A Dual Framework of Feature Selection based on the Fusion of HOG and Pyramid HOG for the Categorization of COVID-19

Mehak Mushtaq Malik[a], *Muhammand Hamza Azam*[b] ✉, *and Farhat Afza*[a]

[a] Department of Computer Science, COMSATS University Islamabad - Wah Campus, Wah Cantt 47040, Pakistan
[b] Centre of Research in Data Science, Computer and Information Sciences Department, Universiti Teknologi PETRONAS, Seri Iskandar, 32610, Perak, Malaysia

✉, [b] muhammad_17007652@utp.edu.my
[b] Centre of Research in Data Science, Computer and Information Sciences Department, Universiti Teknologi PETRONAS, Seri Iskandar, 32610, Perak, Malaysia

ABSTRACT

COVID-19 was initially detected in Wuhan, China. The virus spread all over the world at a rapid speed. COVID-19 is an infection that may cause infections in the respiratory system and the lungs. In order to diagnose COVID-19, chest X-rays have been utilized extensively. The purpose of this research is to create a computer-vision-based method for identifying COVID in chest X-rays. In the proposed model handcrafted features known as HOG and PHOG are derived and fused. After features have been fused, the Binary Grey Wolf Optimization technique is used with an Entropy-Based Optimization Algorithm to select the most significant features possible. The proposed model results are evaluated on a benchmark X-ray dataset that gives greater than 99% accuracy. The proposed model performed better compared to existing published works in this domain.

Keywords: *COVID; Classification, Entropy; HOG; PHOG; SVM*

1. INTRODUCTION

In the present era, the outbreak of a unique newly born coronavirus disease (COVID-19) was introduced which affected many countries worldwide. It is a deadly virus that extends predominantly through droplets generated when a sick person coughs or sneezes or fluid that is released from the nose. The peak indications of COVID-19 are the hawk, weariness, breakdown of breathing, and high temperature. Particular societies may cultivate more unadorned forms of the syndrome, known as pneumonia. Currently, "COVID-19" is also known as "innovative," as it is a new draining in the household of viruses. The proportion of infected patients is increasingly high. The growing number of COVID-19 cases is the major reason behind increasing the anxiety ratio. Furthermore, the observation and investigations of COVID-19 play a vital part to prevent this disease. This disease habitually attacks the mature (people above 50 age) and those infected by the dangerous disease. The overall spread of infectious cases is unsafe for elder people and also those already suffering from some critical disease[1].

Currently, the authors analysed the chest X-ray dataset containing both the positive and negative images of COVID-19. They extract their deep features using different deep neural networks and then classify them using an SVM classifier. Besides the comparison of different statistical values of different neural networks, they showed that RESNET50 showed better results from the perspective of accuracy and provides an accuracy of about 95.38%. The authors discussed the clinical findings of pathological images regarding COVID-19 which can help in scientific research [2]. The authors analysed clinical findings in asymptotic patients of COVID-19 in Italy and suggested that the major transmission of Covid-19 is through asymptotic carriers which transmit the disease to other persons without even showing proper symptoms of the disease. A new neural network architecture, COVIDX-Net, is proposed to detect COVID-19-positive images from a dataset of chest X-ray images. COVIDX-Net is based on seven already existing CNNs. They showed that the best performance in

the detection of COVID-19 is obtained by using VGG19 and DenseNet-201. Machine learning [3] methods were used to differentiate between infected/healthy person's lung X-rays with high accuracy. In the literature, several methods have been presented for COVID-19 detection using X-rays [4]. However, these still, limitations exist in this domain due to several factors irregular shape, size of the lesions, and prominent features extraction for accurate classification [5]. The core contributions steps are manifested as:

- Shape based features such as HOG and PHOG are extracted from the input images then these features are serially fused.
- The optimal features are selected out of fused features vector using Binary Grey Wolf with iEntropy Based Optimization approach.

The article organization is as: Section II discusses the related work, proposed method steps are elaborated in Section III, results and discussion are describe in Section IV and conclusion of the article is written in Section V.

2. RELATED WORK

COVID-19 [6] is an alarmingly fast-spreading worldwide epidemic [7, 8]. Feature extraction is a crucial step in image processing for identification and classification [9, 10]. Multiple multi-resolution characteristics for COVID-19 were analysed by Ismael et al. [5] using data from chest x-rays and CT scans. They used entropy for function extraction and achieved a 99.92% precision. FUSI-CAD model [11], is the convergence between hand-crafted and deep elements. In which textual handcrafted feature called the discrete wavelet transform (DGT) and then DGT and GLCM features are fused to create a single vector that provides 99.0% accuracy [12]. The texture features are derived from chest x-ray images, that provides 79.52% accuracy [13]. Ozturk et al. [14] demonstrated that the amount of data available for COVID-19 is insufficient to train CNN models effectively, prompting them to use handcrafted features for detection. The GLCM, GLRLM, and DFTA features are extracted and fused into single vector then applied PCA for the selection of optimum features. This method is tested on local dataset that provides 86.54 % accuracy [6]. Kang et al. [15] extracted various handcrafted features from CT images and assign labels. Elazir et al. [16] used orthogonal momentum features derived from chest x-ray images to detect COVID-19 that provides accuracy of 98.9%. Ankita et al. [17], can distinguish between stable, tuberculosis, and pneumonia with 95.9% accuracy. Furthermore, denseNet-161 [18] is used to classify the COVID-19/pneumonia with a 98.9% accuracy. The ResNet-18 [19], is used for the detection of COVID19 severity rate with 76 % accuracy. The long-term short memory (LTSM) model is designed for COVID-19 classification with 99.4% accuracy [20].

3. METHODOLOGY

In proposed methodology HOG (Histogram of Oriented Gradients) [40] and PHOG (Pyramid of Histogram of Oriented Gradients) [41, 42] features are extracted. After that extracted features of HOG and PHOG are fused to create strong vector that fed to Binary Grey Wolf Optimization algorithm [43, 44] with Entropy-Based Optimization Algorithms [45] for selecting the core features. Fig. 2 represents the architecture of the proposed model.

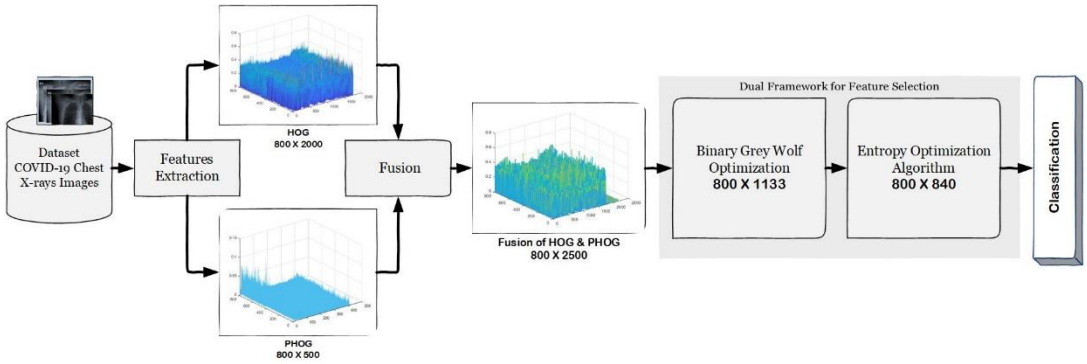

Fig. 1. Flowchart of the Proposed Methodology

The input images are resized with the dimension of 64 x 64 as shown in Fig. 2. Resize function mathematically express as in Eq. (1):

$$R=\sum_{i=1}^{n}(I_i, \text{Scale}) \quad (1)$$

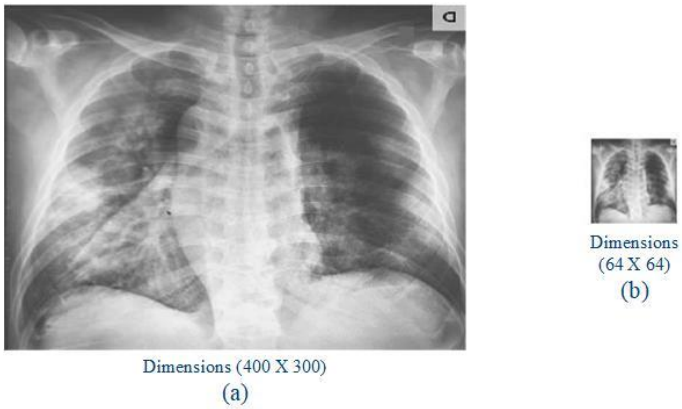

Fig. 2. (a) Original Chest X-ray Image (b) Resize Chest X-ray Image

HOG and PHOG are well-known shape features are used for feature extraction. In HOG feature extraction, gash areas are labeled through the circulation of gradients. Two-D gradients "da" and "DB" are planned by purifying the gashed area. The HOG is a well-organized mode to mine features by way of the pixel colors for constructing an object. With the responsiveness of image gradient vectors, resizing, and color stabilization. Division of the image into numerous 8x8 pixel cells. In all cells, the measure of these 64 cells is binned and added into buckets of unsigned direction [48]. While PHOG features remained multiplied as they involve local as global latitudinal material. The benefit includes integration with machines of X-ray for motorized recognition [49]. Histogram visualizations of HOG and PHOG is shown in Fig. 3. In mathematical terms, Hog features express as Eq. (2):

$$da = \partial \text{ fun (a, b)} \div \partial a \quad (2)$$

Similarly, we have Eq. (3):

$$db = \partial \, fun \, (a,b) \div \partial b \quad (3)$$

The weighting parameter and orientation of the image is "$wfun(a,b)$", so in Eq. (4):

$$wfun(a,b) = \sqrt{da^2 + db^2} \quad (4)$$

Now in term of the angle, we express in Eq. (5):

$$\vartheta(a,b) = tanh^{-1}(da^2 \div db^2) \quad (5)$$

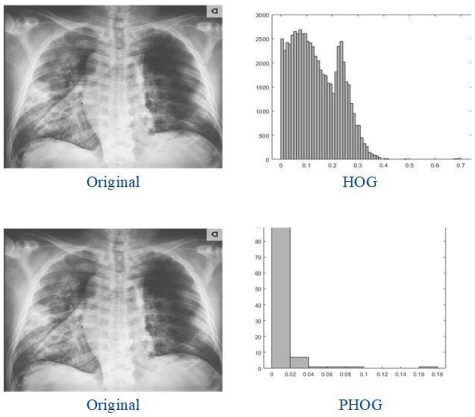

Fig. 3. Histogram visualization.

Binary Grey Wolf Optimization algorithm is performed with Entropy-Based Optimization. The Binary Grey Wolf Optimization updates the wolf's position by turning a binary vector into a position, as shown in Eq. (6):

$$Z^q(x+1) = \begin{cases} 1, if\ D\left(\frac{Z^1 Z^2 Z^3}{3}\right) \geq R_D \\ 0,\ otherwise \end{cases} \quad (6)$$

Where, R_D is a random vector, q is the dimension and D is the sigmoid function is shown in Eq. (7),

$$D(t)=\frac{1}{1+\exp(-10(t-0.5))} \quad (7)$$

Secondly, grey wolves are assessed for fitness. The fitness-based option of three leaders, alpha, beta, and detail. Now one by one wolf, the Z1, Z2, and Z3 are computed. Besides, wolf fitness is assessed and the alpha, beta, and delta positions are modified. In the end, the alpha solution is selected for optimal results. In general, in comparison to other metaheuristic optimizations, BGWO is better, modular, and flexible. However, BGWO is also restricted to optimal local restrictions. BGWO uses the 3 best solutions (leaders) to update the position, which ensures that any wolf attempts to switch to leadership positions. The entropy [21] is calculated in each discretization bin over the scattering of feature values as shown in Fig 4. The values are ignored by entropy based on similar features.

$$\text{Entropy} = -\sum i \sum j [e(i,j) \log(e(i,j)) \quad (8)$$

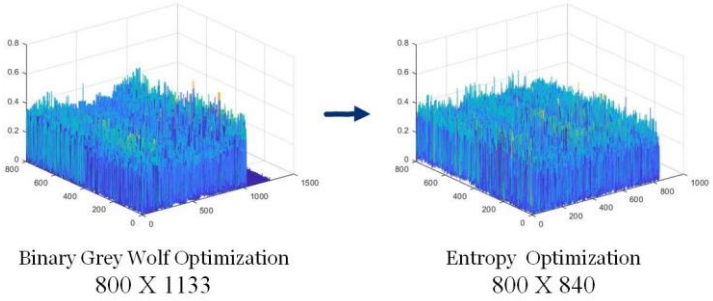

Fig. 4. Features Visualization using Mesh Plot

In the proposed model HOG features are extracted from x-rays images with 1*2000 dimension while PHOG features are extracted with 1*500 dimension. The extracted features are fused together with 1*2500 dimension and fed to binary grey wolf optimization for core features selection. After that selected 1*1133 core features

are further passed to the entropy optimization method in which 800*840 best features are selected out of 1133 features and passed to the classifiers for COVID-19 classification.

4. RESULTS AND DISCUSSION

The proposed method performance is evaluated on publicly benchmark x-rays COVID-19 imaging dataset. This dataset is publicly available on the Kaggle website [22] [26] [23] and description about the dataset is mentioned in Table1.

Table 1. Description about dataset

Classes	CPLXI	CDXD	CCXSD	Total Images	Flip Horizontal
COVID-19 Images	70-Images	70-Images	60-Images	200-images	400-images
Normal Images	28-Images	25-Images	147-Images	200-images	400-images

The feature vector length of 1*1133 is obtained from binary Gray wolf optimization model that is further passed to entropy method for best features selection in three times. In each time we obtained different length of features vector such as 1*840, 1* 550 and 1*250 that are shown in Table 2.

Table 2. Information about the best of three experiment

Experiments	Total Number of Features
	PHOG + HOG
1	840
2	550
3	250

Table 2 depict the selection of optimum features obtained after employing entropy that are further passed to the classifiers for the analysis of prediction results.

4.1 EXPERIMENT# 1: COVID-19 Classification based on Core Features with the Length of 1*840

In this experiment, data is divided into training and testing using 10-fold and supplied to the different kernels of the SVM for COVID-19 classification based on 1*840 optimal features vector as mentioned in Table 3.

Table 3. Classification Results based on 1*840 Core Selected Features Length

Classifier	Accuracy	Sensitivity	Specificity
L-SVM	99.6%	99.26%	99.89%
Q-SVM	99.9%	99.75%	99.91%
FG-SVM	97.3%	99.95%	94.79%
MG-SVM	99.5%	99.01%	99.89%
CG-SVM	99.0%	98.04%	99.96%

Table 3 presents the classification outcomes, in which we achieved accuracy of 99.9% on Q-SVM, 99.6% on L-SVM, 97.3% on FG-SVM, 99.5% on MG-SVM and 99.0% on CG-SVM. Out of different kernels of SVM Q-SVM performed better compared to others.

4.2 EXPERIMENT#2: COVID-19 Classification based on Core Features with the Length of 1*550

Similarly in this experiment, classification results are also computed on 1*550 selected features vector as mentioned in Table 4.

Table 4. Classification Results based on 1*550 Core Selected Features Length

Classifier	Accuracy	Sensitivity	Specificity
L-SVM	98.8%	97.56%	99.89%
Q-SVM	99.6%	99.91%	99.26%
FG-SVM	89.0%	99.88%	81.97%

MG-SVM	99.3%	98.52%	99.82%
CG-SVM	98.5%	97.09%	99.89%

Table 4 depict the classification outcomes, in which achieved accuracy is 98.8% on L-SVM, 99.6% on Q-SVM, 89.0% on FG-SVM, 99.3% on MG-SVM and 98.5% on CG-SVM. In this experiment Q-SVM gives better results compared to others.

4.3 EXPERIMENT 3: COVID-19 Classification based on Core Features with the Length of 1*250

In this experiment, Core Selected Features length of 1*250 are passed to the different SVM kernels. The achieved outcomes are also presented in confusion matrix in Table 5.

Table 5. Confusion Matrix

Class	COVID	NORMAL
COVID	**400**	
NORMAL		**394**

Table 5 shows the binary classification results such as COVID-19/normal classes. The proposed model classification outcomes are mentioned in Table 6.

Table 6. Classification results based on 1*250 core selected features length

Classifier	Accuracy	Sensitivity	Specificity
L-SVM	98.9%	97.80%	99.90%
Q-SVM	99.9%	99.75%	99.91%
FG-SVM	87.6%	99.83%	80.55%
MG-SVM	99.1%	98.28%	99.33%
CG-SVM	98.1%	97.00%	99.23%

In Table 5, we achieved accuracy of 98.9% on L-SVM, 99.9% on Q-SVM, 87.6% on FG-SVM, 99.1% on MG-SVM and 98.1% on CG-SVM. In this experiment, Q-SVM also gives highest accuracy compared to others.

4.4 Comparison of the proposed classification results with existing methods

The proposed method results are compared to the existing methods as mentioned in Table 7.

Table 7. Proposed Method Results Comparison

Ref#	Year	Results
[24]	2022	98.31% Sensitivity
[25]	2021	96.73% Accuracy
Proposed Model		99.9% Accuracy, 99.75% Sensitivity

The comparison results are presented in Tab.6 in which the ensemble stacked CNN model is employed for COVID19 classification. This model is evaluated on X-rays images that provide 98.31% sensitivity [24]. The modified version of the ResNet-18 model is utilized for classification that gives 96.73% accuracy [25]. As compared to existing works in this research features fusion model is proposed with an optimized features extraction strategy that provides better classification outcomes compared to existing research.

5. CONCLUSION

Classification of COVID-19 is a challenging task, due to the complex patterns of lesions. In this research hand-crafted HOG and PHOG features are extracted and fused. After features fusion, the Binary Grey Wolf Optimization algorithm is used with Entropy-Based Optimization Algorithms for the selection of core features. The selected core features are passed to the SVM for COVID-19 classification. The proposed classification model provides 99.90% Accuracy, 99.75% Sensitivity, 99.9% Specificity, and 1.00 AUC.

REFERENCES

[1] A. Tahamtan, and A. Ardebili, "Real-time RT-PCR in COVID-19 detection: issues affecting the results," *Expert review of molecular diagnostics,* vol. 20, no. 5, pp. 453-454, 2020.

[2] Y. Fan, J. Liu, R. Yao, and X. Yuan, "COVID-19 detection from X-ray images using multi-kernel-size spatial-channel attention network," *Pattern Recognition,* vol. 119, pp. 108055, 2021.

[3] A. M. Fayyaz, K. A. Al-Dhlan, S. U. Rehman, M. Raza, W. Mehmood, M. Shafiq, and J.-G. Choi, "Leaf Blights Detection and Classification in Large Scale Applications," *INTELLIGENT AUTOMATION AND SOFT COMPUTING,* vol. 31, no. 1, pp. 507-522, 2022.

[4] A. Shelke, M. Inamdar, V. Shah, A. Tiwari, A. Hussain, T. Chafekar, and N. Mehendale, "Chest X-ray classification using deep learning for automated COVID-19 screening," *SN computer science,* vol. 2, no. 4, pp. 1-9, 2021.

[5] H. Farhat, G. E. Sakr, and R. Kilany, "Deep learning applications in pulmonary medical imaging: recent updates and insights on COVID-19," *Machine vision and applications,* vol. 31, no. 6, pp. 1-42, 2020.

[6] A. Muiz Fayyaz, M. Kolivand, J. Alyami, S. Roy, and A. Rehman, "Computer Vision-Based Prognostic Modelling of COVID-19 from Medical Imaging," *Prognostic Models in Healthcare: AI and Statistical Approaches*, pp. 25-45: Springer, 2022.

[7] S. O. Folorunso, J. B. Awotunde, N. O. Adeboye, and O. E. Matiluko, "Data Classification Model for COVID-19 Pandemic," *Advances in Data Science and Intelligent Data Communication Technologies for COVID-19*, pp. 93-118: Springer, 2022.

[8] S. A. Mahmoudi, S. Stassin, M. E. H. Daho, X. Lessage, and S. Mahmoudi, "Explainable Deep Learning for Covid-19 Detection Using Chest X-ray and CT-Scan Images," *Healthcare Informatics for Fighting COVID-19 and Future Epidemics*, pp. 311-336: Springer, 2022.

[9] G. Kumar, and P. K. Bhatia, "A detailed review of feature extraction in image processing systems." pp. 5-12.

[10] D. ping Tian, "A review on image feature extraction and representation techniques," *International Journal of Multimedia and Ubiquitous Engineering,* vol. 8, no. 4, pp. 385-396, 2013.

[11] D. A. Ragab, and O. Attallah, "FUSI-CAD: Coronavirus (COVID-19) diagnosis based on the fusion of CNNs and handcrafted features," *PeerJ Computer Science,* vol. 6, pp. e306, 2020.

[12] R. Das, M. Arshad, P. Manjhi, and S. D. Thepade, "Covid-19 Identification with Chest X-Ray Images merging Handcrafted and Automated Features for Enhanced Feature Generalization." pp. 1-6.

[13] L. Hussain, T. Nguyen, H. Li, A. A. Abbasi, K. J. Lone, Z. Zhao, M. Zaib, A. Chen, and T. Q. Duong, "Machine-learning classification of texture features of portable chest X-ray accurately classifies COVID-19 lung infection," *BioMedical Engineering OnLine,* vol. 19, no. 1, pp. 1-18, 2020.

[14] Ş. Öztürk, U. Özkaya, and M. Barstuğan, "Classification of Coronavirus (COVID-19) from X-ray and CT images using shrunken features," *International Journal of Imaging Systems and Technology,* vol. 31, no. 1, pp. 5-15, 2021.

[15] H. Kang, L. Xia, F. Yan, Z. Wan, F. Shi, H. Yuan, H. Jiang, D. Wu, H. Sui, and C. Zhang, "Diagnosis of coronavirus disease 2019 (covid-19) with structured latent multi-view representation learning," *IEEE transactions on medical imaging,* vol. 39, no. 8, pp. 2606-2614, 2020.

[16] M. A. Elaziz, K. M. Hosny, A. Salah, M. M. Darwish, S. Lu, and A. T. Sahlol, "New machine learning method for image-based diagnosis of COVID-19," *Plos one,* vol. 15, no. 6, pp. e0235187, 2020.

[17] J. Redmon, "Darknet: Open source neural networks in c," 2013.

[18] N. F. P. Setyono, D. Chahyati, and M. I. Fanany, "Betawi Traditional Food Image Detection using ResNet and DenseNet." pp. 441-445.

[19] S. Targ, D. Almeida, and K. Lyman, "Resnet in resnet: Generalizing residual architectures," *arXiv preprint arXiv:1603.08029*, 2016.

[20] M. Z. Islam, M. M. Islam, and A. Asraf, "A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images," *Informatics in Medicine Unlocked*, vol. 20, pp. 100412, 2020, 2020.

[21] C. M. Vastrad, "Important Molecular Descriptors Selection Using Self Tuned Reweighted Sampling Method for Prediction of Antituberculosis Activity," *arXiv preprint arXiv:1402.5360*, 2014.

[22] D. Kermany, K. Zhang, and M. Goldbaum, "Labeled optical coherence tomography (oct) and chest x-ray images for classification," *Mendeley data,* vol. 2, no. 2, 2018.

[23] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, "Covid-19 image data collection: Prospective predictions are the future," *arXiv preprint arXiv:2006.11988*, 2020.

[24] M. Gour, and S. Jain, "Automated COVID-19 detection from X-ray and CT images with stacked ensemble convolutional neural network," *Biocybernetics and Biomedical Engineering,* vol. 42, no. 1, pp. 27-41, 2022.

[25] R. A. Al-Falluji, Z. D. Katheeth, and B. Alathari, "Automatic detection of COVID-19 using chest X-ray images and modified ResNet18-based convolution neural networks," *Computers, Materials, & Continua*, pp. 1301-1313, 2021.