Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 09 Issue 02
2025

Research Article
https://doi.org/10.56979/902/2025

Application of Data Mining Techniques to Prognosticate COVID-19 Proliferation

Ashar Atique[1], Muhammad Munwar Iqbal[1], Muhammad Idrees[2], Qamas Gul Khan[1], and Farooq Ali[1]

[1]Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan.
[2]Department of Computer Science and Engineering, University of Engineering and Technology Lahore, Narowal Campus, Pakistan.
Corresponding Author: Muhammad Munwar Iqbal. Email: munwariq@gmail.com

Received: June 03, 2025 **Accepted:** August 24, 2025

Abstract: The emergence of COVID-19 in early 2020 rapidly transformed into one of the most serious global health concerns. First reported in Wuhan, China, the virus quickly crossed national borders and spread worldwide. Initial signs of infection, such as fever, cough, and general weakness, often appear mild, yet in many cases the illness progresses to severe complications, including lung impairment, organ dysfunction, or even death. For diagnosis, Reverse Transcription Polymerase Chain Reaction (RT-PCR) continues to be regarded as the benchmark method. Although reliable, this test is costly and often requires up to three days before results are available, which limits its practicality for mass testing during a pandemic. This limitation has created an urgent demand for diagnostic methods that are quicker, more affordable, and equally accurate. Detecting the virus at an early stage is essential, as it not only improves patient recovery but also plays a critical role in slowing transmission within communities. In response to this challenge, the present research applies a customized Convolutional Neural Network (CNN)–based deep learning model to chest X-ray images for COVID-19 detection. The system was designed for multi-class classification and tested using an online dataset. The evaluation results indicate that the model achieved a classification accuracy of 98.87%, highlighting its effectiveness in supporting rapid and reliable COVID-19 screening.

Keywords: CNN Classification; Secretions; Diagnosis; PCR-RT; Covid; Mutations

1. Introduction

Over the past few decades, the incidence and spread of infectious and viral diseases have risen sharply, posing significant challenges to healthcare systems worldwide. One of the most notable examples is the Coronavirus (COVID-19) pandemic, which was first identified in Wuhan, China, in late 2019. Within a short time, it expanded into a global crisis, exerting pressure on healthcare, economic stability, and social structures across nations. COVID-19, caused by the novel SARS-CoV-2 virus, manifests through a broad spectrum of symptoms. While many patients experience mild signs such as fever, sore throat, and cough, severe cases may progress to acute respiratory distress syndrome (ARDS), multi-organ failure, or even death. These outcomes are particularly common among the elderly and those with weakened immune systems. The average duration from infection to death is estimated at approximately 14 days; however, this period varies with factors like age, pre-existing medical conditions, and immune response [1]. The unpredictable nature of viral mutations and their rapid transmission underline the urgency of accurate and timely diagnostic systems to limit further spread and provide prompt treatment.

Although Reverse Transcription Polymerase Chain Reaction (RT-PCR) remains the benchmark for COVID-19 detection, its limitations high cost, dependency on specialized laboratories, and result delays of up to 72 hours, make it impractical for widespread screening. These challenges have reinforced the need for faster, scalable, and cost-effective diagnostic alternatives, ideally through non-invasive approaches. In this context, Artificial Intelligence (AI) and Machine Learning (ML), particularly Deep Learning (DL) architectures such as Convolutional Neural Networks (CNNs), have shown significant promise in medical imaging. CNNs can analyze chest X-ray (CXR) images with high precision, allowing automated distinction between COVID-19 and non-COVID-19 cases. This capability reduces the burden on medical professionals while accelerating clinical decisions [2, 3]. In addition, forecasting models like Prophet and ARIMA have been applied to study infection trends and enhance understanding of virus dynamics [4, 5]

Recognizing the growing importance of chest imaging in COVID-19 diagnostics, this study applies a customized CNN-based deep learning framework to classify chest X-rays, with the objective of minimizing testing costs, reducing dependency on laboratory-based approaches, and delivering faster results. The research introduces an AI-driven diagnostic model built upon a CNN architecture enhanced through transfer learning, trained with annotated chest X-ray datasets, to distinguish COVID-19 infections from normal cases with high precision.

The primary contributions of this research are as follows:

- Development of a customized CNN architecture combined with transfer learning for efficient chest X-ray image analysis.
- Utilization of a publicly available Kaggle dataset to ensure reproducibility and wider applicability.
- Incorporation of preprocessing, normalization, and multi-class classification strategies to improve model generalization.
- Achievement of 98.82% accuracy while reducing training time and preventing overfitting.
- Performance benchmarking against existing methods, demonstrating superior outcomes in key evaluation metrics.

2. Literature Review

Scholars have investigated the application of machine learning (ML) and data mining approaches to support the diagnosis, prediction, and management of COVID-19 transmission[6]. Various approaches, including predictive modeling, image-based classification, and time-series analysis, have been proposed to enhance decision-making in healthcare systems under pandemic stress. One of the early predictive models used the PIBA technique, which estimated daily death rates during outbreaks using publicly available COVID-19 data[7]. This model employed mortality data from Wuhan and was adapted to forecast trends in other Chinese cities and Korea, estimating a death rate of 1.6% in early-stage patients[8]. Similarly, the ARIMA model was applied to Italian health data, achieving a 93.75% accuracy in predicting recurrence trends and 84.4% accuracy in modeling recoverable cases.

Numerous types of data mining techniques were experimented on the pandemic. To detect short-term trends, researchers primarily applied statistical time-series models, including ARIMA and SARIMA, in the initial stages of the outbreak because they are easy to apply and can rapidly identify short-term trends. They use previous values of daily or weekly cases to have an idea of what can occur in the future. These methods worked quite effectively in the case of very short predictions but failed in situations when the dynamics of the virus altered abruptly, i.e., when new restrictions were imposed by governments or when new types of the virus emerged. Consequently, researchers resorted to more adaptable machine learning algorithms with the ability to learn nonlinear and complex relationships among two or more variables. Random forests, support vector machines, and gradient boosting techniques gained popularity because it is possible to use numerous input features like population density, temperature, testing rate, and mobility data to make a more accurate prediction. Mathematical techniques have also been used to understand the spread of COVID-19. A tree-based model explored the effects of isolation and undetected transmissions (hidden nodes) in the virus's spread[9]. Results showed that quarantine and lockdown measures significantly curbed transmission, especially in regions like India. Another study used the ARIMA model with the Johns Hopkins dataset to predict the epidemic's trend, while further research explored COVID-19's indirect effects. For example, the virus's effect on pregnancy was analyzed using a meta-analysis approach, focusing on complications such as premature birth, preeclampsia, miscarriage, and neonatal asphyxia[10]. The findings indicated a 90% likelihood of COVID-19-positive mothers developing pneumonia and an increased risk of severe perinatal outcomes. With increased computing and data availability, one of the most popular ways to forecast COVID-19 was deep learning. Neural networks, and in particular recurrent neural networks such as LSTM (Long Short-Memory) and GRU (Gated Recurrent Unit), were popular due to their ability to learn the patterns that change with time. These networks have been used to forecast new cases, hospitalizations, and deaths daily, depending on past data sequences. Convolutional neural network (CNN) models were also adjusted to process spatiotemporal data, whereby the aim was to forecast how the disease diffuses regionally, instead of just through time. These models usually worked better than the conventional statistical techniques in cases where there was sufficient data. They were, however, also more complicated and consumed a lot of computing power. Moreover, since they are not as transparent in their inner mechanisms, they could be more difficult to read than simpler models for the public health officials.

ML has proven effective in diagnosing other infectious diseases, such as swine flu and Clostridium difficile infection (CDI). Neural Networks and Support Vector Machines (SVMs) were applied to expedite swine flu detection, where Neural Networks demonstrated superior accuracy and faster response compared to SVMs[11]. In another study, ML was applied to predict complications in CDI patients using hospital data from the University of Michigan. The model, trained on post-diagnosis data spanning over two years, achieved high confidence in forecasting severe outcomes like ICU admission and mortality[12]. Recent studies also highlighted how regression-based ML techniques can support sustainable pandemic response strategies. Regression algorithms have been used to diagnose COVID-19 with increasing accuracy[13, 14]. However, many studies focused more on predictive outcomes than on interpretability or the visual explanation of virus proliferation [15, 16]. These limitations present an opportunity to develop more robust, explainable, and visually traceable models—gaps that the present research aims to address. Experiments also involved ensemble models, a combination of predictions of multiple varied algorithms to obtain a more stable prediction. The practice of ensemble forecasting became a common practice in most countries, particularly in national and regional COVID-19 forecasting centers. The point is that each model will not be able to model every aspect of a quickly changing pandemic, but that an average between several models can be better calibrated. The strategy minimized the mistakes of model biasing by each individual and made the policymakers have more confidence in the forecast. Ensemble methods were also malleable enough to incorporate data-based and mechanistic models and combine the assets of each. There was a vast array of sources of data that were used to forecast COVID-19. The most frequently used data sets were the official statistics of confirmed cases, hospitalizations, and deaths. Nevertheless, to enable forecasts to be more receptive to shifts in human behavior, researchers started to include mobility information on mobile phones and applications, weather-related factors, including temperature and humidity, as well as government policy indices that monitored lockdowns and mask-wearing requirements. Early warning indicators of outbreaks encompassed the use of social media data, Google search trends, as well as wastewater samples. These extra characteristics enabled the models to capture the early indications of increased infections prior to their manifestations in verified cases. Nevertheless, the handling of such data was a significant challenge, and the challenges were reported gaps in data, delays, and disparities across nations. The data were often more difficult to clean and preprocess than to create the models themselves.

Recent surveys confirm that modern machine-learning methods generally outperform classical statistical models in forecasting COVID-19 trends. Cheng et al.[17]reviewed over 130 COVID-19 prediction studies and found that hybrid approaches—such as combining neural networks with optimization algorithms significantly improved accuracy. Similarly, Nguyen et al.[18] introduced BeCaked, which integrates an SIRD epidemiological model with an LSTM autoencoder, achieving R^2 values above 0.98 for global forecasts. Chen et al. [19] proposed a hybrid BiGRU-attention model, reporting Adjusted $R^2 > 0.99$ on long-term forecasts. Another advancement was the XGBoost-SIRVD-LSTM model, which used XGBoost for feature selection and an LSTM within a compartmental SIRVD structure, outperforming baseline models on R^2, RMSE, and MAPE [20]. The most common metrics of model performance, root mean square error (RMSE) or mean absolute error (MAE), were often employed to compare the number of cases predicted and the number of cases actually observed. Generally, the research determined that there was no single method that was always yielding optimal predictions. Simple ARIMA models might be good to use in a short-term forecasting period, particularly when the trend is smooth. The deep learning models and hybrid methods produced better results where large and varied datasets were present. Ensemble techniques were more likely to provide the best predictions across various time and space. Nonetheless, it was not easy to compare the studies directly as each one of them employed slightly different sources of data, timeframes, and methods of validation. Literature reviews made the conclusion that machine learning methods had massive potential, but in many cases, their performance difference to traditional models was not so enormous or stable. There were a number of innovations as models got advanced with time. Transfer learning came in handy in cases of smaller regions in which the training data was scarce. Researchers could train a model on global data and fine-tune it locally without having to collect enormous local datasets by pretraining it on global data and fine-tuning it on a particular country or city. Attention mechanisms, which were initially created in the field of natural language processing, were also implemented to predict COVID-19 time series. These enabled models to specialize in the most interesting periods or characteristics during which predictions should be made, which increases the precision in the areas of sudden changes, like the outbreak of new variants. Another significant trend was probabilistic forecasting, where forecasts were made in the form of ranges of uncertainty, rather than just points. The method was particularly useful to policy-makers who had to prepare both the best-case and worst-case scenarios instead of one number. Nevertheless, COVID-19 remained highly hard to predict despite the advancements. The biggest challenge was the nonstationary fact that the underlying conditions continued to vary as people became accustomed to limitations or new variants came up. Models that were trained on previous layers tended to do a poor job on subsequent layers. The other issue was data quality. The reporting systems were not regular in many countries, with backlogs, absence of weekend data, or alteration of test policy. This rendered it difficult to create stable datasets to train the model. Other models were also inflated by the overfitting behavior, in which they were highly effective on historical data but not on new patterns. The complex machine learning models could be helpful at observing any useful pattern, though they seldom gave a causal explanation of what was causing the patterns, and it would be a risky step to rely on the predictions, regardless of how useful, to form the foundation of a social policy.

Beyond hybrid architectures, innovations in input data and training strategies have improved accuracy. Hu et al.[21] developed an attention-enhanced transfer learning LSTM (TLLA) that consistently reduced MAE and RMSE compared to traditional LSTMs. Jiao et al.[22] Incorporated human mobility data into an LSTM-attention model for Japan, significantly reducing long-term forecasting errors. These findings demonstrate that combining epidemiological knowledge, auxiliary datasets, and deep learning improves not only predictive accuracy but also model interpretability. Another level of complexity was brought about by spatial heterogeneity. The population density, as well as the infrastructure and mobility patterns in a particular region, are significantly different, so that the model trained on data in one of the countries cannot be generalized to another without relevant modifications. To overcome this problem, other research adopted hierarchical or multi-scale modeling models that related localized predictions to national trends. The other ones built up groupings that integrated parallel regional models. However, these methods demanded quality local data, which is not always available in detail. The empirical case studies reveal that data mining is useful in operational forecasting. Government agencies and research consortia created national forecast hubs, which take predictions of various models and provide real-time comparative analyses. Hospitals used machine-learning predictions to predict spikes in the patient inflow and to distribute the limited resources like ventilators and ICU beds. Telecommunications-based mobility data helped in identifying potential hotspots before the confirmed cases increased in number and subsequently provided the basis of appropriate interventions. These case studies serve to point out that data-mining services can be used to greatest effect when coupled with human insight and domain knowledge, but not as independent higher-order automation. Ethical and social issues have been simultaneously brought up by the increased data-driven forecasting of COVID-19. The process of mobilizing mobility and health data is often associated with personal sensitive information, increasing the risks of privacy invasions. The strong anonymization protocols and the responsible use of data are one of the primary priorities. There are also problems of equity and bias because the models that are trained on unfinished or biased datasets can generate inefficient predictions of specific communities, especially when they have lower testing rates or limited technological access. Openness about the restrictions of the model used and meticulous information conveyed to the populace about the uncertainties is fundamental to sustaining trust in the forecasting systems. The future research directions are focused on realizing an alternative hybrid model that will combine both the causal inference of epidemiology and the pattern-recognition abilities of machine learning. By developing the capability to build adaptive or continuous learning systems that can automatically revise with the incoming data, the forecast fidelity will improve when the conditions change rapidly.

In summary, prior literature demonstrates the growing role of AI and data-driven models in pandemic management. However, limitations such as class imbalance, lack of real-time adaptability, and underperformance in diverse imaging conditions remain. Our work addresses these gaps by proposing a

CNN-based classification framework using chest X-ray images, evaluated with high-performance metrics on a large and diverse dataset.

Table 1. Related Work

Author & Year	Method Used	Dataset Source	Accuracy (%)	Key Limitation
Khan et al.[23]	Channel Boosted CNN	Chest X-rays	97.94	Limited to small image variations
Hira et al.[24]	CNN-Based Auto Model	Chest X-rays	90.80	No F1 or Recall reported
Gunraj et al.[25]	CovidNet-CT (CNN)	CT Scan images	93.10	CT data—not suitable for mass screening
Majeed et al. [9]	Transfer Learning + CNN	Chest X-rays	96.45	Lacked precision in class imbalance
Wang et al.[26]	COVID-Net (CNN)	COVIDx Chest X-ray Dataset	92.4	The model is biased toward the majority class
Apostolopoulos et al.[27]	MobileNet v2 Transfer Learning	Public X-ray dataset	96.78	Limited data variety
Hemdan et al.[28]	COVIDX-Net (7 CNN models)	COVID-19 X-ray Dataset	90.00	Performance varies significantly by model

3. Proposed Methodology

This study presents a CNN-based method with transfer learning for COVID-19 detection and prediction using chest X-ray images. The workflow includes five stages: data acquisition, preprocessing, transfer learning, CNN training, and classification. X-ray images from public datasets, collected in different formats and resolutions, were preprocessed by converting to RGB (if needed), resizing to 180×180, and normalizing pixel values to 0–1. Transfer learning was then applied using a pre-trained CNN to extract features and reduce overfitting, with the model fine-tuned on the COVID-19 dataset.

During training, the CNN—composed of convolutional, pooling, and fully connected layers—learned image features effectively. In the final stage, the trained model classified input X-rays as either COVID-19 positive or normal.

Figure 1. Flow Diagram of the proposed model

3.1. Dataset Description

For this study, the dataset used was from a Kaggle repository and comprised 5,863 chest X-ray images in JPEG format, collected from patients aged five years and older. To maintain reliability in model training, the images were preprocessed by eliminating unreadable, duplicate, or poor-quality scans. The refined dataset was then split into three parts: 70% for training, 15% for validation, and 15% for testing. Each subset was further organized into two categories: COVID-19 and Normal, a strategy consistent with prior studies [29]. Figure 2 illustrates the distribution of X-ray images across these two classes.

3.1.1. Data Preprocessing

To maintain uniformity and enhance model efficiency, several preprocessing steps were carried out: Image Normalization: Pixel values were scaled within the 0 to 1 range to facilitate smoother convergence during training. All images were fixed to a resolution of 180 × 180 pixels to align with the CNN input layer. RGB Conversion – images not originally in RGB were transformed into three-channel RGB format for consistent data representation.

Data Augmentation: operations such as horizontal flips, zoom adjustments, and slight rotations were applied to strengthen the model's generalization and minimize overfitting. These preprocessing measures ensured that the CNN received clean, standardized, and enhanced input data.

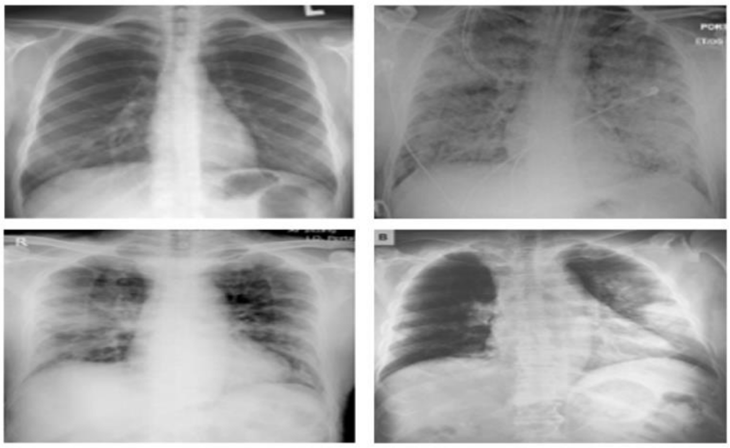

Figure 2. Samples from the Dataset

3.1.2. Transfer Learning

Transfer learning was adopted in this study as an efficient way to reduce training time and improve feature extraction when working with medical imaging data. The approach begins with a pre-trained CNN model, such as VGG16 or ResNet50, that has already been trained on the large-scale ImageNet dataset. These models contain early layers capable of capturing universal visual elements like lines, textures, and basic shapes, which remain useful across many domains. Instead of training a network from scratch, these generalized layers were retained to preserve their feature-detection ability. The higher layers of the model, however, were replaced and a fine-tuned dataset to make the network sensitive to domain-specific patterns. This adjustment allowed the system to adapt existing visual knowledge to the context of medical diagnostics. By leveraging transfer learning, the model not only achieved higher accuracy with relatively limited labeled data but also significantly reduced computational cost and training duration. Moreover, this strategy improves model generalization, resulting being suitable for practical applications where large annotated datasets are often unavailable.

3.1.3. CNN

A Convolutional Neural Network (CNN) is designed for tasks involving image and pattern recognition. Unlike traditional neural networks that primarily rely on fully connected layers and activation functions, CNNs also integrate convolution and pooling layers, which allow them to efficiently capture spatial hierarchies in data. The foundation of CNNs can be traced back to the visual perception studies of K. Fukushima in 1980, which inspired the concept of hierarchical feature extraction. Nearly two decades later, in 1998, Yann LeCun introduced the LeNet architecture, which gained prominence for its effectiveness in handwriting recognition. This marked the beginning of CNNs as a core component of modern computer vision research. The CNN model used in this study is shown in Figure 3.

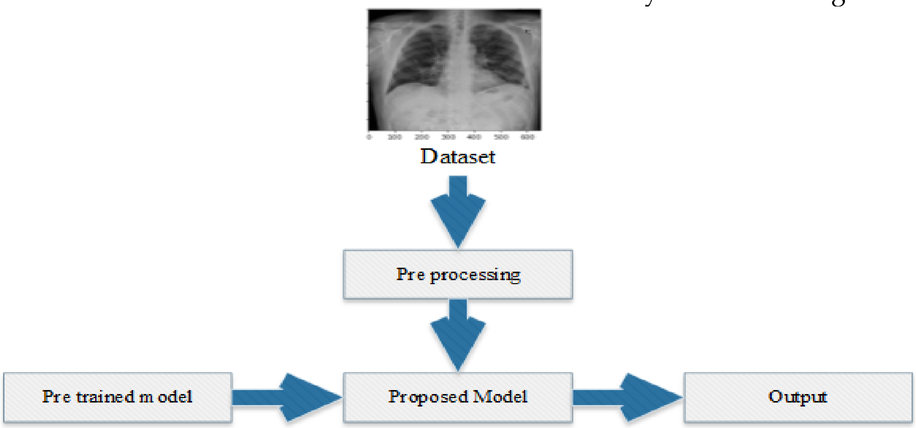

Figure 3. Proposed Model Diagram

The convolutional layer serves as the backbone of a CNN, where the process of feature extraction takes place. In this layer, pixel values from an input image (defined by height and width) are convolved with filters or kernels, resulting in feature maps that typically have reduced dimensions compared to the original input. Several hyperparameters, such as filter size, stride, and padding, must be tuned for optimal performance. To illustrate, convolution operations on 7×7×1 images are depicted in the figures that follow.

Mathematically, the convolution operation can be expressed as in Equation (1):

$$\sum_{i=0}^{i=n} B * \alpha + b \quad (1)$$

Here, B is the input, α the filter, and **b** the bias. The output size after convolution is determined using equation (2).

$$[\frac{n+2i-k}{v} + 1] \quad (2)$$

Here, n is the original image size, **i** corresponds to the padding, k is the kernel dimension, and **v** is the stride value.

Pooling layers are integrated after convolutional operations to progressively reduce the spatial resolution of feature maps. By compressing feature representations, pooling decreases the number of learnable parameters, reduces memory usage, and lowers computational complexity. This process also enhances the model's generalization capability. Key hyperparameters include the pooling kernel size, stride, and padding. The most common pooling strategies are max pooling and average pooling, where max pooling emphasizes the most prominent features while average pooling provides a smoothed aggregation of activations [30].

3.2. Classification

Once training is completed, the model classifies new chest X-ray inputs as either COVID-19 positive or Normal. The model generates probability scores, and a threshold set at 0.5 is applied to assign the final class label.

4. Results and Discussion

The proposed CNN was evaluated on the preprocessed Kaggle chest X-ray dataset, with all inputs resized to 180×180×3. Its architecture started with two convolutional layers and a pooling layer for basic feature extraction, followed by a dropout layer to reduce overfitting. Additional convolutional and pooling layers captured higher-level patterns, after which the output passed through a flattening layer, another dropout, and finally a dense layer for binary classification of COVID-19 and normal cases. ReLU activation, SAME padding, and a consistent 3×3 filter sizes were applied throughout to ensure effective learning while preserving spatial details.

4.1. Evaluation Metric

To thoroughly evaluate the performance of the model, several metrics were applied as given in Equations (3), (4), (5), and (6).

Accuracy: The proportion of correctly classified images over the total number of test images.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

For a comprehensive evaluation of the model's effectiveness, multiple performance metrics were employed:

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

Recall (Sensitivity): The ratio of true positive predictions to the actual positive cases. It reflects the model's ability to detect positive samples.

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

F1 Score: The harmonic average of precision and recall, providing a balanced measure between the two.

$$F1 = 2. \frac{Precision\ .\ Recall}{Recison+Recall} \quad (6)$$

Loss: Binary cross-entropy loss was used to measure the difference between predicted and actual outputs during training.

4.2. Model Performance

The model's performance was assessed over 20 epochs, with accuracy and loss trends shown in Figures 4 and 5. The accuracy curve in Figure 4 indicates rapid convergence, reflecting the benefit of transfer learning in speeding up training and improving results. The loss curve in Figure 5 shows a steady decline, confirming effective learning and error reduction. Sample outputs in Figure 6 further validate the model's ability to distinguish COVID-19 positive cases from normal ones.

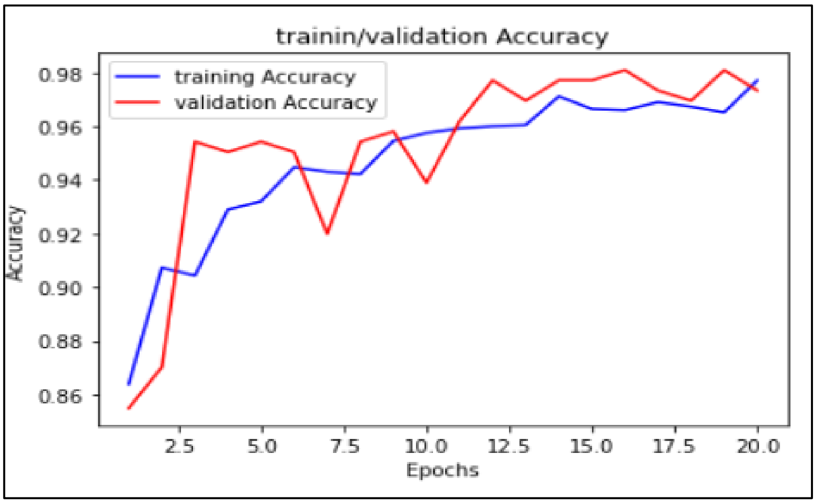

Figure 4. Confusion Matrix of SVM

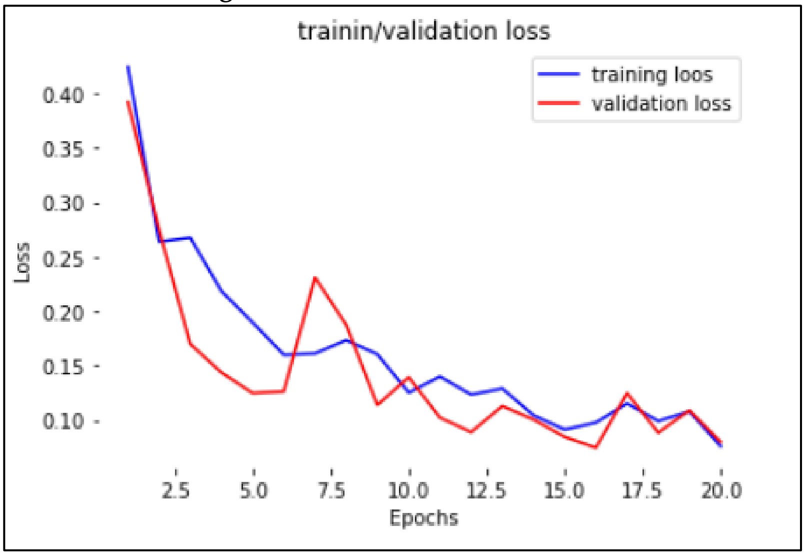

Figure 5. Accuracy of classifiers

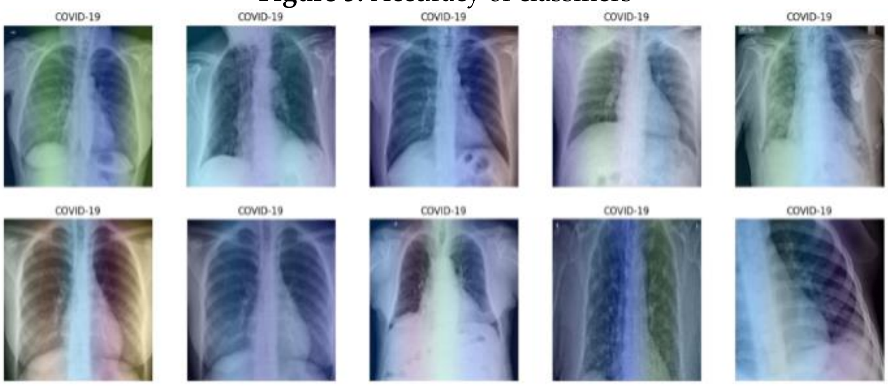

Figure 6. Political Sentiment Analysis

The main crux is that results collectively highlight the robustness of the proposed architecture. The final performance metrics achieved after 20 epochs are summarized in Table 2.

Table 2. Results of the CNN Proposed Model

Epoch	Accuracy	Recall	F1 score	Precision	Loss rate
20	**0.989**	**0.979**	**0.987**	**0.995**	**0.042**

4.3. Comparative Analysis

Table 3 presents a comparison with earlier studies, showing that the proposed CNN with transfer learning achieves superior accuracy, precision, and F1 score, confirming its effectiveness for COVID-19 detection.

Table 3. Result Comparison with Existing Techniques

Author & Year	Method Used	Dataset Source	Accuracy (%)	Key Limitation
Khan et al.[23]	Channel Boosted CNN	Chest X-rays	97.94	Limited to small image variations
Hira et al.[24]	CNN-Based Auto Model	Chest X-rays	90.80	No F1 or Recall reported
Gunraj et al.[25]	CovidNet-CT (CNN)	CT Scan images	93.10	CT data—not suitable for mass screening
Majeed et al.[9]	Transfer Learning + CNN	Chest X-rays	96.45	Lacked precision in class imbalance
Wang et al.[26]	COVID-Net (CNN)	COVIDX Chest X-ray Dataset	92.4	Model biased toward majority class
Apostol Poulos et al.[27]	MobileNet v2 Transfer Learning	Public X-ray dataset	96.78	Limited data variety
Hemdan et al.[28]	COVIDX-Net (7 CNN models)	COVID-19 X-ray Dataset	90.00	Performance varies significantly by model
Proposed Model	Custom CNN + Transfer Learning	Kaggle Chest X-ray Dataset	98.82	Currently limited to binary classes

The incorporation of transfer learning enabled faster convergence and high accuracy with a smaller dataset, making this approach suitable for real-time medical image classification systems where quick and reliable predictions are critical.

5. Conclusion & Future Work

This study aimed to build a cost-effective system for early COVID-19 detection using a CNN-based algorithm with transfer learning. Trained on chest X-ray data, the model classified images as normal or COVID-19 positive with 98.82% accuracy, demonstrating strong diagnostic capability and confirming its effectiveness.

The findings emphasize the promise of deep learning and transfer learning in vital healthcare applications, particularly when data availability is limited. By leveraging X-ray imaging as a low-cost and widely accessible diagnostic option, the proposed method becomes highly suitable for use in resource-constrained regions. Additionally, the framework has potential for integration into rapid screening processes within clinical settings, alleviating reliance on PCR-based tests. For future research, the focus will be on applying and testing additional CNN architectures and AI models to expand the system for diagnosing other viral infections, including flu, SARS, and pneumonia. Another direction will involve investigating post-COVID complications and their detection through longitudinal imaging and predictive modeling.

References

1. Khemasuwan, D., and Colt, H.G.: 'Applications and challenges of AI-based algorithms in the COVID-19 pandemic', BMJ Innovations, 2021, pp. bmjinnov-2020-000648
2. Y. Tao, C.Y., T. Wang, E. Coltey, Y. Jin, Y. Liu, et al.: 'A Survey on Data-Driven COVID-19 and Future Pandemic Management', ACM Computing Surveys (CSUR), 2022
3. A. Srivastava, S.J.: 'Data-driven machine learning: A new approach to process and utilize biomedical data': 'Predictive Modeling in Biomedical Data Mining and Analysis' (Elsevier, 2022), pp. 225–252
4. M. Alazab, A.A., A. Mesleh, A. Abraham, V. Jatana, S. Alhyari: 'COVID-19 prediction and detection using deep learning', International Journal of Computer Information Systems and Industrial Management Applications, 2020, 12, pp. 168–181
5. M. M. A. Monshi, J.P., V. Chung, F. M. Monshi: 'CovidXrayNet: Optimizing data augmentation and CNN hyperparameters for improved COVID-19 detection from CXR', Computers in Biology and Medicine, 2021, 133, pp. 104375
6. P. M. Shah, F.U., D. Shah, A. Gani, C. Maple, Y. Wang, et al.: 'Deep GRU-CNN model for COVID-19 detection from chest X-rays data', IEEE Access, 2021, 10, pp. 35094–35105
7. M. J. Horry, S.C., M. Paul, A. Ulhaq, B. Pradhan, M. Saha, et al.: 'COVID-19 detection through transfer learning using multimodal imaging data', IEEE Access, 2020, 8, pp. 149808–149824
8. N.-u. Rehman, M.S.Z., T. Meraj, H. T. Rauf, R. Damaševičius, A. M. El-Sherbeeny, et al.: 'A self-activated CNN approach for multi-class chest-related COVID-19 detection', Applied Sciences, 2021, 11, (19), pp. 9023
9. T. Majeed, R.R., D. Ali, A. Asaad: 'Covid-19 detection using CNN transfer learning from X-ray images', medRxiv, 2020, 12, pp. 2020
10. M. Polsinelli, L.C., G. Placidi: 'A light CNN for detecting COVID-19 from CT scans of the chest', Pattern Recognition Letters, 2020, 140, pp. 95-100
11. M. Z. Islam, M.M.I., A. Asraf: 'A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images', Informatics in Medicine Unlocked, 2020, 20, pp. 100412
12. R. Karthik, R.M., M. Hariharan: 'Learning distinctive filters for COVID-19 detection from chest X-ray using shuffled residual CNN', Applied Soft Computing, 2021, 99, pp. 106744
13. X. Yu, S.L., L. Guo, S.-H. Wang, Y.-D. Zhang: 'ResGNet-C: A graph convolutional neural network for detection of COVID-19', Neurocomputing, 2021, 452, pp. 592–605
14. A. M. Ismael, A.Ş.: 'Deep learning approaches for COVID-19 detection based on chest X-ray images', Expert Systems with Applications, 2021, 164, pp. 114054
15. A. Kumar, A.R.T., S. C. Satapathy, Y.-D. Zhang: 'SARS-Net: COVID-19 detection from chest x-rays by combining graph convolutional network and convolutional neural network', Pattern Recognition, 2022, 122, pp. 108255
16. R. Kundu, P.K.S., S. Mirjalili, R. Sarkar: 'COVID-19 detection from lung CT-Scans using a fuzzy integral-based CNN ensemble', Computers in Biology and Medicine, 2021, 138, pp. 104895
17. Y. Cheng, L.L., M. W. Mutiso, S. A. Vughs, and S. Chen: 'Survey of COVID-19 forecasting models based on machine learning methods', Expert Systems with Applications, 2025, 260, pp. 125330
18. T. T. Nguyen, Q.V.H.N., D. N. Nguyen, N. D. Nguyen, and J. Campbell: 'COVID-19 forecasting using an improved SIRD model and machine learning', Expert Systems with Applications, 2021, 186, pp. 115716
19. H. Chen, S.Z., J. He, Y. Zhang, and J. Gu: 'Hybrid COVID-19 forecasting model integrating epidemiological features and BiGRU with self-attention', Expert Systems with Applications, 202, 236, pp. 121244
20. R. Jiao, M.S., Y. Kawamura, and S. Araki: 'Spatio-temporal forecasting of COVID-19 cases and hospitalizations using LSTM with human mobility data', Scientific Reports, 2025, 15, (1), pp. 6871

21. X. Hu, Z.L., Y. Zhang, and H. Yang: 'A transfer learning and attention-based LSTM model for COVID-19 epidemic forecasting', Applied Soft Computing, 2025, 156, pp. 111540
22. A. A. Alkhalefah, M.J.H., M. Alazab, and S. Rubaiee: 'XGBoost-SIRVD-LSTM: A hybrid deep learning approach for COVID-19 outbreak prediction', Mathematics, 2024, 12, (14), pp. 2155
23. S. H. Khan, A.S., A. Khan, Y.-S. Lee: 'COVID-19 detection in chest X-ray images using a new channel boosted CNN', Diagnostics, 2022, 12, (2), pp. 267
24. S. Hira, A.B., S. Hira: 'An automatic approach based on CNN architecture to detect Covid-19 disease from chest X-ray images', Applied Intelligence, 2021, 51, pp. 2864–2889
25. H. Gunraj, L.W., A. Wong: 'Covidnet-ct: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest CT images', Frontiers in Medicine, 2020, 7, pp. 608525
26. L. Wang, Z.Q.L., A. Wong: 'COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images', Scientific Reports, 2020, 10, pp. 19549
27. I. D. Apostolopoulos, T.A.M.: 'COVID-19: Automatic Detection from X-Ray Images Utilizing Transfer Learning with Convolutional Neural Networks', Physical and Engineering Sciences in Medicine, 2020, 43, pp. 635–640
28. E. E. D. Hemdan, M.A.S., M. E. Karar: 'COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-ray Images', arXiv, 2020, arXiv:2003.11055
29. T. Ozturk, M.T., E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya: 'Automated detection of COVID-19 cases using deep neural networks with X-ray images', Computers in Biology and Medicine, 2020, 121, pp. 103792
30. A. Hussain, M.A.K., A. Tariq, M. Alhaisoni, R. S. Alnumay, and S. Kadry: 'A lightweight CNN model for efficient COVID-19 detection using chest X-ray images', Computers in Biology and Medicine, 2023, 152, pp. 106408