVID X-ray image data would be useful to develop more adequate models for faster diagnosis. An important task could be distinguishing patients with mild COVID symptoms versus pneumonia symptoms.

The world has witnessed a total or partial lockdown due to the COVID pandemic. Efforts are geared towards multifaceted solutions to combat this pandemic. These leverage advanced research, e.g. in epidemiology [40] and contact tracing apps [17]. AI techniques and radiological imaging could propel accurate diagnostics and counterbalance the problem of low doctor to patient ratios, especially specialized physicians in areas such as remote villages as well as overpopulated regions, e.g. metropolitan cities. ICT (Information and Communications Technology) is deployable in such places to offer better healthcare, e.g. [19].

In line with this, our proposed approach herewith can aid diagnosis of COVID and pneumonia by using chest X-rays for training via transfer learning. However, note that it only provides decision support. Hence, there would still be the need for radiology specialists with their time and effort to manually examine reports. Compared to the number of patient cases, there is limited availability of doctors whose time is precious. Thus, the use of robots can be considered here. Currently, robots are useful for simple medical procedures such as taking temperatures of patients, giving them medical pills and monitoring their health [54]. An example of such a robot appears in Fig. 20 from a publicly available source [43]. Likewise, robots can be trained to conduct medical diagnosis by using the proposed approach in this research and the related literature.

Training robots to perform / assist diagnosis is a challenging task. This could potentially involve humans and robots working together in medical decision support for diagnosis. It would entail several challenges, e.g. reasoning by using commonsense knowledge in suitable applications, making robots navigate paths well in hospitals and other healthcare settings, and considering ethical issues related to robots. Hence, some of this work could benefit from our own related research on advances in commonsense knowledge based human-robot collaboration (HRC) in particular [7–9], specific applications and discussions on robotics [29, 30] as well as the general usefulness of commonsense knowledge in various AI systems [34, 42]. We can therefore conduct detailed investigations on these aspects as topics of future work, thriving upon our expertise in these areas, as evident from the aforementioned publications. Further work on the use of robotics in the diagnosis and/or treatment of COVID would involve substantial work with significant challenges.

Another interesting venture could be the development of mobile application (apps) based on results of this study and related work. Such apps could pertain to symptomatic information, preventive measures etc. and could be orthogonal to other COVID related apps, e.g. contact tracing [25], telemedicine [24], food donation [48], small business recovery [46], health essentials shopping [23], question-based symptom checking [52], and so on. Many of these encompass multimedia data in heterogeneous formats and are helpful in various aspects pertinent to this pandemic and its aftermath. A partial snapshot of such an app, recently developed in our university, is presented in Fig. 21 herewith. This app serves as a recommender on health essentials useful in COVID and beyond, based on analysis of product data including user reviews, product prices and other criteria. Likewise, numerous apps are providing various types of support in terms of dealing with the COVID global pandemic and its recovery phases, thereby also contributing to smart health and smart living. Some of this work receives attention on social media as well as surveyed in a recent article [33] that cites the works of numerous researchers. Accordingly, building more apps with reference to issues pertaining to COVID and related facets, based on some contributions of this study, would make useful impacts on AI and multimedia technology in

Fig. 20 COVID threatens to overwhelm the U.S. health system creating a need for remote services [43]
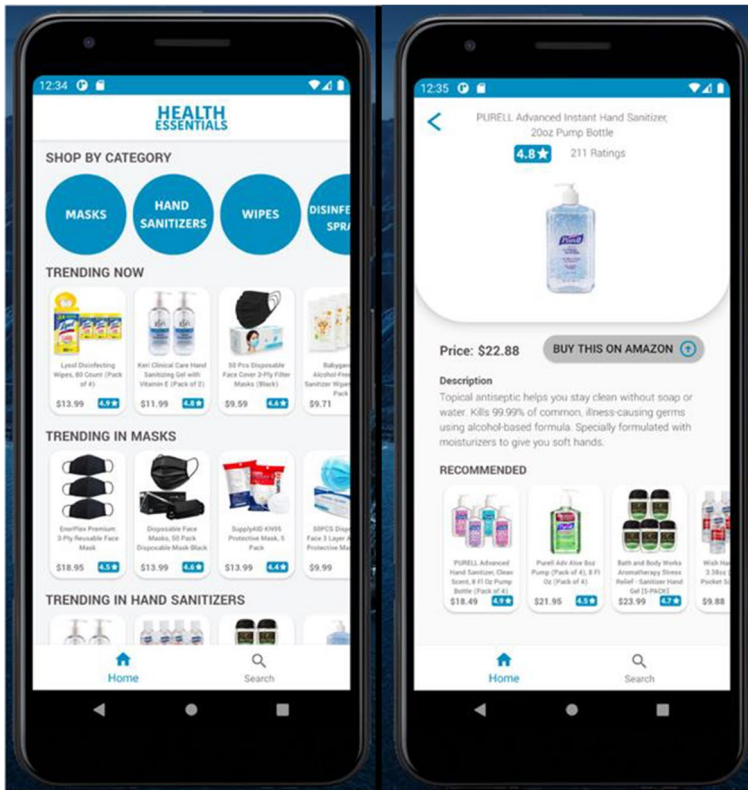
Fig. 21 Health essentials app useful in COVID and beyond [23]

healthcare. Some of this app development can be future work in our own research groups. This can make broader impacts on smart health [3, 45, 47], analogous to the work of various researchers.

6 Conclusions and roadmap

The research in this article focuses on proposing cost-effective, rapid, automated diagnosis of COVID via an AI-based solution deployed over free publicly available image data. We propose a transfer learning based approach using computer vision models to classify chest X-ray images as COVID / pneumonia / normal healthy cases. We harness computer vision models of VGG16, VGG19, and ResNet101 in conjunction with data augmentation techniques to conduct transfer learning, aiming to minimize the number of images as well as training time in the learning process. We achieve impressive results as evident from the observations of accuracy, loss and confusion matrix over unseen test data.

Among the computer models investigated in our study, the VGG19 model, largely, yields the highest accuracy. We can provide some justification for this as follows. The images in this work certainly entail significant nuances and the VGG architecture is

good for learning nuances in the data. This is because the fundamental topology of VGG has multiple 3×3 kernel-sized filters designed such that these small filters lined in a sequence can simulate the effect of large filters. The addition of more layers in VGG19 as compared to VGG16 offers increased depth, which contributes positively to the learning process. In fact, one of the goals of our study was to investigate whether the addition of more layers in VGG19 can have a positive impact on the learning, and our experiments have revealed that it is indeed so. While we considered the use of ResNet101, and obtained interesting results with satisfactory accuracy, the VGG models outperformed ResNet, in our experimental evaluation. We can possibly reason this out considering the fact that the skip-connections approach of ResNet may not always be ideal for learning with high accuracy when there are too many subtleties in the data, as is the case with chest X-ray images (this is different than distinguishing one model of a car from another). Yet it is commendable that we obtain reasonable accuracy with ResNet as well, especially with a greater number of samples, and in that sense, ResNet serves as a benchmark for comparing the VGG models, in the completion of this entire study.

Overall, this study presents the following highlights.

- VGG16 and VGG19 models provide better performance compared to ResNet101.
- The best-case accuracy obtained with this data for VGG16 is 98% for 350 samples.
- The best-case accuracy achieved here by VGG19 is 99% for 100 samples.
- ResNet101 depicts that accuracy increases with number of training samples, e.g. 83% with 500 samples here.
- An interesting observation is that VGG16 and VGG19 offer excellent results with the smaller dataset as well, i.e. with only 50 samples.
- Adding more data samples for experiments with VGG16 and VGG19 only marginally increases accuracy; hence, for practical purposes, as few as 50 samples can be useful for effective learning, thereby presenting a potential best fit.

Our proposed methodology is suitable in AI-based medical diagnosis of COVID, our notable contribution being finding a best fit between the learning time and prediction accuracy. A limitation of this study pertains to the metrics used in the evaluation of the proposed approach. As of now, we utilize only the standard metrics of accuracy, loss and confusion matrix. Another limitation possibly deals with the type of data used in this study: the samples indicate presence or absence of COVID but we have not used data on different mutations of COVID that perhaps may be seen more recently. These limitations, in addition to other aspects of study, can call for further research.

On a final note, we state that research in this paper is potentially useful in decision support for COVID diagnosis and yields interesting applications, providing the scope for future work. We hereby list some future issues on our roadmap.

- Learning with additional data on different types COVID cases, including more recent mutations such as Omicron
- Developing more metrics for evaluation besides the standard metrics of accuracy, loss etc., considering domain-specific issues from bioinformatics, medicine and healthcare
- Implementing applications based on this work within robotics and app development, thriving on our own expertise in these areas and recent work conducted by our research teams

- Executing more detailed comparative studies as needed, especially focusing on more recent variations of the virus
- Investigating other computer vision models in addition to VGG16, VGG19 and ResNet101, based upon existing success stories from the state-of-the-art

This entire work fits into the realms of artificial intelligence in medicine, as well as multimedia data analysis in domain-specific applications. It would appeal to researchers in AI, data science, health informatics and related domains.

Data availability The data used in this study is publicly available on ImageNet.

Code availability The code used in this work is written in Python. This program code is available on GitHub. The code or its relevant parts can be provided to some interested users upon request.

Declarations

Conflicts of interest/Competing interests The authors declare that they have no conflicts of interest / competing interests to the best of their knowledge.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent to publish The authors give the consent to publish this journal article.

References

1. Almezhghwi K, Serte S, Al-Turjman F (2021) Convolutional neural networks for the classification of chest X-rays in the IoT era. Springer's MTAP (Multimedia Tools and Applications) 80(19):29051-29065
2. Apostolopoulos I, Mpesiana T (2020) Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. Phys Eng Sci Med 43(2):635–640
3. Baig MM, Gholamhosseini H (2013) Smart health monitoring systems: an overview of design and modeling. J Med Syst 37(2):1–14
4. Bhatt D, Patel C, Talsania H, Patel J, Vaghela R, Pandya S, Modi K, Ghayvat H (2021) CNN variants for computer vision: history, architecture, application, challenges and future scope. Electronics 10(20):2470
5. Chen Y, Hong WC, Shen W, Huang N (2016) Electric load forecasting based on a least squares support vector machine with fuzzy time series and global harmony search algorithm. Energies 9(2):70. https://doi.org/10.3390/en9020070

6. CNN Architectures: VGG, ResNet, Inception + TL, https://www.kaggle.com/shivamb/cnn-archi tectures-vgg-resnet-inception-tl

7. Conti CJ, Varde A, Wang W (2020) Robot action planning by commonsense knowledge in human-robot collaborative tasks. IEEE IEMTRONICS:170–176

8. Conti CJ, Varde AS, Wang W (2020) Task quality optimization in collaborative robotics. 2020 IEEE international conference on big data, pp 5652–5654. https://doi.org/10.1109/BigData50022.2020.9378498

9. Conti CJ, Varde AS, Wang W Human-robot collaboration with commonsense reasoning in Smart manufacturing contexts. IEEE Trans Autom Sci Eng (TASE). https://doi.org/10.1109/TASE.2022.3159595

10. Du X, Emebo O, Varde A, Tandon N, Nag Chowdhury S, Weikum G (2016) Air quality assessment from social media and structured data: pollutants and health impacts in urban planning. In: IEEE international conference on data engineering (ICDE), workshop on health data management and mining (HDMM), pp 54–59

11. Fan GF, Qing S, Wang H, Hong WC, Li H-J (2013) Support vector regression model based on empirical mode decomposition and auto regression for electric load forecasting. Energies 6(4):1887–1901. https://doi.org/10.3390/en6041887

12. Ghoshal B, Tucker A (2020) Estimating uncertainty and interpretability in deep learning for coronavirus (Covid19) detection. arXiv:2003.10769

13. Hawkins D (2004) The problem of overfitting. J Chem Inf Comput Sci 44(1):1–12

14. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. 2016 IEEE CVPR conference (computer vision and pattern recognition), Las Vegas, NV, pp 770–778

15. Hemdan E, Shouman M, Karar M (2020) COVIDX-Net: a framework of deep learning classifiers to diagnose COVID-19 in X-Ray images. arXiv:2003.11055

16. Huh JH, Seo YS (2019) Understanding edge computing: engineering evolution with artificial intelligence. IEEE Access 7:164229–164245

17. Jhunjhunwala A (2020) Role of telecom network to manage COVID-19 in India: Aarogya Setu. Trans Indian Natl Acad Eng 1-5. PMC7264964

18. Keras Optimizers, https://keras.io/api/optimizers/

19. Khan NUZ, Rasheed S, Sharmin T, Ahmed T, Mahmood SS, Khatun F, Hanifi S, Hoque S, Iqbal M, Bhuiya A (2015) Experience of using mHealth to link village doctors with physicians: lessons from Chakaria, Bangladesh. J Med Inform Decis Mak 15:62. https://doi.org/10.1186/s12911-015-0188-9

20. Kingma D, Ba J (2014) Adam: A method for stochastic optimization. arXiv:1412.6980

21. Kuchana M, Srivastava A, Das R, Mathew J, Mishra A, Khatter K (2021) AI aiding in diagnosing, tracking recovery of COVID-19 using deep learning on chest CT scans. Springer's MTAP (Multimedia Tools and Applications) 80(6):9161-9175

22. Li MW, Wang YT, Geng J, Hong WC (2021) Chaos cloud quantum bat hybrid optimization algorithm. Nonlinear Dyn 103. https://doi.org/10.1007/s11071-020-06111-6

23. Lourenco J, Varde A, Zhou H The Health Essentials App. https://doi.org/10.13140/RG.2.2.29642.49604

24. MD Live Mobile App 4.0 https://www.mdlive.com/mobileapp/

25. NY Department of Health, "Covid Alert NY" (2020) https://coronavirus.health.ny.gov/covid-alert-ny-what-you-need-know

26. Pan SJ, Yang Q (2010) A survey on transfer learning. IEEE TKDE Trans Knowl Data Eng 22(10):1345–1359

27. Park SW, Huh JH, Kim JC (2020) BEGAN v3: avoiding mode collapse in GANs using variational inference. Electronics 9(4):688

28. Park SW, Ko JS, Huh JH, Kim JC (2021) Review on generative adversarial networks: focusing on computer vision and its applications. Electronics 10(10):1216

29. Paulino L, Hannum C, Varde AS, Conti CJ (2021) Search Methods in Motion Planning for Mobile Robots. Intelligent Systems Conference (IntelliSys). Springer Publishers, pp 802–822

30. Persaud P, Varde A, Wang W Can robots get some human rights? A cross-disciplinary discussion. J Robot, Hindawi Publishers, Volume 2021, Article ID 5461703. https://doi.org/10.1155/2021/5461703

31. Pham T (2020) Deep learning of COVID-19 chest x-rays: new models or fine tuning? Health Inf Sci Syst 1–14

32. Pham TD (2021) Classification of COVID-19 chest X-rays with deep learning: new models or fine tuning. Health Inf Sci Syst 9(1):1–11

33. Puri M, Dau Z, Varde A COVID and social media: analysis of COVID-19 and social media trends for smart living and healthcare. ACM SIGWEB, autumn 2021, Article No: 5, pp 1–20. https://doi.org/10.1145/3494825.3494830

34. Razniewski S, Tandon N, Varde A (2021) Information to wisdom: commonsense knowledge extraction and compilation. ACM WSDM (International Conference on Web Search and Data Mining), pp 1143–1146. 10.1145/3437963.3441664

35. Saha I, Gourisaria MK, Harshvardhan GM (2021) Distinguishing pneumonia and COVID-19: utilizing computer vision to mimic clinician efficacy. In: IEEE international conference on artificial intelligence and smart systems (ICAIS), pp 834–841
36. Sethy P, Behera S (2020) Detection of coronavirus disease (COVID-19) based on deep features. arXiv preprint 2020, arXiv: 2020030300
37. Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. J Big Data 6(1):1–48
38. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. ICLR
39. Sinha R, Pandey R, Patnaik R (2018) Deep Learning for computer vision tasks: a review. arXiv preprint, arXiv:1804.03928
40. Sun J, He W, Wang L, Lai A, Ji X, Zhai X, Li G, Suchard MA, Tian J, Zhou J, Vert M, Su S (2020) Covid-19 epidemiology, evolution and cross-disciplinary perspectives. Trends Mol Med 26(5):483–495
41. Tancer J, Varde A (2011) The deployment of MML for data analytics over the cloud. IEEE ICDM (International conference on data mining) workshops, pp 188-195
42. Tandon N, Varde A, de Melo G (2017) Commonsense knowledge in machine intelligence. ACM SIGMOD Record 46, No Dec 2017, pp 49–52. https://doi.org/10.1145/3186549.3186562
43. The Triple-I Blog, https://www.iii.org/insuranceindustryblog/covid-19-spurs-jobs-for-robots-drones-other-technologies/
44. The World Shuts Down (2020) https://www.dailymail.co.uk/news/article-8150501/Map-shows-worlds-population-coronavirus-lockdown.html
45. Tian S, Yang W, Le Grange JM, Wang P, Huang W, Ye Z (2019) Smart healthcare: making medical care more intelligent. Glob Health J 1(3):62–65
46. Torres J, Anu V, Varde A, Duran C (2021) My-Covid-safe-town: a mobile application to support post-Covid recovery of small local businesses. IEEE IEMTRONICS:1–7. https://doi.org/10.1109/IEMTRONICS52119.2021.9422594
47. Varde A, Pandey A, Du X (2022) Prediction tool on fine particle pollutants and air quality for environmental engineering. Springer Nature Computer Science. 3(184). https://doi.org/10.1007/s42979-022-01068-2
48. Varghese C, Pathak D, Varde A (2021) SeVa: a food donation app for smart living. IEEE CCWC, pp 408-413
49. VGG16, VGG19, https://towardsdatascience.com/extract-features-visualize-filters-and-feature-maps-in-vgg16-and-vgg19-cnn-models-d2da6333edd0
50. VGG19, www.mathworks.com/help/deeplearning/ref/vgg19.html
51. Wang P, Chen P, Yuan Y, Liu D, Huang Z, Hou X, Garrison C (2017) Understanding convolution for semantic segmentation, arXiv:1702.08502
52. WebMD, COVID-19 Symptom Checker, https://www.webmd.com/coronavirus/coronavirus-assessment/default.htm
53. Weiss K, Khoshgoftaar T, Wang D (2016) A survey of transfer learning. J Big Data 3(1):9
54. Yang GZ, Nelson BJ, Murphy RR, Choset H, Christensen H, Collins SH, Dario P, Goldberg K, Ikuta K, Jacobstein N, Kragic D, Taylor RH, McNutt M (2020) Combating COVID-19—the role of robotics in managing public health and infectious diseases. Sci Robot 5(40). https://doi.org/10.1126/scirobotics.abb5589
55. Zhang J, Xie Y, Liao Z, Pang G, Verjans J, Li W, Sun Z, He J, Li Y, Shen C, Xia Y (2020) COVID-19 screening on chest X-ray images using confidence aware anomaly detection. arXiv:2003.12338