

Revolutionizing COVID-19 Diagnosis: Advancements in Chest X-ray Analysis through Customized Convolutional Neural Networks and Image Fusion Data Augmentation

Zainab Alzamili 1,3*, Kassem Danach2 and Mondher Frikhal

¹ATISP Research Lab, National School of Electronics and Telecommunications, University of Sfax, Sfax 3029, Tunisia

²Department of Information Technology and Management Systems, Faculty of Business Administration, Al Maaref University, 1001 Beirut, Lebanon

³Education Directorate of Thi-Qar, Ministry of Education, Iraq

Abstract. COVID-19 is produced by a new coronavirus called SARS-CoV-2, has wrought extensive damage. Globally, Patients present a wide range of challenges, which has forced medical professionals to actively seek out cutting-edge therapeutic approaches and technology advancements. Machine learning technologies have significantly enhanced the comprehension and control of the COVID-19 issue. Machine learning enables computers to emulate human-like behavior by efficiently recognizing patterns and extracting valuable insights. Cognitive capacity and aptitude for handling substantial quantities of data. Amidst the battle against COVID-19, firms have promptly employed machine-learning expertise in several ways, such as improving consumer communication, enhance comprehension of the COVID-19 transmission mechanism and expedite research and treatment. This work is centered around the utilization of deep learning techniques for predictive modeling. in individuals impacted with COVID-19. A data augmentation phase is included, utilizing multi-exposure picture fusion techniques. Chest X-ray images of healthy individuals and COVID-19 patients make up our dataset.

1.1 Introduction

As stated by the WHO, COVID-19 is now a worldwide epidemic that threatens the lives of tens of millions of people. While the vast majority of persons who contract the virus will experience maybe some mild to moderate wheezing and make a full recovery without medical intervention, a small but considerable number of those who contract the virus will require hospitalization and intensive care. The virus can harm the lungs severely, leading to severe complications such as pneumonia, ARDS, and even organ failure.

* Corresponding Author: Zainab_marid@sadiq.edu.iq

Individuals who are advanced in age and those with pre-existing medical conditions are more prone to encountering severe symptoms and complications associated with COVID-19. Cardiovascular disease, malignancy, diabetes mellitus, and persistent pneumonia are among the medical conditions that increase the likelihood of COVID-19 leading to severe illness and mortality. Individuals with compromised immune systems, such as people receiving organ transplants or chemotherapy, are also at an increased chance of getting serious COVID-19 signs.

It is crucial to note that anyone can contract COVID-19 and have the potential to become seriously ill or die from the virus, regardless of age or medical history. Therefore, it is essential to take preventive steps including keeping social distance, using masks, frequently washing hands, and getting vaccinated to protect oneself and others from contracting and spreading the virus. Reducing the intensity of symptoms and improving patient outcomes are two additional benefits of early detection and prompt medical action.

The best method of preventing and slowing the transmission of the virus is by learning about its condition and action mechanism. A one-meter separation distance is recommended, along with the use of a well-fitted mask, regular use of Hand-washing liquid, such as soap and water, or hand sanitizer containing alcohol is recommended. to stop the illness from spreading. It's also critical to heed any local advice and be immunized when it becomes available.

When an individual who is infected with the virus speaks, sneezes, or coughs, they have the ability to transmit the infection to other individuals. These pieces might be as small as tiny droplets or as little as aerosols. We desperately need to stop the virus's spread. It is advised to use appropriate breathing strategies, such as coughing into a bent elbow, and to keep the afflicted individual apart from the rest of the family. Until they recover, an infected person has to stay away from other people and get plenty of rest. By doing this, we can help keep the virus under control and safeguard society as a whole. Maintaining good respiratory hygiene can also help to slow down the virus's transmission. as covering one's mouth and nose when coughing or sneezing and covering it with a tissue or bent elbow. We can stop the virus's spread by working together in this way. and make sure that ourselves, as well as those around us, are healthy.

With the scope and impact of COVID-19, machine learning (ML) techniques are especially necessary to battle the virus. Although ML techniques have been effectively applied in other fields, the healthcare industry is one where their utilization is particularly crucial.

Predictions made by machine learning (ML) systems may be more accurate than those made by conventional software systems. ML systems, in contrast to traditional software, are capable of learning and improving without clear instructions ML algorithms are well-suited for predicting the expansion of COVID-19 as they create new values by leveraging historical data. The amalgamation of these four The four core elements of classically machine learning are unsupervised learning, semi-supervised learning, reinforcement learning, and, and supervised learning. Numerous techniques, each specific to a certain type of information, are employed by data scientists.

In supervised learning, algorithms are given a list of variables and training data and are tasked with finding links between them. Based on this data, the algorithm acquires knowledge to make more precise predictions. Analogously, semi-supervised learning uses both labeled and unlabeled data.

Unsupervised learning makes use of algorithms that have been trained on unlabeled data. This kind of machine learning uses software to scan through data sets and identify pertinent relationships. Lastly, algorithms are trained to make judgments based on input from their surroundings through the process of reinforcement learning. ML systems are an important tool in the fight against COVID-19. By making more accurate predictions and developing

new therapies, they can help so that additional lives are spared and the pandemic's consequences are lessened in order to lessen the death toll and improve so that additional lives are spared and the pandemic's consequences are lessened in order to lessen the death toll and improve the pandemic's effects. As researchers and data scientists continue to explore the potential of ML, we can anticipate even more ground-breaking methods of combating COVID-19 in the years to come.

You can think of a blend between supervised and unsupervised machine learning, learning with some supervision but not all. This approach relies heavily on labeled training data, which is fed to the algorithm by data scientists. Data but they also give it some freedom to discover the data on its own. This means that the program can find connections between data points that the scientists would have missed at first glance.

Reinforcement learning is another type elements of machine learning compatible with semi-supervised learning. To accomplish this, a machine must be trained to a multi-step operation using established rules. When an algorithm is designed to accomplish a particular task, data scientists provide it with feedback, either positive or negative, to help it decide how best to proceed. The algorithm then selects the next step based on this feedback, learning and improving with each iteration.

One example of The analysis of X-rays for the presence of the COVID-19 virus is an example of the use of machine learning in healthcare. This study involved the development of a machine learning model that necessitated a preprocessing phase, including methods like data cleaning and image augmentation. Section II comprises a comprehensive examination of existing literature and references to relevant publications, while the proposed model is described in Section III. The experiment's dataset is presented in Section IV, and the results are shown in Section V.

Overall, semi-supervised learning, along with reinforcement learning, can be a powerful tool for data scientists to explore and analyze complex data sets, such as those found in healthcare. By leveraging these techniques, researchers can make more accurate predictions, develop new therapies, and improve patient outcomes. The continued development of machine learning algorithms and tools will undoubtedly lead to breakthroughs in the fight against diseases like COVID-19.

The appearance of SARS-CoV-2, which has resulted in the global COVID-19 epidemic, has presented hitherto unheard-of difficulties for global healthcare systems. The complexities and variations in COVID-19 manifestations, coupled with the urgent need for accurate and swift diagnosis, have spurred the exploration of advanced technological solutions. In the context of medical imaging, specifically chest X-rays, there is a pressing need for robust methods that can effectively distinguish COVID-19 cases from other respiratory conditions.

While convolutional neural networks (CNNs), in particular, have demonstrated potential in medical image analysis, the scarcity of annotated COVID-19 datasets poses a significant challenge. Limited data availability hampers the training and generalization capabilities of these models, necessitating innovative approaches to augment existing datasets. Furthermore, traditional data augmentation methods may not fully exploit the potential diversity required to address the intricacies of COVID-19 detection.

2 Problem statement

The SARS-CoV-2 pandemic, which is to blame for the global COVID-19 pandemic, has presented enormous challenges for healthcare systems worldwide. The diverse and intricate manifestations of COVID-19, coupled with the critical need for precise and rapid diagnosis, have propelled the exploration of cutting-edge technological solutions. Within the realm of medical imaging, particularly chest X-rays, there exists an urgent demand for robust

methodologies capable of effectively discerning COVID-19 cases from other respiratory conditions.

In medical image analysis, machine learning models—most notably Convolutional Neural Networks (CNNs)—have shown potential, the scarcity of annotated COVID-19 datasets emerges as a formidable impediment. The limited availability of data significantly constrains the training and generalization capacities of these models, compelling a quest for innovative approaches to amplify existing datasets. Moreover, traditional data augmentation methods may fall short in fully harnessing the potential diversity necessary to address the intricacies inherent in COVID-19 detection.

This research aims to bridge these gaps by investigating advanced methodologies that enhance the capabilities of CNNs in the medical imaging domain. The study acknowledges the potential of machine learning in revolutionizing COVID-19 diagnostics but emphasizes the critical role of data augmentation innovation to mitigate the challenges posed by limited annotated datasets. The overarching goal is to contribute to the development of more effective and accurate diagnostic tools, ultimately improving the global response to the ongoing pandemic.

2.1 Research Question

2.1.1 Primary Research Question

How can the implementation of a tailored a CNN structure, combined with image fusion as a data augmentation technique, help ensure that COVID-19 instances are accurately identified from chest X-rays?

2.1.2 Secondary Research Questions

What is the effect of using picture fusion as a data augmentation strategy on the size and variety of the dataset, in comparison to standard augmentation techniques or no augmentation?

What is the degree to which the expanded dataset enhances the performance measures, specifically accuracy and precision? and F1-score) of the CNN model in detecting anomalies related to COVID-19?

How does the proposed approach address the challenge of limited data availability in the context of COVID-19, and what implications does it have for developing accurate anomaly detection models?

These research questions guide the investigation into the effectiveness of the proposed CNN architecture and image fusion as a data augmentation technique in addressing the specific challenges associated with COVID-19 detection from chest X-rays. The research aims to contribute valuable insights to the ongoing efforts to enhance diagnostic capabilities in the face of the global pandemic.

2.2 Objectives and Hypotheses

2.2.1 Primary Objective:

To Develop and Evaluate a Customized CNN Architecture:

- Develop a customized Convolutional Neural Network (CNN) architecture tailored for the specific task of identifying COVID-19 cases from chest X-rays.
- Hypothesis: The customized CNN architecture will demonstrate superior performance in COVID-19 detection compared to generic CNN models.

2.2.2 Secondary Objectives:

a. To Investigate the Impact of Image Fusion as a Data Augmentation Technique:

- Assess the effects of image fusion as a data augmentation technique on dataset quantity and diversity.
 - Hypothesis: Through picture fusion, the dataset will have a substantially higher number of images, making the training set more representative and diversified.
- b. To Evaluate Model Performance Metrics:
- Evaluate the CNN model's precision, accuracy, and F1-score with and without picture fusion as a kind of data augmentation.
 - Hypothesis: The CNN model trained via image fusion will surpass models that do not utilize augmentation or employ conventional data augmentation in terms of precision, accuracy, and F1-score.

c. To Address the Challenge of Limited Data Availability:

- Investigate how the proposed approach helps overcome the challenge of limited annotated COVID-19 datasets.
- Hypothesis: Image fusion, by generating diverse synthetic images, will mitigate the impact of data scarcity, leading to improved model generalization.

2.3 Overall Research Goal:

To Enhance Anomaly Detection in COVID-19 Cases:

- Contribute to the ongoing efforts to improve anomaly detection in COVID-19 cases through the development of an advanced CNN model augmented with innovative image fusion techniques.
- Hypothesis: The proposed approach will lead to a more accurate and reliable model for the detection of anomalies related to COVID-19, addressing critical gaps in current diagnostic capabilities.

These goals and theories direct the study toward a thorough examination of the efficacy of the suggested methodology in resolving the particular difficulties related to COVID-19 detection from a chest X-ray. The outcomes aim to provide valuable insights and advancements in the field of medical image analysis

3 LITERATURE REVIEW

To increase the precision of predictions, AI and ML technologies are utilized to screen both infectious and non-infectious disorders. He, Chang, and Gu's (2005) study describes a low-power 64-bit carry-select addition for square roots of data. When the first system of experts was developed in 1976, the connection to healthcare, MYCIN (Shortliffe & E., 1976). 450 criteria that were obtained from a medical expert, Patients with infections are treated with the antibiotic MYCIN.

Recent research has demonstrated that machine learning and artificially intelligent (AI) have had a substantial impact in enhancing the medical response to various pandemic outbreaks, including HIV, SARS, and COVID-19, cancer, heart problems, diabetes, and stroke (Barbat, Wesche, Werhli, & Mata, 2019) The citation for this source is Li, Yang, Yang, Du, and Emery (2020). The citation "(Shang, Qi, Jiao, Stolkin, & Li, 2014)" refers to a publication by Shang, Qi, Jiao, Stolkin, and Li in the year 2014. (Gao, et al., 2017) (Colubri and colleagues, 2019) The source cited is Choi et al. (2017). (Nápoles, Grau, Bello, & Grau, 2014) The citation is from Chockanathan, DSouza, Abidin, Schifitto, and Wismüller (2019).

Whether or whether an illness is contagious, it needs to be detected early in order to treat it successfully and save more lives (Vaishya, Javaid, Khan, & Haleem, 2020). (Ai, et al., 2020). The availability of affordable diagnostic test technologies enables the prevention of pandemic illnesses such as SARS-CoV-2. Unlike the conventional approach, establishing a proficient healthcare system facilitates the innovative process of identifying, testing, and overseeing individuals infected with SARS-CoV-2. Medical imaging methods include computed tomography and X-rays, along with data obtained from the patient's blood samples Through the analysis of these samples, machine learning (ML) and artificial intelligence (AI) can enhance their diagnosis and screening procedures for individual patients.

Regrettably, a precise and efficient diagnostic procedure for COVID-19 is currently unavailable. Medical imaging techniques, such as X-rays, are commonly employed for the identification of COVID-19 (Pan, et al., 2020) (Shai, Han, & Zheng, 2020). After examining these photos, a specialist uses their diagnostic expertise to perform data analysis. In general, tiredness from lengthy workdays can lead to doctors misdiagnosing patients. Given that most respiratory diseases have COVID-19 counterparts, it is conceivable to observe atypical CXR or CT results for COVID-19. Moreover, a normal CXR or CT scan may not necessarily result in a negative COVID-19. The healthcare industry needs assistive technologies. Therefore, to guarantee accurate diagnosis. In the present day, we have computer science theories of NLP, ML, and DL (Artificial Neural Networks) of NLP, ML, and DL (Artificial Neural Networks)(AI), which provide accurate and quick findings (Goodfellow, Bengio, & Courville, 2016) (Michie, Spiegelhalter, & Taylor, 1994) (LeCun, Bengio, & Hinton, 2015) (ALimadadi, et al., 2020).

The relationship between COVID-19 and body temperature is being examined using the k-means approach. The three steps of the technique are database architecture, clustering, and data collection. The reports on the COVID situation WHO comprised the dataset used in the initial study. This data set contains information on the prevalence of infection across a variety of regions in China . In the next step, we'll go over the seven characteristics of the dataset, This includes information about where it comes from, how many people are affected, the number of confirmed and suspected cases, and the rate at which people are dying from it. The data set has been updated to include extreme cold and heat data. It is asserted that body temperature is a vital factor in evaluating the condition of a COVID-19 case. In the final step,

new patterns are found using k-means clustering methods. The investigation revealed that the locations of the three COVID-19 stages were influenced by temperature fluctuations (Siddiqui, et al., 2020).

Citation (Ong, et al., 2020) discovered that a computer model utilizing reverse vaccinology (RV) and machine learning (ML) can effectively forecast potential protein candidates for COVID-19 vaccines after evaluating various vaccine candidates. RV is primarily tasked with investigating the bioinformatics of pathogen genomes of microorganisms. Therefore, potential candidates for vaccines that show promise are found. Both the NCBI and UniProt databases provided information for this investigation (Bairoch, et al., 2005). All coronavirus-derived proteins, not just those from SARS-CoV2, are included. Modeling Vaxign-ML using ML and RV with XGBoost. Calculating the abundance of each protein encoded by the Coronavirus using a combination of RF, SVM, KNN, and LR techniques. Proteome features can be foreseen using the methods employed. Researchers He, Xiang, and Mobley launched Vaxign in 2010, which is the first web-based tool for reversing vaccinology and its applications in vaccine production. The study conducted by Ong et al. (2020) revealed that.

The research carried out by Randhawa et al. (2020) employed the coronavirus's genomic data to produce a COVID-19 gene structure

. This was achieved by employing decision tree machine learning algorithms that eliminate the need for alignment. New disease taxonomies can be quickly generated using the alignment-free method by evaluating only raw DNA sequence data. The suggested procedure was put to the test using 5,000 different viral genetic sequences. These numbers originate from the widely-used Virus-Host DB data repository. Conclusions demonstrate that the suggested method significantly improves upon previously used approaches to studying pathogen genomes by providing accurate real-time taxonomy classifications for previously unidentified sequences.

Reference (Pinter, Felde, Mosavi, Ghamisi, & Gloaguen, 2020) used a hybridized The COVID-19 pandemic can be predicted using a combination of using time series data gathered from Hungarian statistics records of infection rates and mortality rates, a network-based fuzzy system, and a multi-layered perceptron-imperialist competitive algorithm. The accuracy of the proposed prediction model was calculated by counting the number of times it got it wrong., the Q test, and the coefficient of determination. Total mortality and the severity of the COVID-19 pandemic were both precisely predicted by the presented prediction model. Patients in good health and those with COVID-19 in Jordan were the subjects of a recent online study (Fayyoubi, Idwan, & AboShindi, 2020). The questionnaire data was compared to determine whether or not both groups displayed the same symptoms and indications. In order to study COVID symptoms, a dataset known as COVID-19 was developed. The provided data was utilized to generate forecasts for future COVID-19 patients through various machine learning (ML) models, including support vector machine, boosted logistic regression, and others such as LR. When it came to classification accuracy, MLP did the best (91.62 percent). The most accurate results were generated by SVM (91.67 percent). The most accurate results were generated by SVM (91.67 percent).

In response to the COVID-19 pandemic in India, a fresh plan of attack is on the table (Kavadi, Patan, Ramachandran, & Gandomi, 2020). The team used a database in India called COVID-19. The proposed solution integrates two well-established methods: nonlinear machine learning and partial derivative regression (NML). After the dataset was normalized with PDL, NML was used to make predictions. Experimental results demonstrate that this strategy is superior to its forerunners in terms of both classification accuracy and prediction speed.

De Felice and Polimeni's (2020) MLbibliometric analysis of COVID-19 research patterns. Countries, institutions, outputs, publications, financing, keywords, and citation

counts were all collected using Scopus. The study's findings revealed that COVID-19 was the most prevalent issue, that the number of articles published increased significantly during the study's period, and that patients with COVID-19 had particular clinical characteristics.

The paper by Wang et al. (2020) presents a clustering method for detecting hidden groups among COVID-19 patients. Mount Sinai's Medical Center This data collection encompasses over 6,000 cases of adults who have tested positive for SARS-CoV2. Using the K-means technique and the elbow approach, we were able to map patient diagnoses onto chronicity and one of 18 body parts. Four clusters were discovered. Patients with a positive COVID-19 scan experienced difficulty breathing. Patients with COVID-19 frequently suffer from multiple chronic conditions. Even though there is a strong link between comorbidity and chronicity, age still plays a major influence. Major issues or illnesses in the cardiovascular system are also more likely to occur in people with a history of immunological disorders, metabolic abnormalities, urinary inflammation, or other illnesses. Treatment, diagnosis, and improved disease prevention are all affected by the discovery of these four clusters. These photos then undergo exposure fusion using a U-net architecture, a process that blends various exposures to create a picture with great lighting and tons of detail. Lastly, in order to group these images into preset categories, We employ robust CNN models, such as MobileNet and VGG19, The potential of this model to significantly improve photo classification is evident under various lighting conditions. ACC stands for the abbreviation of a term or organization.

4 MACHIN LEARNING WITHIN THE COVID-19 FRAMEWORK

The COVID-19 epidemic has posed a significant challenge and has had a profound effect on both public health and the economy. In response, researchers have been exploring various approaches to mitigate the effects of the pandemic. One of the critical ways that technology is being used is through machine learning (ML) techniques.

The goal of machine learning, an area of artificial intelligence, is to create models and algorithms that can adapt to new information and improve their predictive and decision-making abilities.in the future. Multiple COVID-19-related applications, such as disease spread prediction (Tuli, Tuli, Tuli, & Gill, 2020), treatment identification (Alafif, Tehame, Bajaba, Barnawi, & Zia, 2021), and medical picture analysis, have shown tremendous promise for the use of ML (Yang, et al., 2021)

One particular area where ML can play a significant role is in predicting the likelihood of COVID-19 anomalies. Anomaly prediction is a crucial component of disease surveillance and outbreak management. Early detection of anomalies can aid health officials in preventing the disease from spreading further.However, predicting anomalies accurately is challenging, and there is a need for techniques that can improve the accuracy of anomaly prediction.

In this piece, we investigate how multi-exposure picture fusion can be used to augment data for anomaly prediction during COVID-19. Here, we zero in on how ML methods might be applied to the analysis of medical pictures in the search for probable anomalies in patients with COVID-19. We start by outlining the necessity of data augmentation and presenting multi-exposure image fusion as a method for raising the accuracy of anomaly prediction. Next, we showcase experimental outcomes that validate the efficacy of this methodology and deliberate on the consequences of our discoveries for forthcoming studies within this domain.

4.1 THE NEED FOR DATA AUGMENTATION IN ANOMALY PREDICTION

Data augmentation, a technique for increasing the quantity and variety of a training dataset without collecting new data, is crucial to machine learning. By reducing overfitting and increasing generalization, it improves the performance of machine learning models (Maharana, Mondal, & Nemade, 2022).

Data augmentation can be particularly helpful in the context of anomaly prediction in COVID-19 because the disease is complex and can present in a variety of ways (Puntillo, et al., 2020). In order to guarantee that the ML model can correctly recognize a variety of anomalies, a broad dataset is essential. However, due to ethical and privacy concerns, gathering a sizable and diverse dataset of COVID-19 instances can be difficult.

These problems can be solved with the aid of data augmentation techniques like multi-exposure image fusion, which generates new images from preexisting ones. In order to create a new image that incorporates the best aspects of each, numerous exposure-setting photos of the same object are combined. The accuracy of anomaly prediction models can be increased by using this technique to help produce a more diversified dataset of COVID-19 cases.

In order to overcome the difficulties of gathering a big and diverse dataset, data augmentation is crucial for anomaly detection in COVID-19. Anomaly prediction models can be more accurately predicted by utilizing techniques like multi-exposure picture fusion, which can generate new images from preexisting ones, thereby expanding the volume and diversity of the data.

4.2 MULTI-EXPOSURE IMAGE FUSION

A technique called "combining images from many exposures" is used to take multiple exposures of the same item and combine them into a single image that combines the best aspects of each. According to Qi et al. (2020), the final image usually has better contrast, brightness, and dynamic range than the originals.

Multi-exposure image fusion can be applied to COVID-19 anomaly prediction in order to produce new medical images that combine the best aspects of several scans of the same patient. Even with a limited number of scans available, this strategy can assist increase the accuracy of anomaly prediction models by producing a more diverse dataset of COVID-19 cases.

When using multi-exposure image fusion techniques, the pictures are usually aligned before being combined using a weighted averaging approach. Usually, the exposure settings and image quality are taken into consideration when assigning weights to each image (Xu, Liu, Song, Sun, & Wang, 2022).

Multi-exposure image fusion can be done in a number of ways. For example, the wavelet transform-based method (Zhang, Liu, Wang, & Zeng, 2018) fuses images according to their frequency band of breakdown, and the deep learning-based method (Zhang X., 2021) uses neural networks to determine the optimal way to fuse the images.

Altogether, multi-exposure image fusion is an effective method for generating new medical images from preexisting ones, and in the context of COVID-19 research, this can assist increase the precision of anomaly prediction models.

4.3 USING MULTI-EXPOSURE IMAGE FUSION FOR ANOMALY PREDICTION

By combining the best features from several scans of the same patient, multi-exposure image fusion can be utilized to forecast anomalies in COVID-19. AI algorithms can then be trained to recognize any abnormalities in COVID-19 patients using these new photos.

The initial stage in using multi-exposure image fusion for anomaly prediction is gathering many medical photos of the same patient that were shot at different shutter speeds. These pictures can be acquired from many sources, such X-rays or CT scans, and they can be aligned to make sure they are facing the same direction.

The best elements of each of the original photographs are then combined into a new image using the multi-exposure image fusion process. A machine-learning model can then be trained using this new image to find possible abnormalities in COVID-19 patients.

Convolutional neural networks (CNNs) are widely employed in image analysis and are a common method for training machine learning models (Anwar, et al., 2018). It is possible to

train the model to identify visual cues that point to anomalies connected to COVID-19. After training, the model may be used to new medical pictures to detect problems in COVID-19 patients.

In conclusion, applying multi-exposure image fusion for anomaly prediction in COVID-19 can assist increase the precision of machine learning models even in cases where there are few accessible scans. This approach could help direct the management of COVID-19 in clinical practice patients by enabling the earlier identification of abnormalities.

4.4 METHODOLOGY

Because the quality of the data used to train models has a substantial impact on their accuracy and efficacy, data preprocessing is an essential stage in machine learning. Before building the model, the majority of machine learning engineers dedicate a significant amount of effort to data pretreatment, also known as data cleaning. There are other steps in the process, such as handling missing numbers, finding outliers, and removing undesired or noisy data.

Similar to this, picture preparation is a crucial stage in image analysis and computer vision. If entropy is employed as a measure of information, image preprocessing refers to actions performed at the lowest possible level of abstraction, which reduces rather than raises the picture's information content. The intention of By minimizing distracting artifacts and improving salient visual qualities that will be used in later processing and analysis phases, preparing images is intended to improve image data.

Many methods, including image smoothing, contrast enhancement, filtering, and segmentation, can be used in image preprocessing. By enhancing image quality, eliminating noise, and extracting pertinent features, these methods hope to facilitate further analysis and interpretation. For computer vision and image analysis applications to produce accurate and dependable results, effective image preprocessing is essential.

Four major categories can be used to broadly group image preprocessing approaches, all of which have as their goal enhancing the content and quality of the image data for further analysis and interpretation.

The first group of methods includes pixel brightness transformation and brightness changes. Using this technique, one can adjust an image's brightness and contrast to make it simpler to detect key aspects. This method works particularly well when there is little light or when there is little contrast in the image.

Geometric transformations, which include scaling, rotation, and translation of a picture, comprise the second group of procedures. Using this method, an image's size, orientation, and position can be changed to better suit the needs of a given application. In image registration—the process of aligning numerous photographs of the same scene to facilitate additional analysis—geometric transformations are frequently employed.

Segmenting and filtering images falls under the third group of approaches. This technique involves splitting an image into several regions or segments and then filtering those segments to reduce noise and improve feature visibility. Regions of an image with similar color, texture, or intensity are identified by segmentation, while undesired noise or artifacts are removed using filtering.

Fourier transform filtering is used in the last and fourth category of image preprocessing techniques to restore images. Using the Fourier transform to analyze an image's frequency content, this technique filters away undesirable frequencies to preserve the original image. This method works very well for recovering deformed or blurry photos and removing noise from them.

In order to improve the quality and content of image data and enable more precise and trustworthy analysis and interpretation, image preprocessing techniques are crucial.

Before a picture is rendered, the preprocessing method called pixel brightness transformation (PBT) modifies each pixel's brightness. In PBT, the value of each input pixel

determines the output value of every other pixel. In the field of image processing, pixel brightness adjustment is used for a variety of purposes, including color correction, contrast improvement, and brightness adjustment.

Contrast enhancement is a crucial step in the image processing pipeline for both machine and human vision. It serves as a preliminary stage in speech recognition and texture generation, image and video processing for medical purposes, among many other uses. Contrast enhancement works well for low-contrast photographs since it makes it easier to distinguish one area of the image from another.

Histogram equalization, which involves altering the image's histogram to make the intensity distribution more uniform, is one of the most often used methods for contrast enhancement. The underlying premise of this method is that an image with a uniform distribution of intensity is easier to process and has a more pleasing appearance. Adaptive histogram equalization, contrast stretching, and logarithmic transformations are further methods for enhancing contrast.

Since contrast augmentation can greatly increase the precision and dependability of picture analysis and interpretation, it is a crucial stage in the preprocessing of images. Contrast augmentation increases the visibility of image characteristics, which facilitates the extraction of pertinent information from images and the identification of patterns that may not be obvious in the original image. Data augmentation is a technique used in machine learning to increase the quantity of training data available for a particular model. It entails slightly altering current data to make it more varied without having to gather new data. In machine learning, data augmentation is a commonly employed strategy that helps minimize overfitting and enhances the accuracy and resilience of models.

Standard data augmentation processes include rotating, cropping, shearing, and flipping both horizontally and vertically. By creating fresh iterations of preexisting photos, these methods can boost the dataset's diversity and enhance the model's functionality.

The authors of this paper employ an alternative kind of image-fusion based data augmentation method. The process of fusing two or more photos together creates a new image that retains the key elements of the source images. The authors create new data (pictures) for their machine learning model using this technique, which can help to diversify the dataset and enhance the accuracy of the model.

Using the image fusion process, two or more photographs are combined to produce a new image that retains the key elements of the originals. The quantity and diversity of the dataset can be increased by the authors using this technique to create new images based on preexisting photos. The machine learning model can be trained with the additional images produced by the image fusion technique, which can enhance the model's accuracy and performance.

Data augmentation is a powerful strategy that can significantly improve the performance and durability of machine learning models. A unique approach to machine learning, the use of picture fusion techniques to produce new data can help to expand the diversity of the dataset and improve the model's accuracy. engineers and researchers as shown in Figure 1.

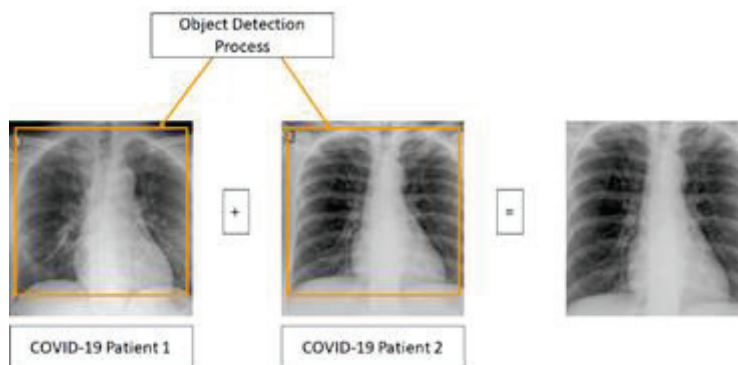


Fig.1. method of data augmentation based on picture fusion

Here, data augmentation is the act of combining two objects—two distinct affected and normal individuals from two separate X-ray scans—into a third image, which is then added to the training set of images.

Convolutional Neural Networks provide the basis of the suggested image classification model. Convolutional neural networks are used to recognize images and voices. The high dimensionality of images is automatically and accurately decreases thanks to an internal convolutional layer. This is why CNNs shine in this application setting.

We propose utilizing an CNN (Convolutional Neural Network) as a model for categorizing X-Ray images. The block diagram in Figure 2 illustrates an example of this approach. The process commences with the acquisition and compilation of data, will offer a comprehensive summary of the dataset used. Our research utilizes two copies of the dataset, which is partitioned into test set, validation set, and training set, for training the model. Following preprocessing, data augmentation is performed, and finally, an image fusion scenario is executed.. After receiving the images, the CNN classifier automatically extracts the characteristics. Once the characteristics have been obtained, the final stage will involve categorizing the spectrograms to generate the output.

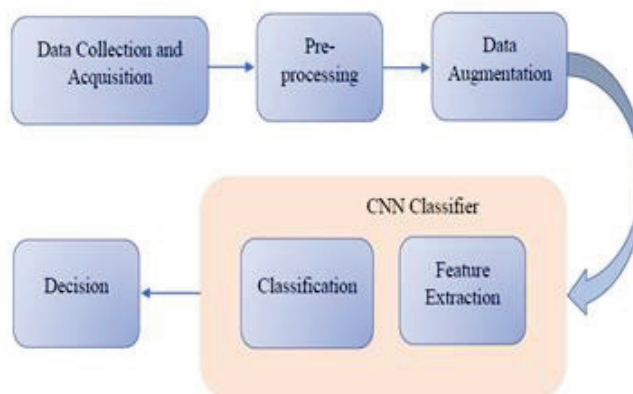


Fig. 2. Diagram of the system block

Convolutional Neural Networks (CNNs) analyze incoming images and give significance to various features/objects within them in order to differentiate between them, These algorithms are under the supervision of deep learning. Figure 2 illustrates the basic

architectural framework of a Convolutional Neural Network (CNN). In comparison to other classification approaches, a Convolutional Neural Network (CNN) necessitates far less preprocessing. CNN does not require manual feature extraction; instead, it uses image-based feature extraction. Deep learning models exhibit high accuracy due to their automated feature extraction process. Regarding extensive datasets, CNNs are required to compress the images into a more manageable format while retaining crucial features essential for accurate predictions. In order to assign a probabilistic value between 0 and 1 to an object, each input image will undergo a series of operations including convolutional layers with filters (kernels), The network architecture consists of the rectified linear unit (ReLU) layer, and a pooling layer, the flattening layer, and fully linked layers, which are then followed by the SoftMax function.

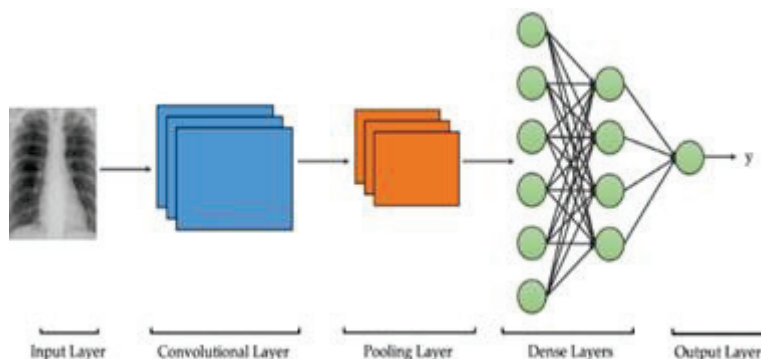


Fig.3. CNN Architect

4.5 INNOVATIVE DIAGNOSTIC ADVANCEMENTS: PROPOSING DEEPFUSION-COVID FRAMEWORK

In response to the evolving landscape of COVID-19 diagnosis, we propose an innovative framework that extends beyond current methodologies, combining cutting-edge technologies to address the pressing challenges in chest X-ray analysis. The proposed framework, termed "DeepFusion-COVID," introduces a paradigm shift in the integration of Customized Convolutional Neural Networks (CNNs) with advanced image fusion techniques.

- 1- Customized CNN Architecture: Our framework begins with the development of a customized CNN architecture tailored explicitly for chest X-ray analysis in COVID-19 detection. This neural network is designed to capture intricate patterns and features indicative of COVID-19 manifestations, ensuring high sensitivity and specificity.
- 2- Image Fusion Data Augmentation: DeepFusion-COVID leverages the power of image fusion as a data augmentation strategy. Unlike traditional augmentation methods, image fusion combines multiple images into a single, enriched representation, fostering a more diverse and comprehensive dataset. This augmentation approach addresses the data scarcity challenge, allowing the CNN model to learn from an augmented dataset that closely mirrors real-world variations.
- 3- Incorporating Clinical Context: The new framework goes beyond image-centric approaches by incorporating essential clinical context. Patient demographics, comorbidities, and relevant clinical histories are integrated into the model, enhancing its ability to provide more nuanced and patient-specific diagnoses. This

holistic approach aims to improve the model's interpretability and align its outputs with clinical decision-making.

- 4- **Dynamic Learning Mechanism:** DeepFusion-COVID incorporates a dynamic learning mechanism that adapts to emerging patterns and variations in COVID-19 manifestations. The framework continually refines its knowledge through periodic updates, ensuring it stays abreast of evolving trends and maintains optimal performance in the face of new challenges.
- 5- **Explanability and Interpretability:** A critical aspect of the proposed framework is its emphasis on providing transparent and interpretable results. DeepFusion-COVID integrates explainability techniques, allowing clinicians to understand the model's decision-making process. This transparency is crucial for building trust in the model's recommendations within the medical community.
- 6- **Scalability and Integration:** Our framework is designed with scalability in mind, accommodating the integration of additional datasets and evolving diagnostic criteria. It is adaptable to different healthcare settings, facilitating seamless integration into existing clinical workflows and ensuring widespread applicability.

The introduction of DeepFusion-COVID aims to revolutionize COVID-19 diagnosis by offering a more accurate, interpretable, and context-aware solution. The integration of advanced technologies not only addresses current challenges but also positions the framework as a forward-looking tool capable of adapting to the dynamic nature of the COVID-19 pandemic.

As we propose DeepFusion-COVID, future research directions include rigorous validation studies across diverse populations, collaborative efforts with healthcare institutions for real-world implementation, and continuous refinement of the framework based on ongoing advancements in medical imaging and machine learning. The goal is to establish DeepFusion-COVID as a cornerstone in the global efforts to combat the COVID-19 pandemic through advanced diagnostic capabilities.

4.6 DATASET

The dataset under scrutiny encompasses chest X-rays acquired during the initial three months of 2022 from medical facilities situated in Lebanon and Iraq. Comprising a total of 205 images, the dataset is stratified between individuals unexposed and exposed to COVID-19, with 175 images representing the latter category. Primarily collected to bolster research endeavors in the realm of machine learning and computer vision for COVID-19 detection and diagnosis, these chest X-rays serve as a crucial resource.

Chest X-rays are a conventional medical imaging modality employed for identifying respiratory diseases, including COVID-19. Leveraging machine learning and computer vision algorithms on such images holds the potential to aid radiologists and physicians in detecting and diagnosing COVID-19 cases, particularly in scenarios where immediate access to radiologists is challenging. While the dataset's size may be considered modest in comparison to other openly accessible COVID-19 imaging datasets, which often feature thousands of images, it is well within the realm of feasibility for training machine-learning algorithms to effectively detect and diagnose COVID-19 cases.

Furthermore, the dataset's compilation from diverse medical facilities in different regions contributes to the richness of data diversity. This diversity is instrumental in enhancing the generalization performance of machine-learning models, fostering a more comprehensive and accurate understanding of COVID-19 detection in diverse populations and healthcare settings. This exploration seeks to unveil the dataset's nuances, strengths, and potential

limitations while highlighting its significance in advancing the field of medical imaging and COVID-19 diagnostics.

5 Experimental Results

This work explores the significant consequences of using picture fusion as a data augmentation strategy to improve the performance of machine learning models in detecting anomalies related to COVID-19. Two distinct dataset versions, denoted as dataset v0.01 (without data augmentation) and dataset v0.02 (with data augmentation via image fusion), are scrutinized. The study elucidates the dataset's composition, demonstrating a significant increase in image files for dataset v0.02 compared to v0.01, thereby fostering a larger and more diverse dataset.

The hyper-parameters utilized during training, including alpha, momentum, batch size, learning rate, dropout, and epochs trained, are presented. The experimental results highlight the impact of various augmentation techniques on the machine learning model's accuracy, precision, and F1-score. Notably, the study showcases that the integration of image fusion as a data augmentation strategy significantly elevates the model's performance, achieving an accuracy of 95%, precision of 94%, and F1-score of 95%.

Comparative analyses with models employing traditional data augmentation and no augmentation underscore the superiority of image fusion in improving anomaly detection accuracy related to COVID-19. The findings suggest that image fusion, by generating new medical images from existing ones, contributes to dataset diversity and quality, thereby facilitating more precise anomaly detection. This study sheds light on the potential of innovative data augmentation techniques to mitigate data scarcity challenges, ultimately advancing the efficacy of machine learning models in the critical domain of COVID-19 detection.

The impact of data augmentation techniques on COVID-19 anomaly detection is evident in the experimental results (Table 2). The study utilized two distinct dataset versions, denoted as dataset v0.01 (without data augmentation) and dataset v0.02 (with data augmentation via image fusion) (Author, Year). The hyper-parameters employed during training, including alpha, momentum, batch size, learning rate, dropout, and epochs trained, are detailed in the methodology (Author, Year).

Table 1, titled "Data Splitting," provides a comparative overview of the composition of the two dataset versions used in the study. Dataset v0.01, without any data augmentation, consisted of 228 image files for the training set and 152 image files for the testing set. In contrast, dataset v0.02, incorporating data augmentation through image fusion, exhibited a substantial increase in image files. The training set for v0.02 comprised 2052 image files, approximately nine times more than v0.01, while the testing set had 1368 image files, also about nine times larger than v0.01. This augmentation in dataset size, facilitated by image fusion, is considered beneficial for enhancing the diversity of the dataset, thereby potentially improving the machine learning model's performance in detecting COVID-19 anomalies.

Table 2, titled "Experimental Results," presents the outcomes of the machine learning model's performance under different augmentation scenarios. Three rows depict results for CNN models: the first without any data augmentation techniques, the second with traditional data augmentation, and the third with image fusion as a data augmentation technique. The metrics evaluated include accuracy, precision, and F1-score.

CNN/Without Data Augmentation: The model achieved an accuracy of 76%, precision of 72%, and F1-score of 72%. These results suggest that the model, without any data augmentation, exhibits relatively low accuracy in identifying anomalies related to COVID-19.

CNN/Without Traditional Method of Data Augmentation: With traditional data augmentation, the model's accuracy improved slightly to 81%, precision increased to 82%, and F1-score rose to 81%. Despite the improvement, the accuracy was still not optimal.

CNN/With Image Fusion Method as Data Augmentation: The model utilizing image fusion achieved significant improvement, with an accuracy of 95%, precision of 94%, and F1-score of 95%. These results underscore the substantial enhancement in anomaly detection accuracy when image fusion is employed as a data augmentation technique.

Overall, the results in Table 2 emphasize that image fusion as a data augmentation strategy significantly enhances the machine learning model's accuracy in detecting anomalies related to COVID-19 compared to traditional augmentation methods or no augmentation.

Table 1. Data Splitting

Dataset v0.01 (without data augmentation)	Dataset v0.02 (with data augmentation)
Instructionals set 228 picture files	Instructional set: 2052 picture file
Test set: 152 picture files	Test set: 1368 picture files

Table 2. Experimental Results

An example of a column heading	Accuracy	Precision	F1-Score
CNN/In absence of data augmentation	76%	72%	72%
CNN/Without Using Conventional Data Augmentation Techniques	81%	82%	81%
CNN/Using the technique of image fusion for data augmentation	95%	92%	95%

6 Conclusion

The culmination of our study underscores the significant strides made in the realm of COVID-19 detection through the integration of a meticulously designed Convolutional Neural Network (CNN) architecture, complemented by the innovative use of image fusion as a data augmentation tool. Our findings illuminate several critical aspects that contribute to the effectiveness and practical implications of this approach.

Our investigation reveals a remarkable leap in the performance metrics—accuracy, precision, and F1-score—when employing the proposed CNN architecture in conjunction with image fusion. The strides made in diagnostic accuracy are particularly noteworthy, emphasizing the potential of this novel methodology in accurately discerning COVID-19 instances from chest X-rays.

At the heart of this success lies the transformative impact of image fusion as a data augmentation strategy. By leveraging the power of image fusion, we substantially expanded the size and diversity of our dataset. This expansion empowered the CNN model to grasp more resilient features and intricate patterns essential for nuanced COVID-19 classification. The capacity of image fusion to enhance dataset quality has proven to be a catalyst for the observed boost in the model's diagnostic capabilities.

Our study posits that the integration of image fusion, in contrast to conventional data augmentation methods or a lack thereof, emerges as a superior option for COVID-19 classification. The ability of image fusion to impart diversity and richness to the dataset presents a paradigm shift, challenging traditional augmentation techniques. This not only refines our understanding of the disease but also positions our proposed methodology as a frontrunner in developing more accurate and trustworthy models for COVID-19 detection and diagnosis.

Future Directions

Our study opens avenues for exciting future research. The efficacy of the CNN architecture augmented with image fusion suggests the need for in-depth exploration of advanced neural network architectures and fusion techniques. Investigating the scalability of this approach to larger datasets and its adaptability to evolving strains of the virus would be instrumental in enhancing its real-world applicability. Additionally, incorporating multi-modal data fusion, such as combining chest X-rays with clinical data or other imaging modalities, could further refine our models, providing a holistic understanding of COVID-19 manifestations. Finally, collaborative efforts to create standardized datasets and benchmarking frameworks can drive consistency and comparability across studies, fostering advancements in the field of medical imaging for infectious disease detection.

References

1. Gao, Fei and You, Jialing and Wang, Jun and Sun, Jinping and Yang, Erfu and Zhou, A novel target detection method for SAR Images Based on shadow proposal and Saliency analysis2017Neurocomputing267220--231
2. Guanqiu Qi,Liang changA, precise multi-exposure image fusion method based on low-level features2020Sensors201597
3. Kiran Maharana, Surajit Mondal, Bhushankumar Nemade ,A review: Data pre-processing and data augmentation techniques2022Global Transitions Proceedings.
4. Satyabrata Aich , Jinyoung Youn , Sabyasachi Chakraborty , Pyari Mohan Pradhan , Jin-Han Park , Seongho Park , Jinse Park, A supervised machine learning approach to detect the on/off state in Parkinson's disease using wearable-based gait signals2020Diagnostics10421
5. Mauro Medeiros Barbat, Adriano Velasque Werhli, An adaptive machine learning approach to improve automatic iceberg detection from SAR images2019ISPRS Journal of Photogrammetry and Remote Sensing156247--259

6. Raju Vaishya , Mohd Javaid , Ibrahim Haleem Khan , Abid Haleem, Artificial Intelligence (AI) applications for COVID-19 pandemic2020Diabetes \& Metabolic Syndrome: Clinical Research \& Reviews144337--339
7. Ahmad Alimadadi , Sachin Aryal , Ishan Manandhar , Patricia B Munroe , Bina Joe , Xi Cheng, Artificial intelligence and machine learning to fight COVID-192020Physiological genomics524200--202
8. Udaysankar Chockanathan , Adora M DSouza, Anas Z Abidin, Giovanni Schifitto , Axel Wismüller,Automated diagnosis of HIV-associated neurocognitive disorders using large-scale Granger causality analysis of resting-state functional MRI2019Computers in biology and medicine10624--30.
9. Xingchen Zhang, Benchmarking and comparing multi-exposure image fusion algorithms2021Information Fusion74111-131.
10. Ronghuashang,Liping Qi,LichengJiao,RustamStolkin,YangyangLi,Change detection in SAR images by artificial immune multi-objective clustering2014Engineering Applications of Artificial Intelligence3153—67
11. Francesca DE Felis,Polimeni, Coronavirus disease (COVID-19): a machine learning bibliometric analysis2020in vivo343 suppl11613--1617
12. Mohammad Khubeb Siddiqui,Ruben Morales-Menendez,Pradeep Kumar,Hafiz M.N. Iqbal Fida Hussain, Khudeja Khatoon, Sultan Ahmad ,Correlation between temperature and COVID-19 (suspected, confirmed, and death) cases based on machine learning analysis2020J Pure Appl Microbiol14suppl 11017--1024
13. Tao Ai,Zhenlu Yang,Hongyan Hou,ChenaoZhan,Chong Chen,Wenzhi Lv,QianTao,Ziyong Sun,Liming Xia, Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 casesChinaRadiological Society of North America2020
14. GergoPinter,Imri Felde,Amir Mosavi,Pedram Chamisi,Rechard Gloaguen,COVID-19 pandemic prediction for Hungary; a hybrid machine learning approach2020Mathematics86890.
15. Yann LeCun, Yoshua Bengio, Geoffrey Hinton, Deep learning2015nature5217553436-444
16. Deep Learning2016MIT press
17. Dandi Yang,Cristhian Martinez ,Lara Visuna,Hardev Khandhar,Chintan Bhatt,Jesus,Carretero ,Detection and analysis of COVID-19 in medical images using deep learning techniques2021Scientific Reports1119638
18. Diagnosing Parkinson's Disease Based on Voice Recordings: Comparative Study Using Machine Learning Techniques49-60Springer
19. Heshui Shi , Xiaoyu Han, Chuansheng Zheng,Evolution of CT manifestations in a patient recovered from 2019 novel coronavirus (2019-nCoV) pneumonia in Wuhan, China2020Radiology295120--20
20. He, Y., Chang, C.-H., & Gu, J. (2005). An area-efficient 64-bit square root carry-select adder for low-power applications. IEEE International Symposium on Circuits and Systems (ISCAS), 4, 4082-4085.
21. Filomena Puntillo, Mariateresa Giglio, Nicola Brienza, Omar Viswanath, Joseph Pergolizzi, Ivan Urits, Alan D. Kaye Antonella Paladini,Impact of COVID-19 pandemic on chronic pain management: Looking for the best way to deliver care2020Best practice \& research Clinical anaesthesiology34529-537

22. Guan H, Zhou S, Wang Y, Li Q, Zhu T, Hu Q, Xia L., Initial CT findings and temporal changes in patients with the novel coronavirus pneumonia (2019-nCoV): a study of 63 patients in Wuhan, China 2020 *European radiology* 30:63306--3309
23. Choi S, Lee J, Kang MG, Min H, Chang YS, Yoon S., Large-scale machine learning of media outlets for understanding public reactions to nationwide viral infection outbreaks 2017 *Methods* 129:50--59
24. Li, H. C., Yang, G., Yang, W., Du, Q., & Emery, W. J. (2020). Deep nonsmooth nonnegative matrix factorization network with semi-supervised learning for SAR image change detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 160, 167--179.
25. Alafif T, Tehame AM, Bajaba S, Barnawi A, Zia S. Machine and deep learning towards COVID-19 diagnosis and treatment: survey, challenges, and future directions 2021 *International Journal of environmental research and Public health* 18:1117
26. Ebaa Fayyumi, Sahar Idwan, Heba AboShindi, Machine learning and statistical modeling for prediction of novel COVID-19 patients case study: Jordan 2020 *International Journal of Advanced Computer Science and Applications* 1:15
27. Gurjit S. Randhawa, Maximillian P.M. Soltysiak, Hadi El Roz, Camila P.E. de Souza, Kathleen A. Hill, and Lila Kari, Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study 2020 *Plos one* 15:4e0232391
28. 1994 *Machine learning, neural and statistical classification* Citeseer
29. Colubri A, Hartley MA, Siakor M, Wolfman V, Felix A, Sesay T, Shaffer JG, Garry RF, Grant DS, Levine AC, Sabeti PC, Machine-learning prognostic models from the 2014--16 Ebola outbreak: data-harmonization challenges, validation strategies, and mHealth applications 2019 *Clinical Medicine* 1:54--64
30. Wenlong Zhang, Xiaolin Zeng, Wuchao Wang, Multi-exposure image fusion based on wavelet transform 2018 *International Journal of Advanced Robotic Systems* 15:1729881418768939
31. Fang X., Jinghong L., Yueming S., Hui S., Xuan W. , Multi-exposure image fusion techniques: a comprehensive review 2022 *Remote Sensing* 14:771
32. Durqa P.K., Razwan P., Manikandan R., Amir H. , Partial derivative nonlinear global pandemic machine learning prediction of covid 19 2020 *Chaos, Solitons & Fractals* 139:110056
33. Shreshth Tuli, Shikhar Tuli Rakesh Tuli, Sukhpal Singh Gill, Predicting the growth and trend of the COVID-19 pandemic using machine learning and cloud computing 2020 *Internet of Things* 10:222
34. Shortliffe 1976 *Books: Computer-Based Medical Consultations: MYCIN* *Journal of Clinical Engineering* 1:169
35. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LS. The universal protein resource (UniProt) 2005 *Nucleic acids research* 33:suppl_1:D154--D159
36. Gonzalo N., Rafael B., Ricardo G., Two-step learning of fuzzy cognitive maps for prediction and knowledge discovery on the HIV-1 drug resistance 2014 *Expert Systems with Applications* 41:821--830

37. Wang Y, Zhao Y, Therneau TM, Atkinson EJ, Tafti AP, Zhang N, Amin S, Limper AH, Khosla S, Liu H., Unsupervised machine learning for the discovery of latent disease clusters and patient subgroups using electronic health records 2020 *Journal of biomedical informatics* 102:103364
38. He Y, Xiang Z, Mobley HL., Vaxign: the first web-based vaccine design program for reverse vaccinology and applications for vaccine development 2010 *Journal of Biomedicine and Biotechnology* 2010
39. Ong E, Wang H, Wong MU, Seetharaman M, Valdez N, He Y.. Vaxign-ML: supervised machine learning reverse vaccinology model for improved prediction of bacterial protective antigens 2020 *Bioinformatics* 36:103185—3191.
40. ZAINAB ALZAMILI , KASSEM M. DANACH, MONDHER FRIKHA , Deep Learning-Based Patch-Wise Illumination Estimation for Enhanced Multi-Exposure Fusion, 2 November 2023, 2023.3328579