COVID-19 Pneumonia Level Detection using Deep Learning Algorithm

**Abbas M. Ali · *Kayhan Zrar Ghafoor ·
Aos Mulahuwaish · Halgurd S.
Maghdid · Safwan Mawlood Hussein ·
Marwan A. Mohammed**

Abstract The first COVID-19 confirmed case is reported in Wuhan, China and spread across the globe with unprecedented impact on humanity. Since this pandemic requires pervasive diagnosis, it is significant to develop smart, fast and efficient detection technique. To this end, we developed an Artificial Intelligence (AI) engine to classify the lung inflammation level (mild, progressive, severe stage) of the COVID-19 confirmed patient. In particular, the developed model consists of two phases; in the first phase, we calculate the volume and density of lesions and opacities of the CT images of the confirmed COVID-19 patient using Morphological approaches. In the second phase, the second phase classifies the pneumonia level of the confirmed COVID-19 patient. To achieve precise classification of lung inflammation, we use modified Convolution Neural Network (CNN) and k-Nearest Neighbor (kNN). The result of the experiments show that the utilized models can provide the accuracy up to 95.65% and 91.304 % of CNN and kNN respectively.

Abbas M. Ali · Marwan A. Mohammed
Department of Software Engineering, Salahaddin University-Erbil, Iraq.
E-mail: abbas.mohamad@su.edu.krd

Kayhan Zrar Ghafoor
Department of Software Engineering, Salahaddin University-Erbil, Iraq; School of Mathematics and Computer Science, University of Wolverhampton, Wulfruna Street, Wolverhampton, WV1 1LY, UK. kayhan@ieee.org.

Aos Mulahuwaish
Department of Computer Science and Information Systems Saginaw Valley State University 7400 Bay Rd, Science East 174 University Center, MI 48710. amulahuw@svsu.edu.

Halgurd S. Maghdid
Department of Software Engineering, Faculty of Engineering, Koya University, Kurdistan Region-F.R.Iraq. First.Last@koyauniversity.org.

Safwan Mawlood Hussein
Computer Engineering Department, Tishk International University, Iraq. Safwan.mawlud@tiu.edu.iq.

*Kayhan Zrar Ghafoor is the corresponding author. kayhan@ieee.org.

1 Introduction

The novel COVID-19 started in Wuhan city, Hubei Province of China in November 2019. In December 2019 the virus has been announced by the world health organization (WHO) can cause a serious respiratory disease characterized clinically by fever, cough and lung inflammation. Even though the virus has initially appeared in China, it has been recognized now in numerous countries around the world [1], [2]. WHO announced this virus as a public health emergency on January 30, 2020. This is because of its fast spread out through person-to-person, as well as because most infected people have no immunity against this virus.

The COVID-19 spreads among humans and many various species of animals such as cats, camels, bats, and cattle. Primary cases of COVID-19 had links with live animal shops and seafood at the epicenter in Wuhan, suggesting animal-to-person route of transmission. Reports show a daily rise of confirmed cases since the outbreak began. Consequently, When the number of confirmed cases reached 118,000 with surpassing 4000 deaths, The WHO on March 11, 2020, announced COVID-19 as a pandemic [3]. Not long ago, Italy passes China in the number of mortality.

COVID-19 and human coronaviruses are classified below the Coronaviridae family. When these viruses infect people it causes Severe Acute Respiratory Syndrome (SARS) or moderate cold Middle East Respiratory Syndrome (MERS) [4]. SARS likewise is a viral respiratory disease caused by SARS-associated coronavirus (SARS-CoV), it is first reported in 2003 in Southern China and spread out in several countries around the world. Furthermore, in Saudi Arabia, MERS virus cases were first reported causing 858 deaths. This virus is believed to have appeared in bats, this belief depending on the examination of virus genomes [5].

The severity of coronavirus disease 2019 symptoms have been classified from very mild to severe, symptoms may appear after the first two days of exposure are fever, headache, dry cough, shortness breath in addition to loss smell or taste and symptoms might be changing after day 14 to persistent pain or pressure in the chest, confusion and breathing trouble. Many techniques apply to diagnosis COVID-19 such as the Computed Tomography (CT) scan and Nucleic Acid Test (NAT). The NAT is used to recognize precise nucleic acid sequence and species of organism, mainly a bacteria or virus which causes disease in tissue, blood, or urine. Even though NAT and detection kits play important roles in identifying COVID-19, CT scans still the most efficient and practical method for detecting the severity of the lung inflammation expected to be correlated with COVID-19 [6]. The embodiment of radiographic appearance of pneumonia for clinical diagnostic standard in Hubei province approved by the National Health Commission of China [7], which confirms the importance of CT scan images in the diagnosis of COVID-19 pneumonia severity.

Some experts believe that congregate COVID-19 cases with others may lead to more infections and hospitals become a hotspot of infection while they are waiting to have CT scan Image testing. In addition, the radiologists'

number is quite fewer than the number of patients which may result in late diagnosis and quarantine of infected people and less effective treatment of patients [7]. Consequently, for instance, in Italy, hospitals have had to give priority to patients with more severe manifestations such as high fever and breathing difficulty over others with less severe symptoms [8].

The fast transmission of COVID-19 and the increase in demand for diagnosis, has encouraged researchers to evolve more intelligent, highly sensitive, and effective diagnostic methods might help stem the spread of coronavirus disease 19. The diagnosis method which is managed by the radiologists is the manual measure of the lung infection quantity. Furthermore, Al-based automated pneumonia diagnosis is used to distinguish the density, the size, and the opacities of the lesions in COVID-19 confirmed cases. These algorithms are capable to analyse CT scan images outcome in a little while comparing to other available ways [9].

The chest CT scan images utilize by the radiologist for the purpose of monitoring the confirmed cases varying from early to serious stages [9]. The rapid progression of lung infection demands several CT scan images, understanding and analysing these Images is taking too much time and challenging task, particularly when the measurements of the lesion should be done manually on the CT scan and X-ray images. Therefore, the evolvement of the intelligent algorithm is urgently needed to precisely and automatically detect COVID-19 cases. Moreover, it is essential to establish a complete and ready-to-experiment dataset for the research community.

In this paper, we propose an Artificial Intelligence (AI) engine to classify the lung inflammation level of the COVID-19 confirmed patient. In particular, the developed model consists of two phases; in the first phase, we calculate the volume and density of lesions and opacities of the CT images of confirmed COVID-19 patient using Morphological approach. In the second phase, the second method classifies the pneumonia level of the confirmed COVID-19 patient. To achieve precise classification of lung inflammation, we used modified Convolution Neural Network (CNN) and k-Nearest Neighbor (kNN). The result of the experiments shows that the utilized models can provide accuracy up to 95.65% and accuracy by using the modified CNN whereas.

The rest of this paper is organized as follows. Section II provide the literature review on recent advances of developed AI systems for COVID-19 detection. This is followed by presenting an overview of the proposed approach and details of the designed algorithm. Section IV presents the detail of the dataset, materials and experimental results. Finally, Section V concludes the paper.

2 Related Work

In this section, the review of state-of-the-arts methods used for the diagnosis of novel COVID-19. There are many methods that are utilized to distinguish viral pneumonia in dubious cases.

Because of the new development of COVID-19, there are a small number of up-to-date reviews in this field. Even though, the existence of very rare literature on the diagnosis of this virus. In an effort, based on deep learning the researchers developed an Al engine to identify COVID-19 by using super-resolution CT images [1]. However, their suggested model depends exclusively on CT images. In reliance on the newest research [10], the detection of COVID-19 outcomes, are more credible when multiple methods are used together. In another try, a smart reading system for CT image, developed by Ping An Insurance Company of China Ltd [10], which can read and examine in a brief period of time.

As mentioned earlier, the typical way for diagnosing COVID-19 pneumonia is by using a CT scan. In [1], the authors have conducted a research intended to build a diagnose system depending on deep learning. The system works in diagnosing COVID-19 with super-resolution CT scan images, which helps radiologists with their work and assists to control the pandemic. It was fascinating that they could able to collect 46,096 unknown images from 106 admitted patients, including 51 cases with confirmed COVID-19 pneumonia. Also, 55 admitted control patients of other diseases have collected in the Renmin Hospital of Wuhan University in Wuhan city, Hubei Province of China. These CT scan Images were collected in order to develop and train their suggested model. Moreover, to assess and compare the accuracy of radiologists versus COVID19 pneumonia with the performance of their model.

It is crucial to point out that their suggested deep learning model has displayed an equal performance compared with the radiologists. Furthermore, the proficiency of radiologists could also develop while performing their clinical practice; which is importantly needed with such outbreaks status when cases rapidly increasing. Therefore, we can point out an argument here, owning such a diagnosing system provides multiple potentials to reduce the heavy responsibilities of radiologists, in addition, to provide an in-time diagnosis of COVID-19. Moreover, getting an early detection of such extremely infectious pandemic will further help in the plans of isolation and remedy, and finally helps countries in restraining and ending this pandemic [1].

There was extremely important argue in [1], related to the importance of CT scan Images in the detection of COVID-19 cases, which can be deemed a lot faster way in comparison with the traditional method utilizing the nucleic acid detection. Besides its effectiveness in diagnosing the disease, it can assess the severity level of pneumonia [11]. With all 140 laboratory-confirmed COVID-19 cases, CT results were reported positive. Moreover, CT scan was capable to identify these positive cases even within their early stage, which shows its efficiency [12,13]. On the other hand, the 5th version of COVID-19 diagnostic manual has reported by the National Health and Health Commission of China, the pneumonia radiographic features are integrated the clinical diagnostic standard in Hubei province.

Thereafter, for the purpose of highlight the effectiveness and proficiency of CT scan images in COVID-19 diagnosis, in a single day, as was reported, 14,840 direct cases of the infection were mentioned by 13th of February 2020

in Wuhan, this number including 13332 events of clinical diagnosis. All these proofs emphasized the importance of CT scan images in COVID-19 pneumonia diagnosis.

Considering the above-mentioned discussion, the researchers in [1] were primarily concentrating on accomplishing a model that could nearly diagnosis COVID-19 in the same way that radiologists do, but with shorter time. They could attain a similar performance to that expert radiologist with 65% lesser time required in diagnosing cases compared to in-clinic radiologist time. This will allow the patients and suspected cases to have a self-check system, as a result, so they save time and prevent direct contact that may consequently transmit the virus to the doctors and nurses.

On the other hand, the authors in [9] have mentioned that till the time of writing their report, there is no such automatic set of tools to clinically identify the stage of COVID-19 infection. For this goal, they have presented a deep learning (DL)-based system for the purpose of automatically segmenting and identifying the spots of infection of the COVID-19 patients, moreover to get a complete view of the lung to be isolated from the chest CT scan. To perform image segmentation, a DL-based network named VB-Net was suggested. Authors in [9] have proposed a 3-D convolutional neural network which corporations V-Net [14] together with the bottleneck structure presented in [15]. The adopted VB-Net is exhibited in Fig. 1, which requires two paths, the dashed boxes highlighting the utilized bottleneck elements within their V-shaped network.

In contrast, the authors in [9] proposed DL-based segmentation based the "VB-Net" neural network to extract all COVID-19 infection areas out of CT scan images. They have trained their gadget with 249 COVID- 19 cases and validated it with other 300 new COVID-19 cases. We can see that they have additionally brought a human-in-the-loop (HITL) method to be used at some point of coaching segment of their model to speedup procedure of manual definition of CT scan images. This strategy has helped radiologists all through the manner of tagging some footnotes on each case's scanned image. The proposed model was once implemented and tested across the conventional way of diagnosing the COVID19 to locate that the proposed system could minimize four minutes after three iterations of model updating, in contrast with the fully guide delineation; which normally may also take time diverse between 1 to 5 hours.

In other attempt, Adrian Rosebrock in [16] has proposed a DL model for diagnoses COVID19 the usage of Keras library and TensorFlow training platform. Using in total 50 images, which has been equally divided into 25 positive COVID-19 and different 25 negative X-rays images; they have build, educate and validate their model. Out of the performed experiments, their proposed model ought to diagnoses COVID-19 with average accuracy of 90-92%, which was once applied on their testing set with 100% sensitivity and 80% specificity due to the restricted data. It is additionally indispensable highlighting that having a exceedingly robust CONVID-19 diagnoses system should be executed with the aid of a multi-modal, which processes multi factors such as patient
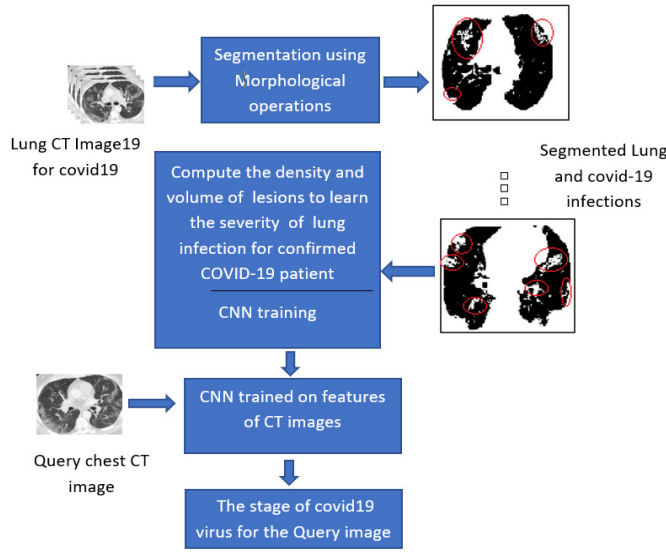
**Fig. 1** Phase transition diagram to classify the severity of COVID-19 patient.

vitals, population density, geographical location and some others. Therefore, having a diagnoses system relying only on X-ray images will not be that reliable to ending such excessive threat pandemic of COVID-19 [16].

3 The Proposed Approach

The proposed work contains two main phases; the first is calculating the weight for the virus stages scene using a Morphological approach for the CT scan images and the second is that Recognition Stages of the COVID-19 virus. CNN and kNN have been built with feeding CT scan images for classification purpose. A detailed explanation of the feature extraction algorithm is proposed in the next section, by considering the processing and clustering of the features.

3.1 Weight Calculation Phase

In this phase the pre-processing and clustering of the features are calculated and fed to the CNN. This process is utilized to segment the lesion of the infected lung and to enhance the image by spotting the infected lung regions. Subsequently, extracting important features for the lungs including lesion regions. These features Naïvely was merging to enhance the feature description of the infected lung. The steps of the proposed algorithm for extracting features vectors for the images is illustrated in the diagram is presented in Figure 2 the process will apply for the al images that needed to train the system on
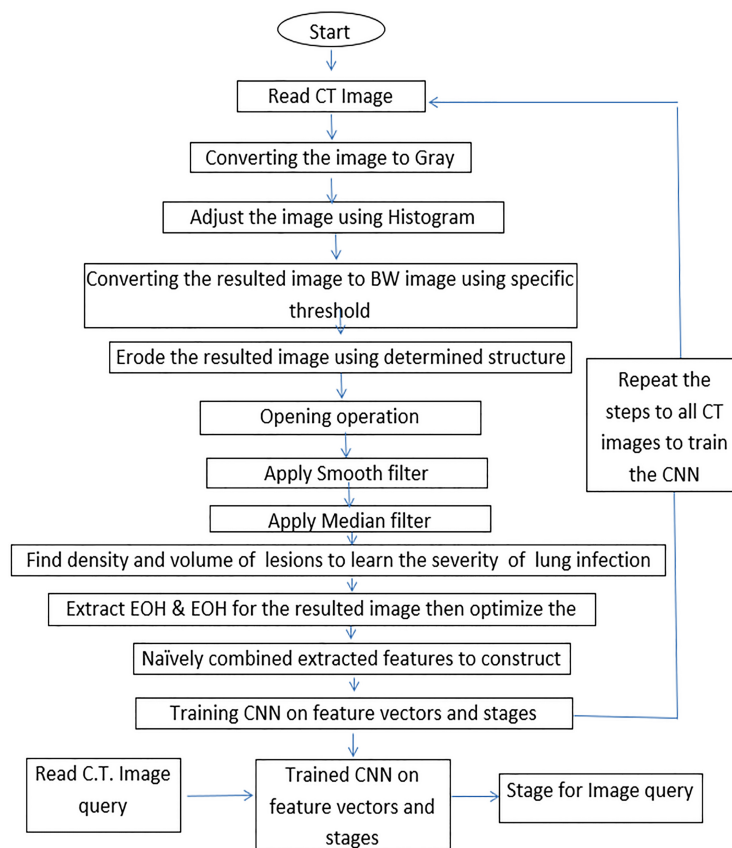
Fig. 2 Feature vectors extraction using morphological approach.

the cases inside the CT images to be classified the other cases based on these trained images.

Pre-processing mechanism is applied based on morphological approach by memorizing a set of features for CT scan images. These features are useful to diagnose the severity of the lung inflammation (mild, progressive, and sever). In addition, to normal CT scan cases based on weight represented by the white pixels disappeared in the two sides of lung.

The most of morphological operations have done to make the lesion regions in the lungs to be clearer, this will affect on the extraction of the feature vectors. The morphological operations on the original images and the result it will round the lesion regions caused by COVID-19 to be white pixels in the black lung regions, this let the calculations of the infected regions by the COVID-19 become very trivial. Figure 3 shows the image after applying the operations of the steps which are listed Figure 2. Two algorithms Histogram of Gradient (HOG) and Edge of Histogram (EOH) are extracted the significant features to be fed to the CNN. The volume and density of the lesions in the

left and right lungs depends on the duration where COVID-19 remain in the lungs. For example, when the infected lung image containing lesions of small volume and density indicates mild stages of the disease, while the large lesion regions of infected lungs indicate to sever stage for the confirmed COVID-19 patient.

3.1.1 Histogram of Gradient (HOG)

The CT scan images is subdivided into windows of images (cells), these windows represent small spatial windows or regions (cells), the histograms for each region will generate a local histogram of these partition gradient directions which will be accumulated together. The combination of histograms of the regions separately will form the whole image representation. Thereafter, the normalization for these local histograms could be established or applying some measurements on these local histograms to differentiate the local responses for the CT images considered. This leads to greater spatial region that can be exploited for normalized significant descriptors such as Histogram of Oriented Gradient [17]. The promise results and realization of these descriptors make the technique efficient to analyse these CT scan images and combining it naively with EOH features after applying morphological operations to extract more distinctive descriptor for lesions positions in the lung. Figure 4, Shows representation of features construction for the image. The CT image will be subdivided into cells in small areas, these areas connected to each other. Then, each cell containing pixels casting channel orientation based on resulted value of gradient computation. from 0 to 180 degrees, the channels of the histogram is spreading equally. Afterwards, the collection of these histograms of gradient directions gives the features vector for the image.
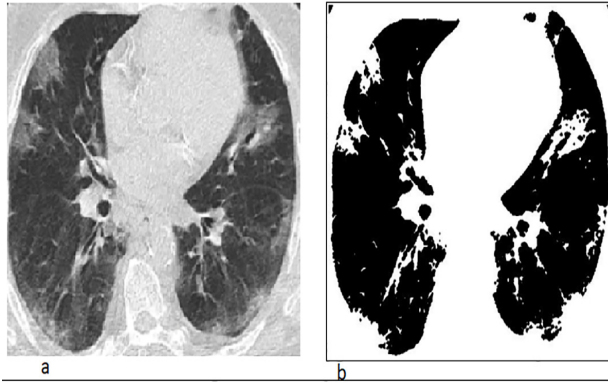
Fig. 3 CT scan images after Morphological Operation a: the original image b:the processed image.
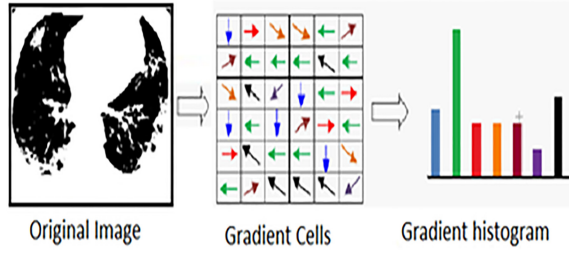
Fig. 4 CT scan images after Morphological Operation a: the original image b:the processed image.

3.1.2 Edge of Histogram (EOH) Method

The directions of edges inside the CT scan images give significant features that indicate some extra lesions in the lung, this coming from local orientation of edges inside CT image which can take it as descriptors. Edge orientation histograms (EOH) is exploited to extract most important features for describing the two side of the lung. The histograms are built by starting with calculating the orientation of the edges inside the image. This will be performed by filtering the edge inside the image. This filtering process uses two kernels: $[-1\ 0\ 1]$ And $[-1\ 0\ 1]^T$ to obtain filtered CT images are directed by dx and dy respectively. In this process, two arguments are considered, the direction α and the magnitude (M) for the pixels edges inside the image. These are computed by Equations 1 and 2 respectively.

$$\alpha = \arctan \frac{dy}{dx} \tag{1}$$

$$M = \sqrt{dy^2 + dx^2} \tag{2}$$

The bins related to edge orientation in the histogram are equally distributed through direction of the edge. Further, the number of histogram bins depends on the edge magnitude, then the calculation of the histogram will be related to the weighted vote [18]. Figure 5 illustrate the impact of the edge detection over COVID-19 confirmed cases and normal CT images. Combining HOG & EOH features Naively Learning visual elision recognition is constructing more variant features for the lung as expressed in equation 3:

$$NV(x, f_{1:nm}) = Hg(1:n), Eh(1:m) \tag{3}$$

Where x is the CT scan image processed morphologically, Hg is the histogram of oriented gradient, and Eh is the Edge histogram.

3.2 Convolutional Neural Network (CNN)

This section presents the design of the proposed CNN algorithm used to identify the level of severity of COVID-19 confirmed cases and details of pro-

posed CNN model implementations. The proposal includes two main algorithms which CNN architecture and a kNN algorithm.

The result of extracting feature vectors (NV), converted to the square matrix to be used as input to the CNN for training. For validation and testing of CNN, CT images of the dataset is processed according to the same steps illustrated in Figure 2, the attained features from morphological phase are prepared to feed CNN. In this paper 15 layers has been used to train the CNN to identify the level of severity of COVID-19 cases. These layers comprise, 1 layer for input , 3 of them as convolution layers with 3x3 filter size, and 2 layers for max pooling, 3 layers for batch normalization, 3 layers for rectified liner unit, 1 layer for fully connected, 1 layer for softmax and 1 layer for classification.

1-Input layer: This layer is responsible to read features, crop and resize CT images to make them ready for next layer [19]. The CT images are collected from different repositories. Thus, the mechanism of preprocessing on CT images make the results more accurate.

2- Convolution layers: these layers considered as major building blocks of CNN that they represent the applications of filters to input layer. In each layer, there are convolutions and kernel with a stable stride run over the C.T. complete image. Here the interior features of images will be extracted and passed to the pooling layer Equation(4) represents the convolution operation [20]:

$$y^j = max(0, b^j + \sum_{i=1} k^{ij} * x^i) \qquad (4)$$

Where k^{ij} is the convolution kernel between i_{th} output y map and input x map and the $*$ sign indicates to convolution operation.

3- Max Pooling Layer: This layer is a type of operation that typically added to CNN following individual convolution layers. This layer is shrinking dimensionality of big size CT images by reducing number of pixels in the output of previous convolution layers. In every window, maximum value will be retained; eventually the finest fits will be kept of every feature in the window. The max pooling process is illustrated in Equation 5.

Fig. 5 Edge description a: Processed image b: Edge detection image.

$$y_{j,k}^i = \max_{(m \geq 0, r \geq n)} y_{jxr+m, kxr+n}^i \tag{5}$$

4- Rectified Linear Units Layers (ReLU): In this layer, the negative numbers resulted of the pooling layer will be considered zeros. This layer make the CNN to reach steady state. It is represented by Equation 6:

$$F(x) = \max(0, x) \tag{6}$$

5- Batch normalization layer: This layer is considered as an important part of CNN to reduce number of training epochs. This subsequently stabilizing the learning process of CNN. It supplies an introduction for input feed-forwarding and calculating gradients with regard to the parameters [21].

6- Softmax layer: This function characterize a CNN to provide a solution for classification problem. It identifies the discrete probability P allocation for K classes. This can be indicated by $\sum_{i=1} pk$. Suppose x is the activation, and θ is weight parameters at the softmax layer, then o is the input to the softmax layer,

$$F(x) = \sum_{1=1}^{n-1} \theta_i * x_i \tag{7}$$

Subsequently

$$p_k = \frac{e^{o_k}}{\sum_{k=0}^{n=1} e^{o_k}} \tag{8}$$

Therefore, the prediction of the class would be:

$$\hat{y} = argmax pi \tag{9}$$

for $i \in 1...N$

3.3 k-Nearest Neighbor (kNN)

The kNN is simple and common method to categorize the natural scene objects using supervised learning algorithm. The process is built on one notion that the similarity between observations and groups which are relating to each other [22]. The algorithm has two phases; which are training and testing. The process of training phase is building the training dataset by means of a set of cases holding training pattern with its related class.

In the process of testing phase, the query begins with a given unlabeled point and the algorithm generates a set of k closest or nearest (results) scores corresponding to the trained input patterns. The similarity results of two feature vectors are estimated by using distance measure such as Euclidian or Manhattan. Finally the classification or decision is established by tagging the class of a tested pattern based on the majority voting method

4 Experimental Setup

This section presents the results of the proposed CNN and kNN architecture. The feature vectors are constructed naively using 89 features (81 extracted from HOG which 8 retrieved from EOH).

4.1 CT Scan Dataset

In this work, we used CT scan images dataset that are collectively available in a GitHub repository [23]. It was constructed by a group in [23] on several confirmed COVID-19 patients. The COVID-19 dataset comprises 186 CT scan images of pneumonia Lungs. In particular, the images are clearly shows the variation of lung inflammation during the course of COVID-19 patient between 1–21 Days. The scanned images size are 346x442x3 uint8. It is clear that we currently do not have plenty of COVID-19 images publicly available to the research community to conduct intense investigation and there is an immediate need to collect more radiology images which can be accessible by the research community.
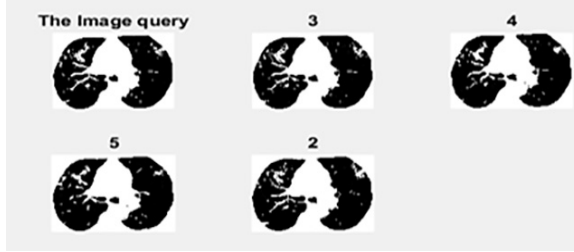
Fig. 6 The 4 distances between the Query image and the other images.

4.2 COVID-19 Severity Classification using kNN

The experiments are conducted, using kNN, on the CT images which are different from the training dataset for feature vector extraction. The Euclidean's distance is applied on the training and testing feature vectors. These distances are sorted out to determine the smallest distance by using tested CT image. Figure 6 shows the result of checking the query test image and the results of the kNN algorithm on the data set.

The precision p of the N retrieved CT images for the query Q CT image is expressed as follows:

$$P(Q, N) = \frac{|Ir|Rank(Q, Ir) \leq N \ such \ that \ Ir \in g(Q)||}{N} \qquad (10)$$

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        22               95.6522 %
Incorrectly Classified Instances       1                4.3478 %
Kappa statistic                        0.9345
Mean absolute error                    0.1141
Root mean squared error                0.2054
Relative absolute error               25.3949 %
Root relative squared error           43.0624 %
Total Number of Instances             23

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall  F-Measure  MCC     ROC Area  PRC Area  Class
                 1.000    0.067    0.889      1.000   0.941      0.911   0.967     0.932     a
                 0.857    0.000    1.000      0.857   0.923      0.898   0.964     0.948     b
                 1.000    0.000    1.000      1.000   1.000      1.000   1.000     1.000     c
Weighted Avg.    0.957    0.023    0.961      0.957   0.956      0.938   0.978     0.960

=== Confusion Matrix ===

 a b c   <-- classified as
 8 0 0 | a = a
 1 6 0 | b = b
 0 0 8 | c = c
```
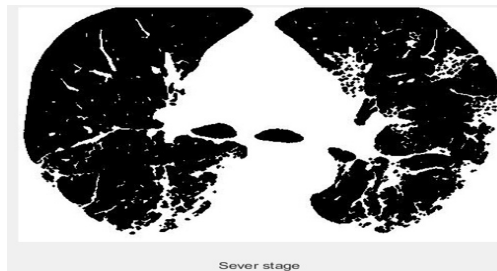
Fig. 7 Diagnosing the COVID-19 confirmed patient pneumonia stage.

Where $g(Q)$ represents the group stages of the virus for the query image and Ir is the retrieved CT image. The results for all created distance feature values are sorted; then, the minimum values are used to be the best matching virus stage.

4.3 COVID-19 Severity Classification using Deep Learning

Sever stage

Fig. 8 Diagnosing the COVID-19 confirmed patient pneumonia stage.

The CT images which are collected from GitHub dataset [**?**] are divided into two groups: testing and training datasets. Moreover, Weka Multilayer perceptron's deep learning R_j has been used to determine the accuracy of the proposed approach. Figure 7 shows the accuracy performance measure to find the level of severity of the lung infection. Further, Figure 8 classified the query CT image as a sever level of lung inflammation. This is because the volume and density of lesions are very high.

Table 1 Detection accuracy of CNN and kNN

Model	EOH	HOG	HOG +EOH
CNN	34.782 %	87.5 %	95.652 %.
kNN	78.260 %	87.5 %	91.304 %

4.4 Results & Discussion

In this work, the accuracy of pneumonia level detection during a course of confirmed COVID-19 patient has been calculated. To achieve this goal, morphological operations and Deep Learning are applied on the extracted feature vectors by using HOG and EOH. GitHub dataset repository is utilized to extract feature vectors. GitHub Covid19 dataset includes a set of CT scan images for confirmed COVID-19 patients, whereby the whole images containing two sides Lungs for the patients, the dataset consists of several stages (early stage, progressive, and sever) collected from different patients. The dataset composed of 186 CT images of size 372x556x3 uint8. Morphological operations are applied to convert the dataset to logical images., In this operation, noises within the images are filtered out to make the lesions and opacities on the infected lungs appear more clearly .The feature vector constructed from HOG and EOH to get higher invariant features for classification. When CNN is used to find the severity stage of the lung infection, we obtained the accuracy of 95.6 %. Further, when CNN applied on HOG features without EOH, we achieved 82.608 %, which is lower accuracy while EOH achieved less than the others, which is 34.782 %. Table 1 presents the detection accuracy of different features used for classification. The results assures the improvement of detection accuracy through changing the features and classification methods. As shown in Table 1, CNN outperforms KNN using HOG with EOH.

5 Conclusion

In this paper, CNN and kNN have been used to detect the level of pneumonia in the lung by using CT Scan images. This work is adopted to improve the accuracy of diagnosing lung inflammation severity in confirmed COVID-19 patient. we build a complete pre-processed dataset of CT scan images which are taken from GitHub repository. After extensive experiments, the results show that the proposed CNN with combined HOG and EOH significantly outperforms kNN. Furthermore, when morphological operation based on EOH is used, the result is not promising. This indicates that Multi-scale features for medical image recognition are better than a single scale features. The morphological operation is more useful to filter out noises and unnecessary features in the classification, which results to high accuracy of lung inflammation detection.

6 Compliance with Ethical Standards

Conflict of Interest: The authors declare that they have no conflict of interest. Moreover, this research was not funded by any funding agency.

References

1. J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, S. Hu, Y. Wang, X. Hu, B. Zheng et al., "Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography a prospective study," medRxiv, 2020.
2. M. Holshue, C. DeBolt, and L. S. First, "Novel coronavirus in the united states," N Engl J Med, vol. 2020, p. 31, 2019.
3. "Cnn health," https://edition.cnn.com/2020/03/11/health/coronaviruspandemic-world-health-organization/index.html, Accessed in March, 2020.
4. "Novel coronavirus (2019-ncov). world health organization." https://www.who.int/emergencies/diseases/novel-Coronavirus-2019., Accessed in March, 2020.
5. "Summary of probable sars cases with onset of illness from 1 november 2002 to 31 july 2003. world health organization." https://www.who.int/csr/sars/country/table20040421/en/., Published April 21, 2004.
6. D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong et al., "Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in wuhan, china," Jama, 2020.
7. National health commission of china. (March, 2020). [Online]. Available: http://www.chinadaily.com.cn/m/chinahealth/index.html
8. Canada broadcast cooperation. (2020, March). [Online]. Available: https://www.cbc.ca/player/play/1709650499517
9. F. Shan+, Y. Gao+, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, D. Shen, and Y. Shi, "Lung infection quantification of covid-19 in ct images with deep learning," arXiv preprint arXiv:2003.04655, 2020.
10. Bioworld. (2020, March). [Online]. Available: https://www.bioworld.com/articles/433530-china-uses-ai-in-medical-imaging-to-speed-up-covid-19-diagnosis
11. Z. Y. Zu, M. D. Jiang, P. P. Xu, W. Chen, Q. Q. Ni, G. M. Lu, and L. J. Zhang, "Coronavirus disease 2019 (covid-19): A perspective from china," Radiology, p. 200490, 2020.
12. C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu et al., "Clinical features of patients infected with 2019 novel coronavirus in wuhan, china," The Lancet, vol. 395, no. 10223, pp. 497–506, 2020.
13. N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei et al., "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in wuhan, china: a descriptive study," The Lancet, vol. 395, no. 10223, pp. 507–513, 2020.
14. F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in 2016 Fourth International Conference on 3D Vision (3DV). IEEE, 2016, pp. 565–571.
15. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
16. B. P. Roquette, H. Nagano, E. C. Marujo, and A. C. Maiorano, "Prediction of admission in pediatric emergency department with deep neural networks and triage textual data," Neural Networks, 2020.
17. Y.-G. Jiang, C.-W. Ngo, and J. Yang, "Towards optimal bag-of-features for object categorization and semantic video retrieval," in Proceedings of the 6th ACM international conference on Image and video retrieval, 2007, pp. 494–501.
18. I. K. Timotius and I. Setyawan, "Evaluation of edge orientation histograms in smile detection," in 2014 6th International Conference on Information Technology and Electrical Engineering (ICITEE). IEEE, 2014, pp. 1–5.

19. N. Sharma, V. Jain, and A. Mishra, "An analysis of convolutional neural networks for image classification," Procedia Computer Science, vol. 132, pp. 377–384, 2018.

20. Z. Chen, A. Jacobson, N. S ¨ underhauf, B. Upcroft, L. Liu, C. Shen, I. Reid, and M. Milford, "Deep learning features at scale for visual place recognition," in 2017 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2017, pp. 3223–3230.

21. F. Schilling, "The effect of batch normalization on deep convolutional neural networks," 2016.

22. A. Bin Abdullah, Supervised learning algorithms for visual object categorization. Universiteit Utrecht, 2010.

23. J. P. Cohen, P. Morrison, and L. Dao, "image data collection," arXiv 2003.11597, vol. 6, 2020.