

Covid19 CT Image Classification Based on CNN Model

Chunlin Zhang

Northwest University, Chang'an District, Xi'an City, Shaanxi Province, China

2088729895@qq.com

Abstract. The COVID-19 pandemic, caused by the highly infectious SARS-CoV-2 virus, has underscored the urgent need for rapid and accurate diagnostic tools to alleviate healthcare resource constraints. This study presents a deep recognition model based on Convolutional Neural Networks (CNNs) for the classification of COVID-19 images. Utilizing a database of chest X-ray images, the model was trained and validated, achieving a classification accuracy of 89% on the test set. Our findings demonstrate the effectiveness of CNNs in extracting complex features, enhancing diagnostic efficiency, and reducing misdiagnoses and missed diagnoses in COVID-19 cases. Through comprehensive analysis of training and validation accuracy and loss curves, the model shows strong generalization capability without significant overfitting, aided by early stopping mechanisms.

Keywords: COVID-19, medical image, image classification, CNN, deep learning.

1. Introduction

COVID-19, caused by the highly infectious SARS-CoV-2 virus, has posed an unprecedented challenge to global public health [1]. The pandemic led to a surge in infection cases, resulting in severe complications for many patients and, in numerous cases, tragically, loss of life. This crisis highlighted the fragility of global public health systems, particularly in the areas of epidemic monitoring and early warning systems. Inadequate and delayed responses in many countries contributed to the rapid spread of the virus, underscoring the need for improved pandemic preparedness.

During the pandemic, healthcare systems worldwide faced critical resource shortages. Additionally, weak public health infrastructure, combined with limited laboratory testing capacity and delayed vaccine development, further exacerbated the spread of the virus. Against this backdrop, rapid and efficient diagnostic tools became essential. Emerging technologies such as remote diagnostic systems, automated testing equipment, and AI-assisted medical imaging can significantly enhance diagnostic efficiency, reduce patient waiting times, and alleviate pressure on healthcare resources.

Artificial intelligence (AI) in medical imaging has increasingly gained attention as a promising tool. Through an in-depth examination of AI's applications and advancements in medical imaging diagnosis, this study aims to provide stronger support for addressing potential future public health crises. Computer vision, which explores how to equip machines with visual perception capabilities, utilizes cameras and computational devices to replicate human vision in tasks such as object recognition, tracking, and measurement. Through techniques like image processing, feature extraction, and machine learning, AI systems can be developed to extract valuable information from images or multidimensional data. The close relationship between computer vision and medical image analysis reveals a broad range of

applications, including image segmentation [2], registration, reconstruction, and 3D visualization. These techniques offer clinicians clearer and more precise diagnostic information, ultimately enhancing the accuracy and efficiency of clinical diagnoses [3].

However, traditional medical imaging classification methods have significant limitations. First, they face constraints in feature extraction. Traditional methods often rely on manual feature extraction, requiring specialists to select image features such as edges and textures based on their expertise. This process is time-consuming and subject to human judgment, which may lead to the omission or inaccuracy of critical features. Moreover, manually extracted features may fail to comprehensively represent the complex information in medical images, particularly when handling high-dimensional data, which significantly reduces classification accuracy. Second, traditional models often lack generalization ability. These methods are typically trained on specific datasets, causing them to perform poorly when applied to images from different hospitals or obtained using different equipment. This overfitting limits the effectiveness of traditional methods on new, unseen samples, hindering their clinical utility. Third, traditional approaches often have low computational efficiency. Many of these methods exhibit high computational complexity when processing large-scale medical imaging data, especially when analyzing a large number of images individually. This increases diagnostic time, making real-time processing and rapid feedback challenging, which can delay clinical decision-making.

In response to these challenges, this study leverages deep learning techniques to construct a deep recognition model based on convolutional neural networks (CNNs) [4]. The model achieved an accuracy of 89% on the test set, demonstrating its effectiveness and reliability in COVID-19 CT image classification. Our findings also confirm the potential of deep learning methods in medical image analysis, showing that they can effectively extract and learn complex features, thereby improving classification performance. This study provides clinicians with a robust auxiliary tool, offering a valuable reference for the rapid and accurate diagnosis of COVID-19 cases. Furthermore, it lays the groundwork for applying deep learning methods to image classification tasks for other diseases in the future.

2. Previous works

In the early stages, medical image classification primarily relied on traditional machine learning methods, typically encompassing three steps: image preprocessing, feature extraction, and classifier design. During the image preprocessing phase, commonly used techniques included filtering and wavelet transformation, which were applied to remove noise from medical images and enhance their quality. Techniques such as histogram equalization and contrast stretching were used to improve the visual quality of medical images, facilitating subsequent analysis. Additionally, standardization procedures, such as adjustments to image size and resolution, enabled the comparison and analysis of image data from various sources. In the feature extraction phase, researchers utilized traditional methods based on texture, shape, and color to extract key information from medical images. These features often describe local or global properties of the images, such as edges, corners, and texture patterns. Finally, in the classifier design phase, classifiers were constructed based on these extracted features, using models such as linear regression, Support Vector Machines (SVM), Random Forest (RF), and XGBoost. These classifiers learned the relationship between features and target classes from the training data, allowing for the classification of new medical images.

Linear regression is a fundamental machine learning algorithm, primarily used to predict the linear relationship between independent and dependent variables [5]. Its advantages lie in its simplicity, ease of implementation, adaptability, and interpretability. As a foundation for more complex algorithms, such as ridge regression and lasso regression, it aids in understanding more intricate models. However, linear regression also has notable limitations. It assumes a linear relationship between variables, so it performs poorly when the actual relationship is nonlinear. Additionally, linear regression is sensitive to outliers, and when high multicollinearity exists among independent variables, parameter estimates may become unstable, making the predictions unreliable.

SVM are widely used for classification and regression tasks [6]. SVM's strengths include strong generalization abilities and suitability for small sample sizes. By using kernel functions, SVM can transform nonlinear problems into linear ones, effectively avoiding overfitting. However, SVM has high computational complexity, especially when handling large datasets, as the training process requires multiple iterations, increasing computational demands. SVM is also sensitive to parameter selection, which, if not optimized, can lead to suboptimal classification results. Moreover, SVM is sensitive to data scaling and noise, and unnormalized data may lead to biased classification outcomes.

RF is an ensemble learning method that improves predictive performance by constructing multiple decision trees and aggregating their predictions [7]. Its advantages include high accuracy, strong capability to handle large datasets and high-dimensional data, and good noise resistance. Random Forest can assess feature importance, helping to identify key information in the data [8]. However, RF can be prone to overfitting, particularly when the number of trees is too large. Additionally, the model lacks interpretability and requires significant computational resources, especially for large-scale predictions.

XGBoost is a decision tree-based ensemble learning algorithm widely applied for its high accuracy and computational efficiency [9]. XGBoost's advantages include robustness and strong interpretability, allowing it to handle various data types flexibly. However, parameter tuning for XGBoost is relatively complex, and it is prone to overfitting. XGBoost is also sensitive to outliers and requires substantial computational resources.

While these traditional machine learning algorithms have achieved notable results in medical image classification, their inherent limitations must be acknowledged. The most prominent challenges include limited model complexity and difficulty in effectively capturing spatial features. Traditional machine learning models often rely on fixed mathematical frameworks, which restrict their performance when handling complex data distributions and make it challenging to capture deep-level patterns. Moreover, these algorithms usually depend on manually designed feature extraction methods, which are time-consuming and often fail to fully capture the spatial structure of the data. To address these challenges, deep learning models have shown unique advantages, as they can automatically learn complex, deep representations of data, significantly improving performance in medical image classification tasks.

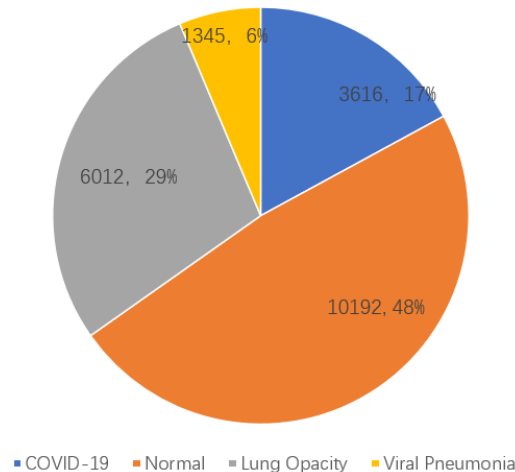


Figure 1. Data distribution.

3. Dataset and data processing

We utilized a chest X-ray image database constructed with data from institutions such as Qatar University in Doha, Qatar, and the University of Dhaka in Dhaka, Bangladesh [10], [11]. This database was designed to enable comparative analysis of imaging characteristics among COVID-19 positive cases, normal lung states, and other types of viral pneumonia. The dataset within this database was released in stages. In the initial release, the database contained 219 COVID-19 positive cases, 1,341 normal lung cases, and 1,345 viral pneumonia chest X-ray images. With the first data update, the number of CXR

images for COVID-19 positive cases increased to 1,200. By the second update, the database was further expanded, including 3,616 COVID-19 positive cases, 10,192 normal lung cases, 6,012 cases of non-COVID-19 lung opacity, while maintaining the original 1,345 viral pneumonia images. Additionally, the database provides corresponding segmented lung images, referred to as "lung masks."

During the data preprocessing phase, we first divided the entire dataset into training and validation sets to ensure effective learning and evaluation of the model. Specifically, 80% of the data was allocated for training, and 20% for validation. Furthermore, to enhance the model's generalization capability, we applied standardization and data augmentation techniques to the images, including rotation, scaling, and flipping. These steps aimed to improve the model's adaptability to different clinical scenarios and enhance its reliability in practical applications. Such preprocessing ensured that the model was trained on diverse data, thereby improving its performance and accuracy.

4. Model

Table 1. Model information.

Layer	Output size	Kernel size	Activation function	Number of parameters
inputlayer	(None,128,128,3)	-	-	0
Conv2D	(None,126,126,32)	1	ReLU	896
MaxPooling	(None, 63, 63,32)	2	-	0
Conv2D	(None, 61,61,64)	1	ReLU	18496
MaxPooling	(None, 30,30,64)	2	-	0
Conv2D	(None, 28, 28,128)	1	ReLU	73856
MaxPooling	(None, 14, 14, 128)	2	-	0
Flatten	(None, 25088)	-	-	0
Dense	(None, 256)	-	ReLU	6422784
Dense	(None, 4)	-	Softmax	1028

In this study, a deep recognition model based on CNN was constructed, as shown in Table 1. The architecture comprises multiple layers, each responsible for specific data processing tasks. The input layer has a data shape of (128, 128, 3), where "None" represents the batch size, which can be dynamically adjusted as needed. The dimensions 128x128 indicate the image height and width in pixels, while 3 denotes the number of color channels, corresponding to RGB.

The convolutional layers in the model consist of three stages. The first convolutional layer has 32 filters of size 3x3, which are used to extract features from the input image. This layer produces an output with the shape (126, 126, 32). The second convolutional layer applies 64 filters of size 3x3 for further feature extraction, resulting in an output shape of (61, 61, 64). The third convolutional layer has 128 filters of size 3x3, aimed at extracting deeper features, with an output shape of (28, 28, 128). Each convolutional layer is followed by a MaxPooling2D layer, which reduces the feature map size, lowering computational complexity while retaining essential information.

The flatten layer converts the multidimensional feature maps into a one-dimensional vector with an output shape of (25088), preparing the data for subsequent fully connected layers. In the fully connected layers, the first layer contains 256 neurons and processes the output from the flatten layer, with an output shape of (256). The final output layer has 4 neurons responsible for the classification task, with an output shape of (4), representing the predicted probabilities for each of the four classes.

During model training, the cross-entropy loss function is used to measure the difference between predictions and true labels. The optimizer chosen is Adam, which combines the advantages of adaptive learning rate and momentum, accelerating convergence and improving model performance. The hyperparameters are set with 37 epochs to ensure sufficient learning on the training data, and a batch size of 128 to balance computational efficiency with model performance.

5. Results

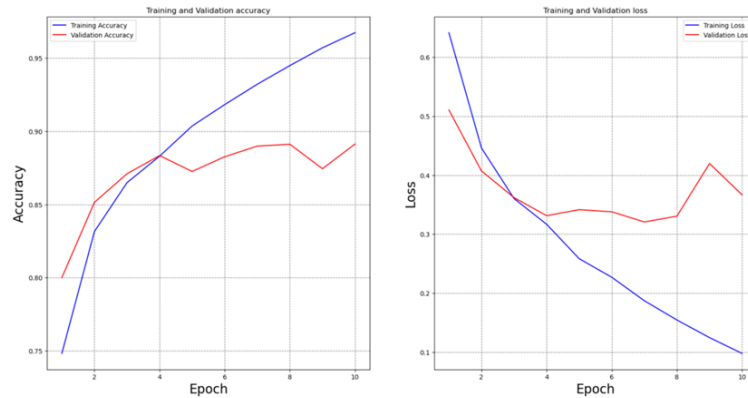


Figure 2. Accuracy and loss curve.

Figure 2 presents the accuracy and loss curves for both the training and validation sets, providing a comprehensive evaluation of the model's robustness and generalization performance. The results highlight the effectiveness of using a CNN architecture, a key innovation in this work, which enables the model to extract and learn spatially hierarchical features from the data. The accuracy curve (left panel) shows a steady increase in training accuracy (blue), reaching approximately 98% by the 10th epoch. The validation accuracy (red) rises rapidly in the early epochs and stabilizes around 90%, demonstrating strong generalization capability. This suggests that the CNN effectively captures discriminative features from the training set while avoiding overfitting on the validation set. The ability of CNNs to automatically learn and refine feature representations through convolutional layers is a major factor contributing to these results. The loss curves (right panel) reveal a similar pattern. The training loss (blue) steadily decreases, indicating consistent error reduction, while the validation loss (red) decreases initially and then stabilizes without any significant increase. This stability is further evidence of the model's ability to generalize well to unseen data. The early stopping mechanism plays a crucial role here, halting training once the validation loss plateaus and preventing overfitting.

Incorporating CNNs into this framework is a pivotal innovation that leverages their strength in feature extraction and hierarchical representation learning, enabling the model to achieve high performance on complex datasets. The spatially invariant properties of CNNs further enhance their robustness, particularly when dealing with diverse data distributions. The model not only demonstrates robust learning and generalization across training and validation sets but also showcases the advantages of CNN-based architectures. The combination of CNNs and early stopping ensures optimal performance, reliability, and scalability. Future work could build on this innovation to enhance adaptability to larger and more complex datasets, further unlocking the potential of CNN-based approaches.

6. Conclusion

This study constructed a classification model for COVID-19 images based on a CNN model. By analyzing accuracy and loss curves for both training and validation sets, we validated the effectiveness of CNN in feature extraction, improving diagnostic efficiency, and reducing misdiagnoses and missed diagnoses. The results indicate that as training iterations increase, the model's training accuracy gradually improves, while the validation accuracy stabilizes. This demonstrates that the model achieves strong classification performance in COVID-19 image classification, providing a solid foundation for future applications.

However, certain limitations remain in this study. First, the COVID-19 imaging dataset is relatively small and lacks sample diversity. Insufficient data size and sample diversity directly impact the model's generalization ability, potentially reducing its adaptability when processing COVID-19 cases with varied characteristics. Furthermore, due to data limitations, the model may struggle to comprehensively

cover image features for patients across different disease stages, genders, and age groups in practical applications. These limitations suggest that future research should prioritize expanding and diversifying datasets to enhance the model's applicability and robustness.

For model improvements, in addition to dataset expansion and data augmentation, further optimization can be achieved in several areas. First, multimodal data integration represents a crucial future research direction [12]. Single-image information may not fully capture the pathological features of COVID-19. Therefore, combining clinical data (such as age, gender, disease stage, and laboratory test indicators) with other imaging data (e.g., CT and MRI) to create a multimodal diagnostic model may offer a more comprehensive approach. This multi-source data fusion could enhance the model's adaptability to diverse cases, improving diagnostic accuracy and robustness. Second, self-supervised learning and transfer learning demonstrate significant potential [13], especially when data samples are limited. Through self-supervised learning, the model can pre-train on unlabeled data, learning general feature representations and reducing dependency on large labeled datasets. Transfer learning, on the other hand, allows knowledge from large-scale pulmonary imaging datasets to be transferred to the COVID-19 classification task, enhancing performance even with small samples. These techniques are particularly beneficial for improving model generalization in data-scarce scenarios. Third, model interpretability is another critical direction for future development. Deep learning models, particularly CNNs, often operate as "black boxes," making their decision-making processes opaque. By incorporating model interpretability techniques, such as Grad-CAM [14] or class activation mapping [15], we can visually highlight the model's focus areas within images, aiding physicians in understanding the model's basis for decisions. This not only enhances the model's credibility but also provides additional diagnostic support for physicians, further optimizing the diagnostic workflow. Additionally, exploring lightweight models and edge computing is valuable. For CNN models to be applicable in real clinical environments, especially in resource-limited settings, computational efficiency and deployment costs must be considered. Techniques such as model pruning, quantization, and knowledge distillation can reduce computational requirements, enabling efficient operation on mobile or edge devices. This could facilitate the use of CNN models in primary healthcare facilities, providing real-time, convenient diagnostic support to a broader patient population. Finally, with advances in medical AI, the integration of personalized medicine and intelligent diagnostic systems is an emerging trend. In the future, the COVID-19 diagnostic model could be integrated with other diagnostic systems, forming an end-to-end intelligent diagnostic system that automates the entire process from data acquisition and image processing to diagnostic recommendations and treatment planning. Such a system would provide personalized diagnostic recommendations based on individual patient characteristics and disease progression, further improving diagnostic accuracy and treatment outcomes.

In summary, this study validates the potential of CNNs in COVID-19 image classification. With continuous expansion of datasets, optimization of model structures, and integration of multimodal technologies, CNNs hold promising applications in COVID-19 diagnosis. In practical applications, CNN models can serve as auxiliary diagnostic tools for physicians, automatically analyzing and classifying image data to provide comprehensive and accurate diagnostic information. As technology advances, CNN models will gradually integrate with other medical technologies, such as gene sequencing and biomarker detection, forming a more comprehensive and precise diagnostic system. This integration will not only improve early diagnosis rates and treatment outcomes for COVID-19 but also help reduce mortality and complication rates, making a greater contribution to global public health.

References

- [1] C. Bao, X. Liu, H. Zhang, Y. Li, and J. Liu, "Coronavirus disease 2019 (COVID-19) CT findings: A systematic review and meta-analysis," *Journal of the American College of Radiology*, vol. 17, no. 6, pp. 701–709, Jun. 2020, doi: 10.1016/j.jacr.2020.03.006.
- [2] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI*

- 2015, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.
- [3] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, “A review of medical image data augmentation techniques for deep learning applications, ” *Journal of Medical Imaging and Radiation Oncology*, vol. 65, no. 5, pp. 545–563, 2021, doi: 10.1111/1754-9485.13261.
- [4] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, “A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects, ” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 6999–7019, Dec. 2022, doi: 10.1109/TNNLS.2021.3084827.
- [5] S. A. A. Shah, U. Nadeem, M. Bennamoun, F. Sohel, and R. Togneri, “Efficient image set classification using linear regression based image reconstruction, ” presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 99–108. Accessed: Oct. 31, 2024. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2017_workshops/w6/html/Shah_Efficient_Image_Set_CVPR_2017_paper.html
- [6] S. Y. Chaganti, I. Nanda, K. R. Pandi, T. G. N. R. S. N. Prudhvith, and N. Kumar, “Image classification using SVM and CNN, ” in 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), Mar. 2020, pp. 1–5. doi: 10.1109/ICCSEA49143.2020.9132851.
- [7] A. Bosch, A. Zisserman, and X. Munoz, “Image classification using random forests and ferns, ” in 2007 IEEE 11th International Conference on Computer Vision, Oct. 2007, pp. 1–8. doi: 10.1109/ICCV.2007.4409066.
- [8] A. Parmar, R. Katariya, and V. Patel, “A Review on Random Forest: An Ensemble Classifier, ” in International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018, J. Hemanth, X. Fernando, P. Lafata, and Z. Baig, Eds., Cham: Springer International Publishing, 2019, pp. 758–763. doi: 10.1007/978-3-030-03146-6_86.
- [9] R. U. Bhagwat and B. Uma Shankar, “A novel multilabel classification of remote sensing images using XGBoost, ” in 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), Mar. 2019, pp. 1–5. doi: 10.1109/I2CT45611.2019.9033768.
- [10] T. Rahman et al., “Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images, ” *Computers in Biology and Medicine*, vol. 132, p. 104319, May 2021, doi: 10.1016/j.compbimed.2021.104319.
- [11] M. E. H. Chowdhury et al., “Can AI help in screening viral and COVID-19 pneumonia?, ” *IEEE Access*, vol. 8, pp. 132665–132676, 2020, doi: 10.1109/ACCESS.2020.3010287.
- [12] M. A. Azam et al., “A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics, ” *Computers in Biology and Medicine*, vol. 144, p. 105253, May 2022, doi: 10.1016/j.compbimed.2022.105253.
- [13] K. Weiss, T. M. Khoshgoftaar, and D. Wang, “A survey of transfer learning, ” *J Big Data*, vol. 3, no. 1, p. 9, May 2016, doi: 10.1186/s40537-016-0043-6.
- [14] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, “Grad-CAM: Why did you say that?, ” Jan. 25, 2017, arXiv: arXiv:1611.07450. doi: 10.48550/arXiv.1611.07450.
- [15] H. Jung and Y. Oh, “Towards better explanations of class activation mapping, ” presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 1336–1344. Accessed: Oct. 31, 2024. [Online]. Available: https://openaccess.thecvf.com/content/ICCV2021/html/Jung_Towards_Better_Explanations_of_Class_Activation_Mapping_ICCV_2021_paper.html