International Journal on
"Technical and Physical Problems of Engineering"
(IJTPE)
Published by International Organization of IOTPE

ISSN 2077-3528
IJTPE Journal
www.iotpe.com
ijtpe@iotpe.com

June 2022 | Issue 51 | Volume 14 | Number 2 | Pages 190-199

LUNGCOV: A DIAGNOSTIC FRAMEWORK USING MACHINE LEARNING AND IMAGING MODALITY

S. Kumar [1, 2]   H. Kumar [1]

1. Department of Computer Engineering, J.C. Bose University of Science and Technology, Faridabad, Haryana, India
sunilymca1@jcboseust.ac.in, htanwar@jcboseust.ac.in
2. Department of Computer Science Engineering, School of Engineering and Technology, CSJM University, Kanpur, India

Abstract- The COVID-19 is throwing the entire globe into a terrible predicament, bringing living things to a terrifying standstill all across the planet and taking millions of human lives. As we have seen, there is a breakdown in the healthcare sector during pandemics since we are unable to meet all the required demands. This circumstance mandates the automation of the procedure for diagnosing COVID-19 with imaging techniques. Modality of imaging Chest X-rays/CT scans is less expensive and speedier tests that may aid in the timely identification of COVID-19 patients suffering. With AI's rapid growth in popularity in the healthcare industry, it can assist in boosting productivity and efficiency. As a result, AI-based image processing and diagnosis procedures are pursued. The proposed LungCov methodology helps in understanding the diagnostic technique for COVID-19 from an input image to an artificial intelligence-based processed image. AI provides the optimized image which can be further used to classify the image as a COVID-19 positive patient or a healthy one. One technology that might aid in this endeavor is machine learning. The main objective is to improve the healthcare system's care and service to society.

Keywords: Corona Virus, Deep Learning, Image Processing, X-ray, CT Scan.

1. INTRODUCTION

Lung diseases affect people of all ages, and they can be caused by a variety of factors. The novel Corona Virus Disease 2019 (COVID-19), also known as SARS-CoV-2, and named the 2019 novel Coronavirus (2019-nCoV) is a transmissible lung infection that transfers to humans in exponential form. COVID-19 began to spread in December of 2019 in China, quickly spread to India in March of 2020, and then to the rest of the world as an epidemic. COVID-19 has proliferated in around all over countries throughout the planet, with a substantial number of afflicted individuals and fatalities. The pandemic has affected most countries in the world to date. The world economy and healthcare are suffering as a result of this. As per WHO's report stated till 18 February 2022, the global COVID-19 positive cases are 418650474 and the total deaths are 5856224. The USA, India, Brazil, Russia, France, Turkey, UK, Spain, Germany and Italy, are among the top ten countries most seriously affected by COVID-19 with positive cases represented in Figure 1 [3].

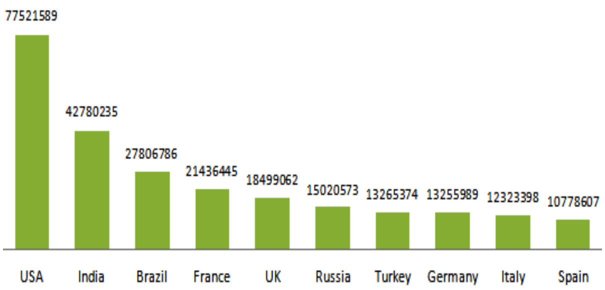

Figure 1. Recent cases of COVID-19 infection in the top ten countries [3]

The COVID-19 is the deadliest illness that affects the majority of countries. The top ten nations most affected and their mortality rate is presented through Figure 2 [3].

In humans, the lungs are the primary organs of the respiratory system, expanding and contracting hundreds of times every day to take in oxygen. As part of the air that we breathe, oxygen enters the lungs. It travels to the blood arteries deep within our lungs before spreading to the rest of our body. When we consume oxygen, our bodies produce carbon dioxide as a waste product. When we exhale, we expel carbon dioxide [5].

Lung illnesses are a leading cause of mortality worldwide. People suffering from lung illness typically have trouble breathing. Lung disease can arise when the lungs experience any sort of issue. Because of the wide range of lung disorders, it is important to use image processing to identify lung diseases before they progress from acute to chronic. The diagnosis and prognosis of lung illnesses such as COVID-19 is a challenging task for clinicians, which is why lung disease detection or diagnosis is an essential problem.

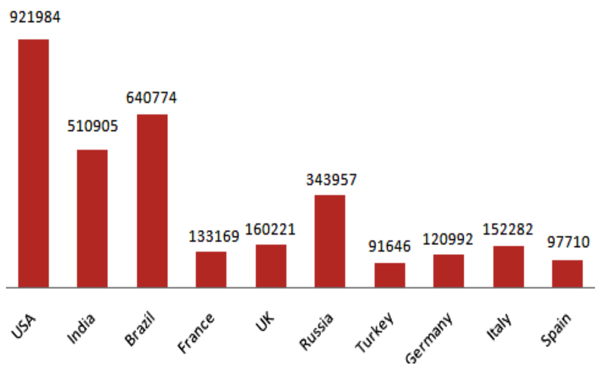

Figure 2. Recent cases of COVID-19 deaths in the top ten countries [3]

Finding efficient solutions to COVID-19-related problems is a prominent research area. Artificial Intelligence (AI) and/or Machine Learning (ML) and/or Deep Learning (DL) already garnered increased research efforts toward tackling difficult challenges in a variety of areas, including healthcare, psychology, finance, and engineering [6].

It may also be useful in studies aimed at improving the accuracy of COVID-19 diagnosis and prediction. Clinical trials, optimizations, and improved diagnostic procedures, like medical imaging and image processing techniques, might assist AI. AI has the potential to assist medical approaches; it has not been widely used and well-adapted to help healthcare systems tackle COVID-19. A field that might benefit from AI's beneficial contribution is an image-based clinical diagnosis, which can provide a quick and accurate diagnosis of COVID-19, perhaps saving a patient's life. On the other hand, the growing amount of data is a problem that investigators and decision-makers must cope with [2].

2. RELATED WORKS

The majority of ordinary people have a poor understanding of disease/illness diagnosis and prognosis; researchers from many fields of machine learning and the healthcare industry have investigated the COVID-19 identification and prediction. This section discusses a variety of different techniques and/or approaches used in existing COVID-19 diagnosis coupling machine learning and imaging modalities (Table 1).

Ophir Gozes, et al., concentrated on a deep learning model that takes CT images as input with insufficient data and looks for cases of COVID-19. When a case is categorized as positive, the system generates a score, which is utilized as a corona score in the research and indicates the severity of the condition. They proposed a system in which they worked on per slice analysis of the CT scan for each patient through a sequence of 2D processing. To do so, they did it through some steps in step one. To acquire the lungs' 2D ROI segmentation phase was conducted. In step two, to evaluate individual segments as healthy or abnormal, a 2D ROI classifier is used. In step three, to build a localization map, a multi-scale model Grad-Cam is utilized. In the fourth step, the localization maps from all segments were then combined to form a 3D concatenated volume.

In step five, 3D volumetric scoring is introduced which is done by the Corona-score bio-marker. In step six, the score tells the illness severity grade [14].

Table 1. Overview of the literature for COVID-19 diagnosis using machine learning

Author	Imaging Modality	Classification Method	Classified Diseases	Performance Metrics
Ophir Gozes, et al. [14]	CT scan	U-net VGG-16 ResNet-50	COVID-19	AUC - 94.80%
Vishu Madaan, et al. [15]	X-Ray	CNN ReLu Adam	COVID-19	Accuracy - 98.44%
Sakshi Ahuja, et al. [16]	CT scan	Transfer Learning ResNet18 ResNet50 ResNet101 Squeeze Net	COVID-19	Accuracy - 99.4%
Ayturk Keles, et al. [17]	X-Ray	COV19-ResNet COV19-CNNet	COVID-19 Viral Pneumonia Normal	Accuracy - 97.61% Accuracy - 94.28%
Elene Firmeza Ohata, et al. [18]	X-Ray	Transfer learning Mobile Net with SVM	COVID-19	Accuracy - 98.62%, F1-score - 98.461%
Pedro Silva, et al. [19]	CT scan	Efficient-Covid Net	COVID-19	Accuracy - 98.99%
M. B. Jamshidi, et al. [2]	X-Ray CT scan MRI PET etc.	RNN LSTM GAN ELM	COVID-19	---
S. Tao, et al. [4]	X-Ray CT scan SSMI Histopatho-logy	CNN DBN Transfer Learning Ensemble	TB Lung Cancer Pneumonia COVID-19	---

Vishu Madaan, et al., developed a two-phase X-ray image classification named XCOVNet for COVID-19 identification using CNN. Preprocessing the X-ray image dataset, half of which are COVID-19 positive and the other half normal, this job was completed during the first phase. The neural network model is trained in the second phase which is fine-tuned to achieve a classification accuracy of 98.44 percent. They made use of two chest X-ray imaging collections in their research: Dataset-1 comprises 950 X-ray images annotated with over fifteen distinct categories of illness findings; and Dataset-2, which contains 5856 X-ray images are grouped into three categories: bacterial pneumonia, viral, and normal pneumonia. In their study, prior to and after the Max Pooling layer, they employed ReLu as an activation function with the Adam optimizer [15].

Sakshi Ahuja, et al., devised a framework that recognized COVID-19 infection through transfer learning using CT scan images. They proposed a stage-based detection method through the following strategies to improve accuracy: The first phase involved data augmentation, the second stage involved a pre-trained CNN model, and the third stage involved CT scan image anomaly localization [16].

Ayturk Keles, et al., constructed an AI-based inference engine that can transform X-ray equipment into helpful testing apparatus by identifying COVID-19 with deep learning technologies. Unlike prior research in the area, inference engines were created from the roots up, employing fresh and unique deep neural networks with no or previous systems. COV19-CNNet and COV19-ResNet are the names of the two engines. The COV19-CNNet uses convolutional neural network (CNN) architecture, while the COV19-ResNet uses the architecture of ResNet. They focused their research on the complexity of classifying COVID-19 into multiple categories [17].

Elene Firmeza Ohata, et al., used multiple CNN architectures for use as feature extractors for X-ray images by training them on ImageNet. Machine learning techniques such as MLP, KNN, and Naïve Bayes etc. are used to combine with CNNs. They used a linear kernel SVM classifier and the Mobile Net architecture to identify COVID-19 [18].

Pedro Silva, et al., suggested a voting-based technique for screening COVID-19. In this method, images are categorized into different groups in a system of voting. There is a cross-dataset assessment offered to assess the models' resiliency using information from various distributions [19].

M. B. Jamshidi, et al., studied trending deep learning approaches for COVID-19 evaluation and prediction. The researchers used RNN, GAN, LSTM, and Extreme Learning Machine. They conceptualized the realm of AI-based methods, frameworks, and platforms that are suited for dealing with COVID-19 problems are being researched [2].

S. Tao, et al., surveyed the diagnostics of respiratory diseases through deep learning. They mainly used 4 types of image types: X-ray, CT scan, histopathology images and Sputum Smear Microscopy Images (SSMI). Researchers surveyed mainly 4 lung diseases: TB, Lung Cancer, Pneumonia, and COVID-19. They found seven attributes that are common in their analyzed articles. The following features are imaging modality, parameters, data augmentation, types of deep learning algorithm, transfer learning, classifiers' ensemble, and lung diseases classification. The study has some issues, which include data unbalancing due to bias, handling of huge image sizes due to time consumption in processing, limited publicly available datasets, and when employing ensemble methods, there is a high relationship of inaccuracies [4].

3. A UNIFIED VIEWPOINT TOWARDS COVID-19 DIAGNOSIS EMPLOYING MACHINE LEARNING AND IMAGING MODALITY

A simplified strategy is used for the diagnosis of COVID-19 by applying machine learning and image modalities. The identification of COVID-19 is influenced by three major aspects. They are COVID-19, imaging modality and machine learning.

3.1. COVID-19

The COVID-19 is a newly identified lung infection that has been tagged as a global epidemic. COVID-19 positive patients must be identified as soon as feasible to prevent the virus from spreading further. The COVID-19 is generally determined based on symptoms such as fatigue, coughing, fever, and sneezing if the individual has been travelled to a highly contagious location, aged and whether the individual has a background of illness or illnesses that might enhance the likelihood of infection by the COVID-19 virus [23].

The most recent and cutting-edge method COVID-19 is diagnosed via an RT-PCR (real-time reverse transcript polymerase chain reaction) test. Due to the restricted access to diagnostic kits and the rapidly growing number of positive cases, X-ray or CT scans provide an alternative low-cost and quick screening option for patients [1].

COVID-19 is an illness that may be classified into three groups:

- Group 1: This group is for mild cases. It is a form of COVID-19 that is early or asymptomatic and has symptoms such as cough, fever, and headache.
- Group 2: This group is for moderate cases. The patient experiences some shortness of breath and pulmonary problems, such as hypoxia, in this category.
- Group 3: This group is for advanced cases. The patient has hypoxia and shock of this kind. This kind is responsible for the majority of life-threatening situations.

3.2. Imaging Modality

Image-based medical diagnostics, for example, can benefit from AI's valuable input since it allows for quick and accurate identification of COVID-19, potentially saving lives [2]. Not only is AI used for imaging, but it is also used for disease detection via the auscultation approach [27]. X-ray diagnostic technology is widely available around the world and is capable of accurately diagnosing respiratory ailments in individuals. X-rays can be utilized as a manual diagnosis procedure for COVID-19 infection, although this approach takes longer to complete. This process demands the use of efficient and skilled doctors and medical workers [1].

In the assessment of COVID-19, CT scan imaging can also be used as a substitute to the RT–PCR testing whereas X-ray radiography is a little more effective at identifying COVID-19; it is ineffective at detecting the illness in its early stages [17].

The COVID-CT dataset [22] comprises COVID-19 CT scans of infected individuals. SARS-CoV-2 was discovered in 1252 CT scans and was not found in 1230 CT scans, for a total of 2482 scans (Figure 3). The image dataset obtained from the sources contains challenges such as non-standard input image size, non-standard quality, and written textual information about the images [19, 20].

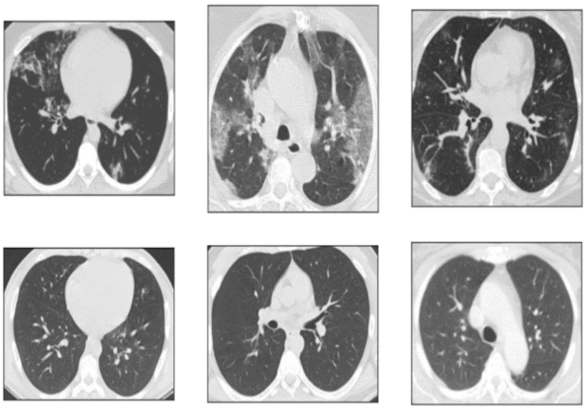

Figure 3. CT scans that are positive for COVID-19 (at the top of the image) and normal CT scans (at the bottom of the image) [22]

3.3. Machine Learning

Manual techniques based on the indications and symptoms examined by doctors have been used to diagnose human diseases since ancient times. A decision (output) can be made based on the disease's signs and symptoms (features). However, as time passes and technology advances, technology assists humans in many interdisciplinary domains. As a result, medical and healthcare industries are increasingly turning to AI, ML, and DL to care and service. ML, in contrast to manual strategies, provides an automated framework for identifying and diagnosing illnesses such as lung disorders as well as COVID-19 [5]. The process of obtaining machine learning-based approaches is to improve care and diagnostics [2].

3.3.1. Advantageous Factors of Machine Learning in Healthcare

Data science, data analytics, the healthcare industry, and research and development all benefit from machine learning [9].

The following are the benefits of ML in healthcare:

a) ML and its algorithms assisting in the diagnosis and prognosis of the disease.

b) Predictive modeling using computational image analysis.

c) Machine learning uses prior experiences, such as medical data, to evaluate data and form conclusions.

d) Drug development and discovery.

3.3.2. Approaches of Machine Learning

Machine learning technologies are used in healthcare applications to give an effective prediction or decision support. There are various Machine Learning methods and their correlations in healthcare as illustrated in Figure 4.

3.3.2.1. Supervised Learning Approach

The Algorithms process on labeled input data with the intended output in the supervised learning method. Machine learning insights are a data-driven approach that helps to enhance the diagnostic result [8]. Classification and regression are the two types of supervised ML.

In classification, the computed result is categorized as a class or category. For example, with a particular CT scan image, examining for the presence or absence of contagion. If COVID-19 contagion is found, the patient is declared as COVID-19 positive. The computed result in regression is expressed as a real or continuous value. For example, the patient's body temperature indicates whether or not he or she has a fever. As a result, it follows that ML in healthcare aids in illness classification or prediction [11].

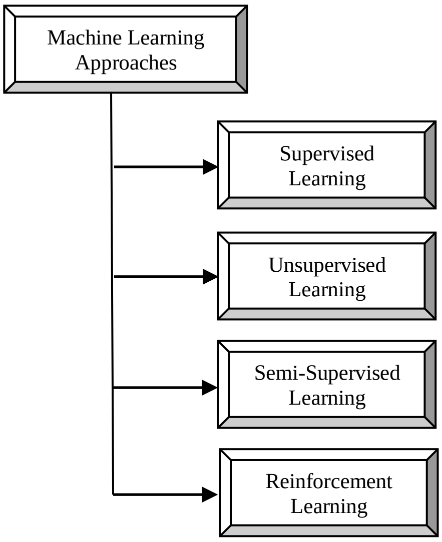

Figure 4. Approaches of machine learning

3.3.2.2. Unsupervised Learning Approach

The learning algorithms are given no labels and are left to identify the structure or hidden patterns in their data on their own. Unsupervised learning seeks out feature learning in the set of given data. This approach tries to build groups or clusters based on data similarities. Unsupervised learning approach includes clustering techniques such as K-Means, association rules mining, and classification in a hierarchical manner etc.

The following are the applications of unsupervised learning in the medical industry:

a) Disease' clustering based on signs.

b) Discovering sufferer with comparable features that help in diagnosis.

c) Medical records pattern recognition.

d) Imaging modality examination, such as X-rays and CT scans [12].

3.3.2.3. Semi-Supervised Learning Approach

Supervised and unsupervised learning can be combined to form semi-supervised learning. Labeled and unlabeled data are both utilized in semi-supervised learning. Labeling data necessitates the use of a human professional or a physical experiment, whereas unlabeled data is often widely available. Due to the massive volumes of medical data demanded investigations, semi-supervised learning can be effective in the healthcare system.

3.3.2.4. Reinforcement Learning Approach

A software agent executes a task in a dynamic environment while using reinforcement learning. The agent aims to obtain the maximum cumulative reward by considering the current condition and inputs from the environment. Agent acquires potential for understanding and discovering the best way to maximize the total reward by doing so. The agent gains knowledge from the whole process, which is essentially a trial-and-error approach that introduces adaptive learning [11]. In general, we follow a set of steps in medical treatment to make the right judgments to get the best outcomes. Reinforcement learning is also an excellent option for healthcare problems in which the majority of decisions are made in a particular sequence [13].

4. PROPOSED LUNGCOV DIAGNOSTICS FRAMEWORK

The proposed LungCov COVID-19 diagnostic methodology is presented (Figure 5), which is based on modalities of imaging like X-Ray and CT scan. The methodology may be able to find whether the given imaging technique corresponds to a healthy or COVID-19 positive patient. The diagnosis of COVID-19 depends on the following steps.

4.1. Lung Diseases

The first step to identifying COVID-19 is to differentiate it from all other lung diseases because every lung disease has its characteristics. The lung diseases have some common symptoms like dry cough, cough with sputum, dizziness, breathlessness, etc., but COVID-19 has its novelty due to the virus. Doctors play a major role in this step to identify and discriminate against COVID-19 infection.

There is a high possibility that COVID-19 and other lung diseases coexist, which might be difficult to distinguish. As a consequence, all signs, symptoms, and medical record data are forwarded to the next step, which may help in the process of diagnosing COVID-19. The ability to produce an accurate COVID-19 diagnostic on time is important for effective treatment in the medical field.

4.2. Imaging Modality

COVID-19 medical record collecting is dependent on the healthcare industry and its authentication. This is a crucial step that involves selecting an appropriate imaging modality and processing it using machine learning techniques.

4.2.1. Identification of Images Dataset

The principal obstacle to conducting research is the absence of complete and accurate data sets of good quality. It is simple to interpret and analyze image datasets that are of excellent quality and have correct annotations. A patient database, a global database, or big data are all examples of datasets. The image dataset acts as an input layer to the ML model [2].

Identifying image datasets from the public database is a significant effort. However, private datasets may be developed if anyone wants to work on anything discreetly. The following are some examples of COVID-19 image datasets.

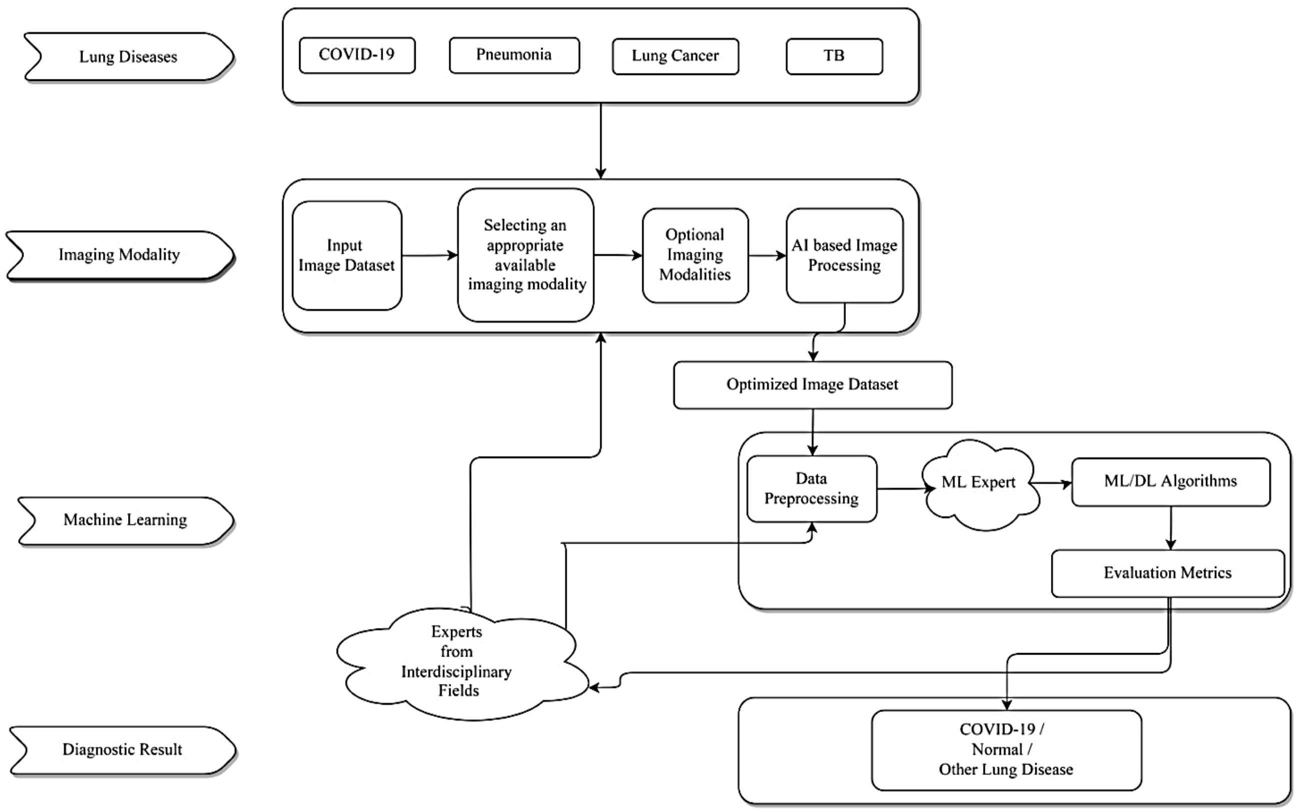

Figure 5. Proposed diagnostics framework for LungCov

Table 2. Publicly accessible image datasets of COVID-19

Image Dataset Name	Imaging Modality	No. of Samples	No. of Patients	Reference
COVID-CT	CT-scan	812	271	[20]
COVID-19 Radiography Database	X-Ray	3616	---	[21]
SARS-CoV-2 CT-scan	CT-scan	2482	120	[22]

As the Table 2 shows, there are several options for various types of imaging techniques from which we must choose one to process and proceed further.

4.2.2. Selecting an Appropriate Available Imaging Modality

To select an appropriate imaging modality, a machine learning-based image selector can be used. The development of an insightful ML or DL based image selector may be finalized, with the primary job of selecting the probable imaging modalities from the system's previous experience. Many imaging approaches may be recommended. The techniques that may be employed in process include X-Ray imaging, computed tomography scan (CT scan), magnetic resonance imaging (MRI), positron emission tomography (PET), and so on [2]. Apart from these imaging modalities, image file formats in the digital imaging and communications in medicine (DICOM) format are used [10].

4.2.3. Optional Imaging Modalities

Every imaging modality has its benefits and specific kinds. Assume we're diagnosing COVID-19 using a collection of X-Ray images; there may be a chance to get better detection from a dataset of CT scans. So, it is a better technique to also work on the available parallel image dataset [2].

4.2.4. AI-Based Image Processing

AI-based image processing approaches reduce process time while also delivering better image processing technology. Artificial intelligence-based image processing can be utilized to determine the optimal and enhance images from the identified image dataset. The images are refined and enhanced using a machine learning approach. A machine learning technology with an effective feature extraction technique can be utilized to help distinguish between COVID-19 and any other lung diseases while deep learning can do this through automated processing. CNN should be the great answer for attaining these targets [28]. The CNN models are being regarded like "black boxes" and much study is being conducted to evaluate and analyze medical imaging. The rationale is that this sort of network is highly competent to represent nonlinear images and is widely used in the processing and diagnosis of medical images [2].

4.2.5. Optimized Image Dataset

An Optimized Image is acquired after the AI-based Image processing stage is completed. To create the dataset of optimized images, each input image is treated separately. The key benefit of completing this step is that we now have a standardized dataset of images, which may aid the system in processing them more efficiently.

4.3. Machine Learning

After getting the optimized images, the main work of diagnosing or classifying COVID-19 comes into focus. An ML or DL can help distinguish between COVID-19 and lung disease when it comes to diagnosis. Deep learning, which facilitates the diagnosis, categorization, and prognosis of COVID-19 using lung imaging, is increasing in tandem with machine learning [2].

4.3.1. Medical Data/Image Preprocessing

Medical data can be created from patient records, documentation, images, and other pieces of information which a machine can interpret. Every image from the identified datasets is, to the greatest extent possible, in the format PNG/JPEG. The dataset's image resolution ranges from low to high pixels for example 224×224 to 4280×4280. All images, although, have been pre-processed using the resize methodology. An input image can be preprocessed by converting it to grayscale, de-noising it with Gaussian blur, median filters, and morphological smoothing in machine learning. The first stage is to prepare the data for the ML model, which is required for data interpretation, data preprocessing [18].

4.3.1.1. Data Interpretation

To make sense of the data, individuals need to comprehend data characteristics and evaluate primary features like the size of the data volume and total numbers of variables etc. are some of the goals of data interpretation.

4.3.1.2. Data Preprocessing

The process of refining and converting raw data is known as data preprocessing. We can also say it is a procedure for reformatting, correcting, and combining data to create enhanced data [2]. Suppose we are working on lung disease and want to segment lung from human body, then computational procedure in ML is followed.

4.3.1.3. Lung Segmentation

Lung segmentation is a task where we have to separate the lungs from all other parts of the human body, like bones, arteries, and all other associated parts. Most of the time, lung segmentation is considered to be a preprocessing step in the diagnosis of lung diseases. The lung segmentation and its structure formation using image processing is a complex procedure since the similarity ratio in the lung region are not the same. The lungs' segmented output can be applied to detect lung disorders such as COVID-19. Using a trained Faster-RCNN, researchers extracted bounding boxes from the lung region into X-ray images [1]. The Region of Interest (ROI), which appears to be a segment carried out to see whether a sufferer is infected with COVID-19, is frequently, calculated using a mathematical model. Texture model, thresholding, and edge-based methods are all examples of lung segmentation algorithms used in machine learning.

A histogram-based threshold estimation approach was presented in which a histogram of the input image was created and evaluated by setting a threshold. The lungs have been extracted from the image after the outer region is detected and discarded according to the threshold. The lung segmentation process included morphological procedures and connected component analysis. The following was the process for producing an automated histogram depending on initial lung segmentation: Create an image histogram from a CT scan slice firstly, and then image thresholding using the histogram image's threshold, complemented image from threshold image, binaries segment, linked components-based refinements, lungs identified, and lungs after noise removal. After the lung segmentation, Regions of Interest (ROIs) were extracted [26].

4.3.1.4. ML Experts

Human influence, as part of machine learning approaches, occurs at this point. ML expert involvement is used as part of machine learning approaches. ML experts, on the other hand, have a unique understanding of ML's capabilities and limits because they are the ones that develop ML systems [2]. ML experts examine and analyze data to identify the best frameworks, relationships, and features. The involvement of ML experts is essential since their expertise and capacities are not present in the ML system's solution. Human involvement is not required for DL methods to work.

4.3.2. Machine Learning/Deep Learning Technologies

This section is at the core of the proposed methodology since it classifies the COVID-19. This section discusses the many methods and technologies used to diagnose COVID-19. ML/DL technologies are being used to infer, assess, and forecast the outcomes of massive datasets. ML/DL techniques were designed to examine disease signs and identify the disease. ML can also be used to identify illnesses relying on imaging modalities. The functionality of ML and DL is different. In the case of machine learning, the functions performed by ML are represented as the following steps:

1. The dataset's division
2. Feature extraction and selection
3. Appropriate ML/DL algorithms must be chosen

4.3.2.1. The Dataset's Division

The ML model is trained using available experienced data, and it may then be used to evaluate fresh data and make predictions using this model.

The following are the steps for the dataset's division:

Step 1: One of the datasets is chosen.

Step 2: Divide the dataset into two subsets for training and testing, such as 70% or 80% for training and 30% or 20% for testing.

Step 3: After the division of the dataset, one portion is reserved for training the machine learning model, while another portion is reserved for testing. The testing dataset is considered as fresh data which will be used in the future for evaluating the effectiveness of the machine learning model, while the training data can be sent to the next phase, which is feature extraction and selection.

Step 4: The preprocessing of the training dataset is an important and unavoidable phase, which was described earlier as data preprocessing.

4.3.2.2. Feature Extraction and Selection

Feature extraction and selection are both important tasks to perform. Extracted features may have some usefulness while some may not. That causes the selection of relevant features. The selected features are supplied to classifiers or ML/DL algorithms. Generally, feature extraction methods used methods such as Gabor, Zernike, Hara lick, Tamura, while CNN and Gray Level Co-occurrence Matrices (GLCM) can be used as feature selection methods. The feature engineering process works in two stages: (1) extracting features from a given image dataset, and (2) selecting features from the extracted features. Improvised Crow Search Algorithm (ICSA), Improvised Grey Wolf Algorithm (IGWA), and Improvised Cuttlefish Algorithm (ICFA) were presented as 3 optimization algorithms. All three algorithms employed attribute selection methods to narrow down a vast number of retrieved attributes to a subset of the best ones. Findings reveal that ICSA, IGWA, and ICFA removed the most insignificant features, accounting for about 71 percent, 52.3 percent, and 40.6 percent of the total retrieved features, respectively. Bio-inspired algorithms, such as genetic algorithms, could be used to select attributes in medical images diagnostics in addition to the suggested methods. [24].

4.3.2.3. Appropriate ML/DL Algorithms Must be Chosen

To understand the diagnostic system of COVID-19 in a better way, here we explore the proposed system of a COVID-19 detection system using machine learning by Deepak Painuli, et al., is presented. They created a database of COVID-19 symptoms, in which they created a rule-based system to process COVID-19 symptoms, which is used as input and can draw some conclusions. COVID-19 symptoms data is considered as a source of raw input data. Afterwards, as part of the data preparation step, feature selection and extraction are carried out [23]. After these phases, the ML algorithms are used for prediction, which may be either predefined or new designed.

Conventional machine learning algorithms like SVM, linear regression, KNN, logistic regression, MLP, decision trees, Xgboost, Gradient Boost, Adaboost, Ensemble and Clustering, as well as the latest machine learning techniques such as transfer learning, can also be used to identify and classify COVID-19. For the classification purpose of lung disease, R. Emhamed, et al., employed some of the popular ML algorithms like KNN, Decision Tree, Random Forest, Xgboost, Gradient Boost, Adaboost, and CNN. The suggested technique employs a CNN that has been used for classification problems and is trained with transformed images via several pre-processing stages [25].

Deep learning methods might be applied in circumstances when ML is challenged by vast or complex data processing. The deep learning technique comprises three phases to diagnose lung diseases: (1) Pre-processing of the image dataset, (2) training of the dataset, and (3) classification

For diagnosis COVID-19, a Chex Net coupled encapsulation and a Dense Net model were developed [1]. RNNs, GANs, ELM, and LSTM are some of the deep learning algorithms used in COVID-19 diagnostics [2].

The procedures and technologies listed above are used to diagnose or categorize COVID-19. In most cases, ML/DL algorithms are utilized to accomplish this task.

4.3.3. Evaluation Metrics

The effectiveness of the learned ML model is assessed using a training dataset, also known as a validation procedure.

Table 3. Confusion matrix for diagnosing COVID-19

		Predicted	
		COVID-19	Normal
Actual	COVID-19	t_p	f_n
	Normal	f_p	t_n

True Positive (t_p) is an abbreviation meaning if a patient is COVID-19 Positive he is detected as COVID-19 Positive. False Positive (f_p) is an abbreviation meaning if a patient is Normal and he is detected as COVID-19 Positive. False Negative (f_n) is an abbreviation meaning if a patient is COVID-19 Positive and he is detected as Normal. True Negative (t_n) is an abbreviation meaning if a patient is Normal, he is detected as Normal [7]. Accuracy, recall, precision, specificity, and f_1 score are only a few of the performance metrics mentioned underneath.

4.3.3.1. Accuracy

This type of performance metric is used for calculating the correct prediction value. The proportion of accurate calculation results to the total available data show the ML performance of the model over unseen data. The accurate classified findings, such as t_p and t_n, are divided by the entire quantity of test cases to get accuracy [29].

$$accuracy = ((t_p + t_n) / (t_p + f_p + f_n + t_n)) \quad (1)$$

4.3.3.2. Precision

Accurate predictive value (t_p) is compared to the actual t_p and f_p values to determine precision. Precision is used to calculate the correct prediction outcomes including all predictions in a positive class [29]. Precision can give answers about how many people we identified as having COVID-19 out of the total number of people who had it.

$$precision = t_p / (t_p + f_p) \quad (2)$$

4.3.3.3. Sensitivity

Sensitivity is another name for recall. On all true predictions, it is used to calculate true positive predictions (t_p) [29]. In actuality, the proportion of people affected by COVID-19 infections out of the total number of people who have been accurately categorized infected with COVID-19. Recall answers: Out of all the persons with lung illness, the accurately predicted people with lung disease.

$$sensitivity = t_p / (t_p + f_n) \quad (3)$$

4.3.3.4. Specificity

It is a metric for determining how many negatives are detected accurately [29]. Specificity refers to the test's capacity to accurately reject healthy individuals who do not have a disease's state, which implies it displays a healthy person to a patient.

$$specificity = t_n / (t_n + f_p) \quad (4)$$

4.3.3.5. f_1-Score

The harmonic mean of precision and recall is used to get the f_1 Score [30]. It is desirable if the model has some type of precision-recall balance.

$$f_1 - score = \frac{2 \times precision \times recall}{precision + recall} \quad (5)$$

4.3.4. Experts from Interdisciplinary Fields

After getting the results from the performance metrics, the obtained results are passed to the imaging modality and machine learning steps via the interdisciplinary experts. Experts from different fields assess the results. As a consequence of the evaluation of the findings, the ML model tuned itself to perform better at diagnosing COVID-19.

4.4. Diagnostic Result

This step is used to verify diagnostic results using stored data and models information. At each step, the processes for detecting COVID-19 gather information or store data. The data is saved and supplied to the input database for optimization and performance purposes, allowing the system to run more efficiently.

5. CONCLUSION

Researchers, scientists, doctors, businessmen, and government officials have all worked hard to find a cure for COVID-19, the world's worst epidemic. They aim to offer an early prognosis using trending technologies that are employed by researchers and scientists.

To combat the COVID-19, a lot of research is being done utilizing artificial intelligence, machine learning, or deep learning. Our study elaborates on how to use machine learning with imaging to diagnose COVID-19. It's also worth noting that the input image data or dataset plays an important function in the diagnosis or classification of COVID-19 because the ML model performs well when given accurate, large, and well-labeled data, but an inefficient dataset leads the ML model to perform badly. This scenario provides a foundation for improving the ML model's performance.

The ML model's functioning is also reliant on information supplied by earlier phases, such as dataset division, data preprocessing, imaging modality (same or different), and the types of algorithms employed. The variance of the ML model's performance is heavily influenced by the fluctuation of each step and phase. The ML model's diagnosis of COVID-19 may not be 100 percent correct, but it does give useful findings in coping with the epidemic.

REFERENCES

[1] K. Mahajan, M. Sharma, L. Vig, R. Khincha, S. Krishnan, A. Niranjan, T. Dash, A. Srinivasan, G. Shroff, "Covid Diagnosis: Deep Diagnosis of COVID-19 Patients Using Chest X-Rays", Thoracic Image Analysis, pp. 61-73, Lima, Peru, October 2020.

[2] M.B. Jamshidi, A. Lalbakhsh, J. Talla, Z. Peroutka, M. Jamshidi, L.L. Spada, M. Mirmozafari, F. Hadjilooei, P. Lalbakhsh, M. Dehghani, A. Sabet, S. Roshani, "Artificial Intelligence and COVID-19: Deep Learning Approaches for Diagnosis and Treatment", IEEE Public Health Emergency COVID-19 Initiative, Vol. 8, pp. 109581-109595, June 2020.

[3] https://covid19.who.int/WHO-COVID-19-global-table-data.csv.

[4] S.T.H. Kieu, A. Bade, M.H.A. Hijazi, H. Kolivand, "A Survey of Deep Learning for Lung Disease Detection on Medical Images: State-of-the-Art, Taxonomy, Issues and Future Directions", Journal of Imaging, Issue 6, No. 12, p. 131, December 2020.

[5] https://pulmonary.pediatrics.med.ufl.edu/centers-programs/asthma-program/normal-lung-function.

[6] I. Sen, M.I. Hossain, M.F.H. Shakib, M.A. Imran, F. A. Faisal, "In Depth Analysis of Lung Disease Prediction Using Machine Learning Algorithms", Machine Learning, Image Processing, Network Security and Data Sciences, MIND 2020, Part II, pp. 226-235, Silchar, India, July 2020.

[7] T. Garg, M. Garg, O.P. Mahela, A.R. Garg, "Convolutional Neural Networks with Transfer Learning for Recognition of COVID-19: A Comparative Study of Different Approaches", AI, MDPI, Vol. 1, No. 4, pp. 586-606, December 2020.

[8] G.A.P. Singh, P.K. Gupta, "Performance Analysis of Various Machine Learning-Based Approaches for Detection and Classification of Lung Cancer in Humans", Neural Computing and Applications, Vol. 31, pp. 6863-6877, October 2019.

[9] S.L. Sharmila, C. Dharuman, P. Venkatesan, "Disease Classification Using Machine Learning Algorithms - A Comparative Study", International Journal of Pure and Applied Mathematics, Vol. 114, No. 6, pp. 1-10, January 2017.

[10] R.N.J. Graham, R.W. Perriss, A.F. Scarsbrook, "DICOM Demystified: A Review of Digital File Formats and their Use in Radiological Practice", Clinical Radiology, Vol. 60, 1133-1140, November 2005.

[11] S. Agrawal, S.K. Jain, "Medical Text and Image Processing: Applications, Issues and Challenges", Machine Learning with Health Care Perspective, Learning and Analytics in Intelligent Systems, Vol. 13, pp. 237-262, March, 2020.

[12] M.L. Littman, "Reinforcement Learning Improves Behavior from Evaluative Feedback", Nature 521, pp. 445-451, May 2015.

[13] O. Gottesman, F. Johansson, M. Komorowski, A. Faisal, D. Sontag, F.D. Velez, L.A. Celi, "Guidelines for Reinforcement Learning in Healthcare", Nature Medicine, Vol. 25, pp. 16-18, 2019.

[14] O. Gozes, M.F. Adar, N. Sagie, A. Kabakovitch, "A Weakly Supervised Deep Learning Framework for COVID-19 CT Detection and Analysis", Thoracic Image Analysis, pp. 84-93, Lima, Peru, October 2020.

[15] V. Madaan, A. Roy, C. Gupta, P. Agrawal, A. Sharma, C. Bologa, R. Prodan, "XCOVNet: Chest X-ray Image Classification for COVID-19 Early Detection Using Convolutional Neural Networks", New Generation Computing, February 2021.

[16] S. Ahuja, B.K. Panigrahi, N. Dey, V. Rajinikanth, T. K. Gandhi, "Deep Transfer Learning-Based Automated Detection of COVID-19 from Lung CT Scan Slices", Applied Intelligence, Vol. 51, No. 1, pp. 571-585, January 2021.

[17] A. Keles, M.B. Keles, A. Keles, "COV19-CNNet and COV19-ResNet: Diagnostic Inference Engines for Early Detection of COVID-19", Cognitive Computation, pp. 1-11, January 2021.

[18] E.F. Ohata, G.M. Bezerra, J.V.S.D. Chagas, A.V.L. Neto, A.B. Albuquerque, V.H.C.D. Albuquerque, P.P. Reboucas, "Automatic Detection of COVID-19 Infection Using Chest X-Ray Images Through Transfer Learning", IEEE/CAA journal of Automatic Sinical, Vol. 8, No. 1, pp. 239-248, January 2021.

[19] P. Silva, E. Luz, G. Silva, G. Moreira, R. Silva, D. Lucio, D. Menotti, "COVID-19 Detection in CT Images with Deep Learning: A Voting-Based Scheme and Cross-Datasets Analysis", Informatics in Medicine Unlocked, Vol. 20, p. 100427, September 2020.

[20] X. Yang, X. He, J. Zhao, Y. Zhang, S. Zhang, P. Xie, "Covid-Ct-Dataset: A CT Scan Dataset About Covid-19", Elsevier Public Health Emergency Collection, Vol. 20, p. 100427, June 2020.

[21] www.kaggle.com/tawsifurrahman/covid19-radiography-database.

[22] E. Soares, P. Angelov, S. Biaso, M.H. Froes, D.K. Abe, "Sars-Cov-2 CT Scan Dataset: A Large Dataset of Real Patients CT Scans for Sars-Cov-2 Identification", medRxiv, May 2020.

[23] D. Painuli1, D. Mishra, S. Bhardwaj, M. Aggarwal, "Forecast and Prediction of COVID-19 Using Machine Learning", Data Science for COVID-19, Elsevier, May 2021.

[24] N. Gupta, D. Gupta, A. Khanna, P.P.R. Filho, V.H.C.D. Albuquerque, "Evolutionary Algorithms for Automatic Lung Disease Detection", Measurement, Vol. 140, pp. 590-608, February 2019.

[25] R.E. Mamlook, S. Chen, H.F. Bzizi, "Investigation of the Performance of Machine Learning Classifiers for Pneumonia Detection in Chest X-ray Images", IEEE International Conference on Electro Information Technology, pp. 98-104, July 2020.
[26] N. Khehrah, M.S. Farid, S. Bilal, M.H. Khan, "Lung Nodule Detection in CT Images Using Statistical and Shape-Based Features", Journal of Imaging, MDPI, February 2020.
[27] M. Mbida, A. Ezzati, "Artificial Intelligence Auscultation System for Physiological Diseases", International Journal on Technical and Physical Problems of Engineering, Vol. 13, No. 4, pp. 97-103, December 2021.
[28] E.H. Hssayni, M. Ettaouil, "Generalization Ability Augmentation and Regularization of Deep Convolutional Neural Networks Using L1/2 Pooling", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 48, Vol. 13, No. 3, pp. 1-6, September 2021.
[29] C.E. Metz, "Basic Principles of ROC Analysis", Seminars in Nuclear Medicine, Vol. 8, No. 4, pp. 283-298, October 1978.
[30] Y. Sasaki, "The Truth of the F-Measure", Teach Tutor Mater, October 2007.

BIOGRAPHIES

Sunil Kumar was born in Kanpur, Uttar Pradesh, India, on September 25, 1988. He earned his B.Tech. at Institute of Engineering and Technology (IET), CSJM University, Kanpur, India in 2008 and the M.Tech. at YMCA Institute of Engineering, Faridabad, India in 2010, which was affiliated by MDU Rohtak, India. He is now pursuing Ph.D. at J.C. Bose University of Science and Technology, YMCA, Faridabad, Haryana, India. His research focuses on identifying lung diseases with the use of machine learning and imaging modalities. He is an Assistant Professor in the Department of Computer Science Engineering at the University Institute of Engineering and Technology (UIET), CSJM University, Kanpur, India. He has over 11 years of teaching in computer science and engineering. He is a member of the All-India Technical Skill Development Council. He qualified for University Grants Commission (UGC)-National Eligibility Test (NET) as an Assistant Professor in Computer Science and Applications, as well as the Graduate Aptitude Test in Engineering (GATE) examination.

Harish Kumar was born in Faridabad, Haryana, India on July 1, 1982. He received his Bachelor and Master degrees in engineering from the YMCA Institute of Engineering in Faridabad, India which was affiliated with MDU Rohtak. He earned his Ph.D. from YMCA University of Science and Technology, Faridabad, Haryana, India. He is an Associate Professor in Department of Computer Engineering at J.C. Bose University of Science and Technology, YMCA, Faridabad, Haryana, India. He is a well-respected educator with a substantial body of research and has more than 16 years of teaching experience. He published more than 30 research papers in various national and international journals. Machine learning, software testing, project management, and computer programming are among his research interests.