Feature selection using regression mutual information deep convolution neuron networks for COVID-19 X-ray image classification

Tongjai Yampaka [a,1], Suteera Vonganansup [a,2,*], Prinda Labcharoenwongs [a,3]

[a] Department of Computer Science, Rajamangala University of Technology Tawan-Ok, Chon Buri, Thailand
[1] tongjai_ya@rmutto.ac.th; [2] suteera_vo@rmutto.ac.th; [3] prinda_lab@rmutto.ac.th
* corresponding author

ARTICLE INFO

Article history
Received March 18, 2022
Revised June 25, 2022
Accepted July 2, 2022
Available online July 31, 2022

Keywords
COVID-19
Medical imaging
Deep neural networks
Regression mutual information
Feature selection

ABSTRACT

Chest radiography (CXR) image is usually required for lung severity assessment. However, chest X-rays in COVID-19 interpretation is required expert radiologists' knowledge. This study aims to improve the COVID-19 X-ray image classification using feature selection technique by the regression mutual information deep convolution neuron networks (RMI Deep-CNNs). The dataset consists of 219 COVID-19, 500 viral pneumonias, and 500 normal chest X-ray images. CXR images were comprehensively pre-trained using DCNNs to extract the very large image features, then, the feature selection could reduce the complexity of a model and reduce the model overfitting. Therefore, the critical features were selected using regression mutual information followed by the fully connected with softmax layer for classification. For the classification of two alternative systems, these networks were compared (ResNet152V2 and InceptionV3). The classification performance for both schemes were 92.21%, 100%, 90% and 91.39%, 100%, 82.50%, respectively. In addition, RMI Deep-CNNs not only improve the accuracy but also reduce trainable features by over 80%. This approach tends to significantly improve the computation time and model accuracy for COVID-19 classification.

1. Introduction

While radiographic pictures (typical CXR) can aid in the early detection of suspected instances [1], [2], they can diagnose with other infectious and inflammatory lung illnesses [3]–[6]. As a result, radiologists find it challenging to differentiate COVID19 from other viral pneumonias. With regard to extracting features from chest X-rays, Deep Convolutional Neural Networks (DCNNs) have been effective. However, DCNNs have numerous drawbacks, including a lack of available datasets and a lengthy processing time. To diagnose pneumonia, Vikash *et al.* [7] presented the idea of transfer learning framework. A better form of the convolutional neural network (CNN), for instance, is the residual neural network (ResNet) model, which prevents distortion as the network becomes deeper and more complicated [8]. Convolutional neural network models have several layers [9], [10].

The numerous pre-trained models have evolved as a result of the pre-trained. For the identification of lung regions and the classification of various forms of pneumonia, Xianghong *et al.* [11] suggested the VGG16 model. COVIDNet was developed by Linda *et al.* [12] to detect COVID19 instances with

an accuracy of 83.5%. Ayrton [13] reported a validation accuracy of 96.2 percent using a short dataset of 339 photos based on ResNet50. The transfer learning approach has been used well in earlier works, although it still involves lots of convolution and maximum pooling processes. To avoid the limitation of complex pre-trained model, this study proposed regression mutual information (RMI) that measure the relationship between features and target class. The general mutual information (MI) is a theoretical metric that can be used to depict relationships between variables, even when those relationships are very non-linear and concealed by highly dimensional data. It is independent of any classifiers. The one with a higher MI between features and target class is more suitable for the classification tasks. Studies on applying MI to enhance DL networks are currently expanding [14]–[18].

When PCR tests suffer some limitations [19], [20], CXR and CT are necessary and readily available even in rather distant areas. A few studies have reported rather promising results for the diagnosis based on CXR imaging [21], [22]. Convolutional neural networks (CNN) architectures for the diagnosis of COVID-19 have proposed by Narin *et al.* [12]. They demonstrated that a pre-trained ResNet50 model achieved an accuracy of 98%. When it was challenging to discriminate between typical pneumonia and COVID-19, Wang *et al.* [23] created COVID-Net to identify CXR images of COVID-19 patients among patients with viral infections, bacterial infections, and healthy individuals. Although a tiny sample size was employed, and no information regarding the method's dependability was provided, COVID-Net managed to attain a PPV of 88.9% and a sensitivity of 80%. Biraja G., *et al.* [21] used the Bayesian technique to CXR-based COVID-19 diagnosis in order to employ uncertainty estimation with intriguing findings. Nevertheless, the samples are insufficient for statistical variability. Our method adds extra COVID-19 samples to existing datasets, followed by a discriminating Normal, Viral pneumonia, and COVID-19, and finally feature selection using regression mutual information. This approach addresses the drawbacks of state-of-the-art methodologies.

Due to the complexity of the general MI, the search technique for adding or removing any feature based on high scores or low scores is often sophisticated. Entropy decreases are measured by mutual information when the target value is present. Mutual information estimators rely on smoothing parameters, the feature selection greedy approach lacks a theoretically supported stopping condition, and the estimation itself is hampered by the estimation's extremely high dimensionality. To address this problem, Regression Mutual Information (RMI) was proposed. In this study, the processes are summarized as follows: First, the experiments show that the transfer learning from ImageNet could be used with other domains with the fine-tuning approach. Fine-tuning is a common technique in transfer learning to perform image classification and recognize classes that they were never trained on when using pre-trained model. Second, the proposed method is an effective model still maintains a high performance when using regression mutual information scheme.

2. Method

2.1. Datasets

The experiment datasets are made up of 219 COVID-19 chest X-ray pictures that were downloaded from Dr. Joseph Cohen's open-source GitHub repository [23]. Additionally, 500 photographs of viral pneumonias and 500 images of normal chest X-rays were chosen from the Kaggle repository "Chest X-Ray Images (Pneumonia)" [24]. Based on previously trained models, all photos in this dataset were scaled to 224x224 and 299x299 pixels. Representative chest X-ray pictures of healthy people, people with viral pneumonia, and those with COVID-19 are shown in Fig. 1.

(a) (b) (c)

Fig. 1. CXR images with differece class (a) normal, (b) viral pneumonia, and (c) COVID-19.

2.2. Experimental setup

The inception V3 is introduced as GoogLeNet in 201. There are various Inception modules that make up the Inception model. The Inception v3 model, which was introduced in 2015, has 42 layers overall and a reduced mistake rate than its forerunners. The final Inception V3 model shows as Fig 2.

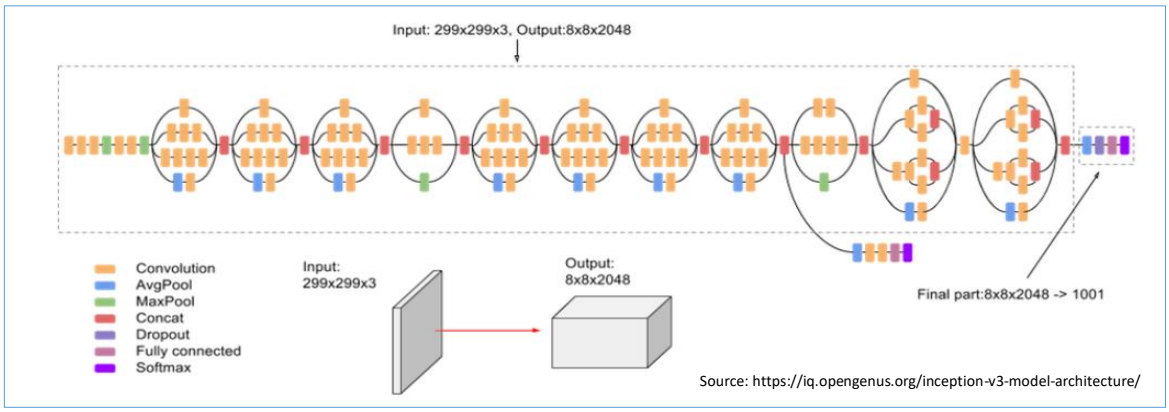


Fig. 2. Inception V3 architecture

In their 2015 computer vision research, initially, presented the ResNet model, which made it possible to train incredibly deep neural networks. ResNet is used to avoid the Vanishing Gradient Problem during backpropagation (Fig 3). The proposed architecture is represented as in Fig. 4.

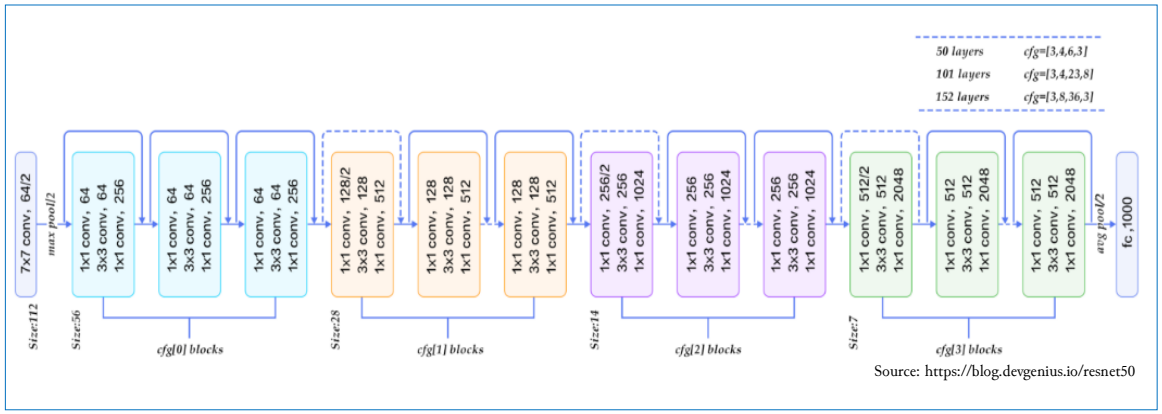


Fig. 3. ResNet architecture.

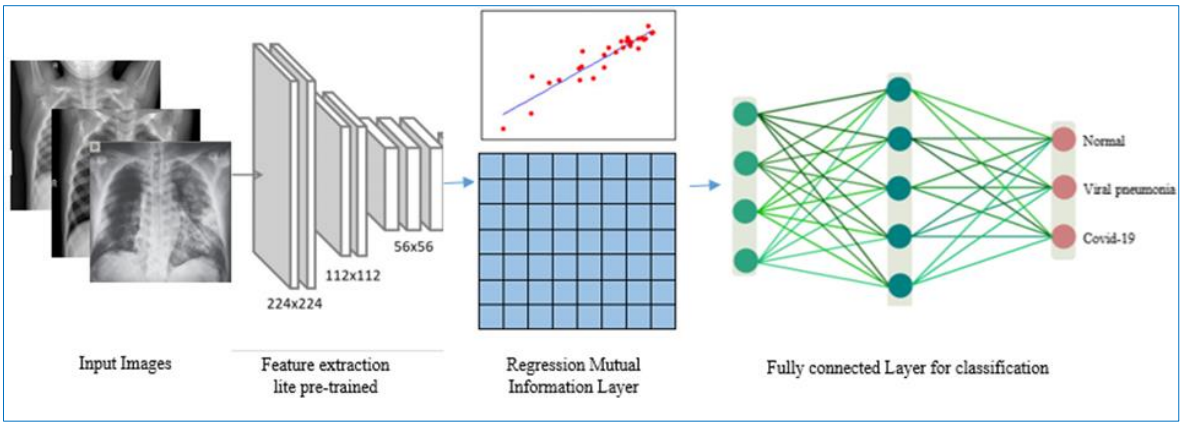


Fig. 4. The schematic representation of proposed architecture

2.3. Performance metrics

In order to evaluate the performance of different pre-trained models, K-fold cross-validation was used to verify the training models. The effectiveness of several networks was compared using three performance measures, including accuracy, sensitivity, and specificity. The predictive formulas were defined as:

$$Sensitivity = \frac{TP}{TP+FN} \tag{1}$$

$$Specificity = \frac{TN}{TN+FP} \tag{2}$$

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \tag{3}$$

The experiment dataset consist of 3-class classification. Unlike binary classification, the performance was measured for each individual class. For example, the formulas of class 1 were defined as:

$$TP = the\ number\ of\ correctly\ predicts\ the\ positive\ class\ 1\ as\ positive \tag{4}$$

$$TN = TN(class_{2,2}) + TN(class_{2,3}) + TN(class_{3,2}) + TN(class_{3,3}) \tag{5}$$

$$FP = FP(class_{1,2}) + FP(class_{1,2}) \tag{6}$$

$$FN = FN(class_{2,1}) + FP(class_{3,1}) \tag{7}$$

2.4. Feature extraction using pre-trained models

Many medical data sets have been effectively classified, segmented, and used to detect lesions using deep learning models. In this study, ResNet50 and InceptionV3 were used to extract the image features. Fig. 5 shows a lot of versions of the x-rays image with different highlighted features. However, some images contained weak information (row 2 column 3). When the features were extracted, the main objective of deep learning is to discover useful representations [25]. For maximizing between the complete input and the encoder output to learn the useful representations, mutual information was proposed to address this problem.

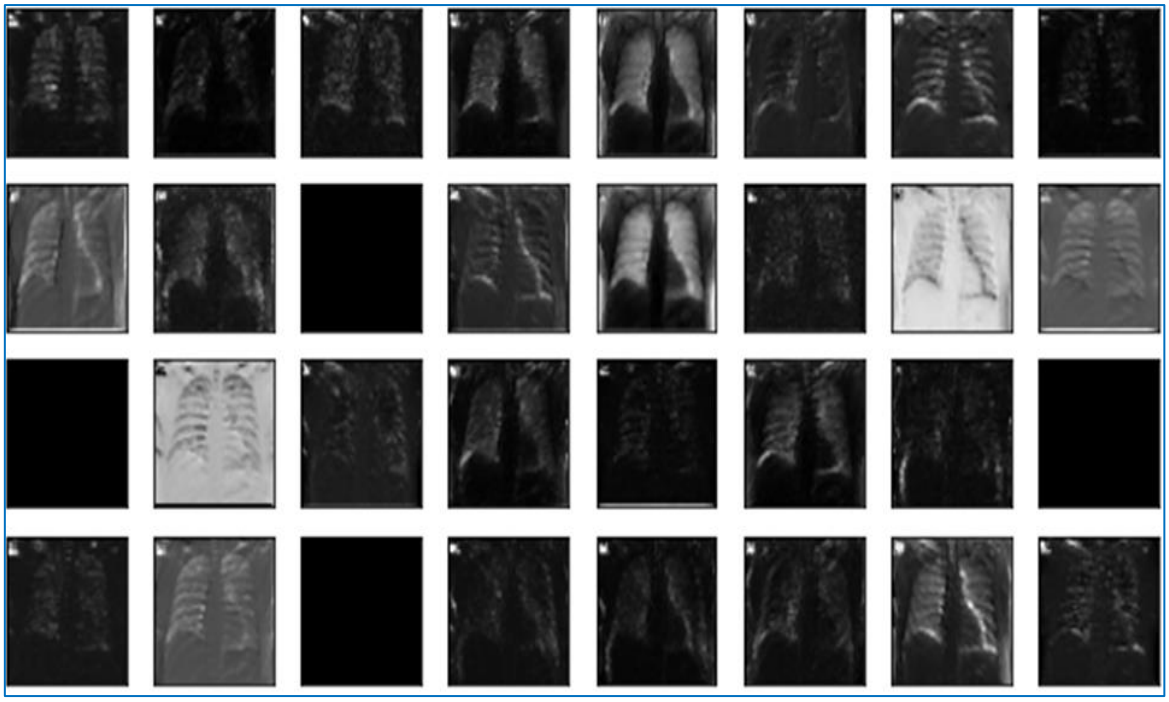


Fig. 5. An example of the activation map of image features

2.5. Mutual Information Evaluation

Information theory can be used to calculate how much information is shared between two variables in a relationship. When one variable is known, the amount of uncertainty in the other variable can be lowered. The amount of information that is unclear can be reduced when another variable is known. When the condition of Y is known, the uncertainty in the state of X is reduced, and the amount of pertinent information increases. Conditional entropy and probability distributions are typically used to calculate the classic mutual information. The pointwise mutual information H (X; Y) pairs estimated posterior knowledge of the number of each dependent pair. A search technique to choose potential feature sets X is the mutual information criterion [26]. The complexity typically dictates how each feature is added or removed based on high or low scores in the search technique. Therefore, the regression over feature and target class were established. An image is encoded using a convolution neuron network until reaching a feature map of M x M feature vectors corresponding to N input patches. These vectors were flattened into a single feature vector, x. In this study, the regression mutual information (RMI) was performed shown as:

$$RMI(X;Y) = 1 - \sum_{i=1}^{n} e_i^2 \tag{8}$$

$$e = y_i - \hat{y}_i \tag{9}$$

$$\hat{y}_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \tag{10}$$

The correlation between the observed outcomes and the observed predictor values is measured by the RMI score, which is expressed as a R square (r2). R square, which typically runs from 0 to 1, is the square of the coefficient of multiple correlation when further regression is incorporated. Then, the new feature set $\hat{X}$ was selected from top-5 RMI score.

Fig. 6 show the regression mutual information evaluation process. For example, the features of image1 (see Fig. 3) were extracted and flatted into 56×56 (56×56 pixels) and 32 patches. The RMI scores were computed and selected from top-5. Suppose the feature vector does not support useful representation. For selecting maximum features, the RMI from the whole input, thus, the feature set was selected only useful input.

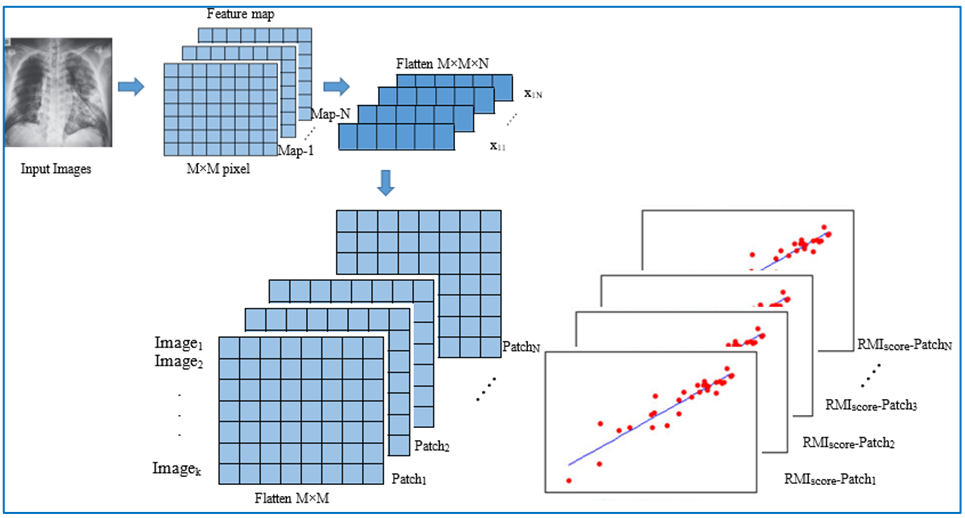


Fig. 6. the mutual information evaluation process. The local feature per image was map and flat in M×M×N size. Then the RMI scores were computed. The feature maps which have top-5 RMI scores were selected.

2.6. Fine-tune and Classification Layer

Training CNN on a small dataset such as medical image often affects the CNN ability to generalize. Therefore, transfer learning network was used to learn features. The final layer (the softmax layer) is often truncated and replaced with new softmax layers. For instance, a pre-trained network on ImageNet has a 1000-category softmax layer. Our experiment is performed with three categories of chest x-ray images. Instead of 1000 categories, the new softmax layer of the network will only have 3. Cross validation was used to fine-tune the back propagation on the network using the pre-trained weights. Then, the new features $\hat{X}$ were parsed through this network to fully connected layer for classification task.

3. Results and Discussion

The overall objective of this research is to demonstrate the utility of our novel RMI approach for COVID-19 diagnosis. Therefore, two sets of experiments were conducted. First, the original models from the pre-trained were trained to classify CXR images. Second, the feasibility of applying RMI to enrich the traditional models was improved diagnosis accuracy.

3.1. Experiment 1: original pre-trained architecture

The input images are fed into the trained ResNet and InceptionV3 to extract image features. Overall results are summarized in Table 1. By using all the CXR features generated by ResNet152V2, we obtain the accuracy of 92.21% (Sensitivity=100% Specificity=90%). With additional features from generated by InceptionV3, we obtain the accuracy of 91.39% (Sensitivity=100% Specificity=82.50%). As seen,

ResNet152V2 performed best on CXR dataset, while InceptionV3 was the lowest performance. Directly compared to InceptionV3, ResNet152V2's diagnosis produces significantly superior outcomes. The outcome may be accounted for by the fact that classifiers need to fit data more precisely using convolutional layers. The ResNet152V2 including RMI obtains the accuracy of 98.77% (Sensitivity=100% Specificity=98.02%), while InceptionV3 including RMI obtains the accuracy of 93.44% (Sensitivity=92.86% Specificity=93.44%).

Table 1. Comparison of classification performances

Model	#Layer	Original			RMI		
		Accuracy	*Sensitivity*	*Specificity*	*Accuracy*	*Sensitivity*	*Specificity*
ResNet152V2	565	92.21	100	90	**98.77**	**100**	**98.02**
InceptionV3	312	91.39	100	82.5	**93.44**	92.86	**91.09**

3.2. Experiment 2: apply RMI to enrich the traditional models for improved diagnosis

In order to study the features that contribute to the goal class, the regression mutual information of each feature set was calculated, and the source picture for each feature was selected. The RMI scores were measured through calculating the regression mutual information. The feature set corresponding in top-5 RMI scores were obtained the final features. Table 2 summarizes the number of trainable features from different models (original vs. RMI).

Table 2. the number of trainable feature from different models (original vs. RMI)

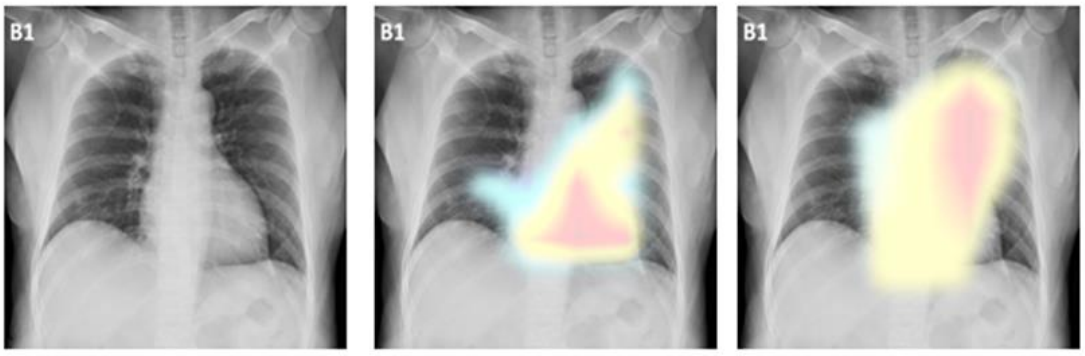
Model	The number of feature		%Reduction	Accuracy	
	Original	*RMI*		*Original*	*RMI*
ResNet152V2	2,048	245	88.00%	92.21%	98.77%
InceptionV3	2,048	320	84.37%	91.39%	93.44%

Table 2 showed that two original models used 2,048 features, while RMI ResNet152V2 used 245 and RMI InceptionV3 used 320, respectively. The reduction contributes 88% of RMI ResNet152V2 and 84.37% of InceptionV3. In addition, RMI not only reduce trainable feature but also improve the accuracy of COVID-19 diagnosis.

3.3. COVID-19 predictions and explanations

The interpretation of features is important not only for the explanation but also for the confirmation of the diagnosis. The important areas assist physician in using their interpretive abilities to diagnose patients more quickly and accurately [27]. Based on the locations where the activation maps overlay the original image, the significant features can be found.

As seen in Fig. 7, the feature map including RMI generated by ResNet152V2 are more accurate than InceptionV3. The rationale is that ResNet152V2 with RMI emphasizes joined features more specifically than specific components. When highlights regions much more precisely, it provides more human-interpretable explanations.



(a) (b) (c)

Fig. 7. The regions of some COVID-19 cases within the lungs are localized. (a) Original COVID-19 CXR image (b) critical regions from ResNet and (c) critical regions from InceptionV3

The confusion matrix of the best model performance is shown in Fig. 8. Table 3 demonstrates that the majority of samples are correctly identified using the original ResNet152V2 with 0.93, 0.92, and 0.92, respectively. For the RMI ResNet152V2 is even slightly higher, yielding 0.94, 0.93, and 0.92, respectively.

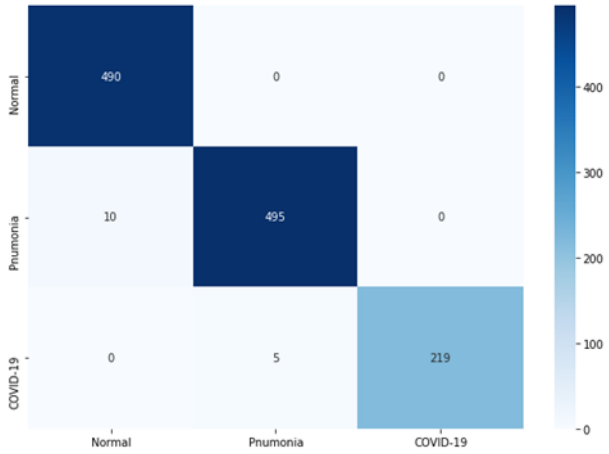


Fig. 8. The confusion matrix of RMI ResNet152V2

Table 3. Most samples are accurately classified with respect to precision, recall, and F1 scores.

Model	Precision	Recall	F1
Original ResNet152V2	0.93	0.92	0.92
RMI ResNet152V2	0.94	0.93	0.92

The positive predictive value (PPV) was established based on these findings to predict whether infected individuals would be diagnosed as positive. Only five of the 224 COVID-19 patient samples in our test set were incorrectly identified as pneumonia, yielding a PPV for COVID-19 cases of 97.76 percent, significantly surpassing a comparable technique [27], [28]. To provide a clearer picture in both the original and RMI scenario, we also report the class-specific measurements in Table 4.

Table 4. Class-specific classification results

Model	Original ResNet152V2			RMI ResNet152V2		
	Precision	*Recall*	*F1*	*Precision*	*Recall*	*F1*
Normal	0.91	1	0.95	0.90	1	0.95
Pneumonia	0.90	0.91	0.90	0.90	0.91	0.90
COVID-19	1	0.81	0.90	1	0.81	0.90

4. Conclusion

This study proposed RMI-DeepCNN to predict COVID-19 based on CXR pictures in this research. On the basis of CXR pictures, two pre-trained models, ResNet152V2 and InceptionV3, were used to predict normal, viral pneumonia, and COVID-19. The best model is RMI-ResNet152V2, which achieves an accuracy of 98.77% (Sensitivity: 100%; Specificity: 98.02%). According to evaluation results, our method outperforms a recent method in with a PPV of 97.76% and recall of 81%. Based on our results, RMI-DeepCNN provides the following proof based on the experiments and findings: First, expanding the feature selection method can still perform better than using only the original features even when a general strategy does not. Second, since precise diagnosis is crucial, models with many trainable parameters and a deeper layer of training can produce correct predictions during inference time. The chosen subset of all features utilizing RMI may be a particular strategy. There are some limitations in this study. First, CXR images for COVID-19 infection cases is insufficient to avoid the overfitting for our models. Second, the diagnoses and localization were not compared accuracies with the radiologists. In future, we intend to overcome these limitations.

Declarations

Author contribution. All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.

Funding statement. None of the authors have received any funding or grants from any institution or funding body for the research.

Conflict of interest. The authors declare no conflict of interest.

Additional information. No additional information is available for this paper.

References

[1] W. Wang *et al.*, "Detection of SARS-CoV-2 in Different Types of Clinical Specimens," *JAMA*, vol. 323, no. 18, pp. 1843–1844, Mar. 2020, doi: 10.1001/jama.2020.3786.

[2] T. Yang, Y.-C. Wang, C.-F. Shen, and C.-M. Cheng, "Point of Care RNA-Based Diagnostic Device for COVID-19," *Diagnostics*, vol. 10, no. 3, pp. 1–3, Mar. 2020, doi: 10.3390/diagnostics10030165.

[3] D. Wang *et al.*, "Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China," *JAMA*, vol. 323, no. 11, pp. 1061–1069, Mar. 2020, doi: 10.1001/jama.2020.1585.

[4] N. Chen *et al.*, "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study," *Lancet*, vol. 395, no. 10223, pp. 507–513, Feb. 2020, doi: 10.1016/S0140-6736(20)30211-7.

[5] Q. Li *et al.*, "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," *N. Engl. J. Med.*, vol. 382, no. 13, pp. 1199–1207, 2020. Available at: Google Scholar

[6] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, Feb. 2020, doi: 10.1016/S0140-6736(20)30183-5.

[7] V. Chouhan *et al.*, "A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images," *Appl. Sci.*, vol. 10, no. 2, pp. 1–17, Jan. 2020, doi: 10.3390/app10020559.

[8] Z. Wu, C. Shen, and A. van den Hengel, "Wider or Deeper: Revisiting the ResNet Model for Visual Recognition," *Pattern Recognit.*, vol. 90, pp. 119–133, Jun. 2019, doi: 10.1016/j.patcog.2019.01.006.

[9] J. M. Ahn, S. Kim, K.-S. Ahn, S.-H. Cho, K. B. Lee, and U. S. Kim, "A deep learning model for the detection of both advanced and early glaucoma using fundus photography," *PLoS One*, vol. 13, no. 11, p. e0207982, Nov. 2018, doi: 10.1371/journal.pone.0207982.

[10] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015, doi: 10.1007/s11263-015-0816-y.

[11] X. Gu, L. Pan, H. Liang, and R. Yang, "Classification of Bacterial and Viral Childhood Pneumonia Using Deep Learning in Chest Radiography," in *Proceedings of the 3rd International Conference on Multimedia and Image Processing - ICMIP 2018*, 2018, pp. 88–93, doi: 10.1145/3195588.3195597.

[12] L. Wang, Z. Q. Lin, and A. Wong, "COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, Dec. 2020, doi: 10.1038/s41598-020-76550-z.

[13] A. S. Joaquin, "Using deep learning to detect pneumonia caused by ncov-19 from x-ray images," *Towards data science*. 2020. Available at: Google Scholar

[14] R. Shwartz-Ziv and N. Tishby, "Opening the black box of deep neural networks via information," *Arxiv preprint*. pp. 1–19, 2017. Available at: Google Scholar

[15] A. M. Saxe *et al.*, "On the information bottleneck theory of deep learning," *J. Stat. Mech. Theory Exp.*, vol. 2019, no. 12, p. 124020, 2019. doi: 10.1088/1742-5468/ab3985

[16] S. Vasudevan, "Dynamic learning rate using Mutual Information," *Arxiv preprint*. pp. 1–11, 2018. Available at: Google Scholar

[17] N. O. Hodas and P. Stinis, "Doing the Impossible: Why Neural Networks Can Be Trained at All," *Front. Psychol.*, vol. 9, p. 1185, Jul. 2018, doi: 10.3389/fpsyg.2018.01185.

[18] R. D. Hjelm *et al.*, "Learning deep representations by mutual information estimation and maximization," *Arxiv preprintiv preprint*. 2018. Available at: Google Scholar

[19] J. Bullock, A. Luccioni, K. Hoffman Pham, C. Sin Nga Lam, and M. Luengo-Oroz, "Mapping the landscape of Artificial Intelligence applications against COVID-19," *J. Artif. Intell. Res.*, vol. 69, pp. 807–845, Nov. 2020, doi: 10.1613/jair.1.12162.

[20] T. Ai *et al.*, "Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases," *Radiology*, vol. 296, no. 2, pp. E32–E40, Aug. 2020, doi: 10.1148/radiol.2020200642.

[21] B. Ghoshal and A. Tucker, "Estimating Uncertainty and Interpretability in Deep Learning for Coronavirus (COVID-19) Detection." pp. 1–14, 2020. Available at: Google Scholar

[22] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. Rajendra Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Comput. Biol. Med.*, vol. 121, no. April, p. 103792, 2020, doi: 10.1016/j.compbiomed.2020.103792.

[23] G. Hub, "Open database of COVID-19 cases with chest X-ray or CT images," 2022. [Online]. Available: github.com.

[24] Paul Mooney, "Chest X-Ray Images (Pneumonia)," *Kaggle*, 2018. [Online]. Available: kaggle.com.

[25] M. I. Belghazi *et al.*, "Mutual information neural estimation," in *International conference on machine learning*, 2018, pp. 531–540. Available at: Google Scholar

[26] H. M. Huan Liu, *Feature Extraction, Construction and Selection*. Boston, MA: Springer US, 1998. Available at: Google Books

[27] L. Wang, Z. Q. Lin, and A. Wong, "COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," *Sci. Rep.*, vol. 10, no. 1, p. 19549, Dec. 2020, doi: 10.1038/s41598-020-76550-z.

[28] M. R. Karim, T. Dohmen, M. Cochez, O. Beyan, D. Rebholz-Schuhmann, and S. Decker, "DeepCOVIDExplainer: Explainable COVID-19 Diagnosis from Chest X-ray Images," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2020, pp. 1034–1037, doi: 10.1109/BIBM49941.2020.9313304.