COVID-19 Diagnostic System Using Medical Image Classification and Retrieval: A Novel Method for Image Analysis

MAHER ALRAHHAL* AND SUPREETHI K.P.
Jawaharlal Nehru Technological University Hyderabad, College of Engineering, Computer Science Department, Hyderabad 500085, India
*Corresponding author: maherrahal92@gmail.com

With the rapid increase in the number of people infected with COVID-19 disease in the entire world, and with the limited medical equipment used to detect it (testing kit), it becomes necessary to provide another detection method that mainly relies on Artificial Intelligence and radiographic Image Analysis to determine the disease infection. In this study, we proposed a diagnosis system that detects the COVID-19 using chest X-ray or computed tomography (CT) scan images knowing that this system does not eliminate the reverse transcription-polymerase chain reaction test but rather complements it. The proposed system consists of the following steps, starting with extracting the image’s features using Visual Words Fusion of ResNet-50 (deep neural network) and Histogram of Oriented Gradient descriptors based on Bag of Visual Word methodology. Then training the Adaptive Boosting classifier to classify the image to COVID-19 or NOTCOVID-19 and finally retrieving the most similar images. We implemented our work on X-ray and CT scan databases, and the experimental results demonstrate the effectiveness of the proposed system. The performance of the classification task in terms of accuracy was as follows: 100% for classifying the input image to X-ray or CT scan, 99.18% for classifying X-ray image to COVID-19 or NOTCOVID-19 and 97.84% for classifying CT scan to COVID-19 or NOTCOVID-19.

Keywords: medical Image Analysis; image classification and retrieval; deep learning; Histogram of Oriented Gradient; Bag of Visual Word; Adaptive Boosting classifier; COVID-19

Received 29 October 2020; Revised 12 February 2021; Editorial Decision 29 March 2021
Handling editor: Fionn Murtagh

1. INTRODUCTION

In the last months of the year 2019, the COVID-19 disease appeared in China, specifically from Wuhan [1]. The common onset symptoms for COVID-19 patients were fever, cough, myalgia or fatigue. COVID-19 complications are severe infections of the respiratory system, acute heart injury and the advanced injuries lead to entering the central care unit and may lead to death. The current tests are mostly based on reverse transcription-polymerase chain reaction (RT-PCR) [2]. One of the main obstacles in controlling the spread of COVID-19 disease is the disadvantages associated with RT-PCR, which are summarized by the following factors:

- The limited efficiency of the test, as its accuracy does not exceed 71% [3,4], meaning that a person may be infected, but his test result will be negative (false result), and therefore, the person will not receive treatment; this leads to the rapid infection transmission among healthy people.
- The relatively large time for the RT-PCR test, where it ranges between 6–8 hours [5], and this has a negative impact on limiting the speed of the spread of COVID-19.
- The high cost of the RT-PCR test [5].
- The shortage in medical equipment needed to perform RT-PCR tests, and thus, the inability to perform the
TABLE 1. Shows the effect of COVID-19 disease on the patient’s lungs by X-ray imaging and CT scan.

A CT scan showing white spots on the lower corners of the lungs, which doctors call ‘the opacity of the ground glass’, and indicates a partial fullness of the alveoli in the lung.

A CT scan of the lung 3 days after the first, the white spots appeared to be increasing.

CT images of the lung of a 33-year-old woman with COVID-19.

Chest x-ray of the patient; she is a woman in her seventies in hospital, with COVID-19 infection (first day of admission).

Chest x-ray of the patient at 8-day, showing the ground-glass opacity at the bases of the lung (white arrows). Consolidation is also evident in the periphery of the left upper and middle regions (arrows marked).

An x-ray of the lung of a 70-year-old woman with COVID-19.

required numbers of daily tests that help to contain COVID-19.

Studies have shown that most cases of COVID-19 disease develop to mild respiratory and constitutional symptoms such as fever, cough, dyspnea, myalgia and fatigue [6,7]. The symptoms vary according to the degree of infection that may be a mild, moderate or severe infection. Studies have also shown that the examination of the chest X-ray reveals the changes that occur on the patient’s lungs 4–5 days after the onset of symptoms [8], whereas the computed tomography (CT) scan reveals these changes 2 days after the onset of symptoms [9,10] or when symptoms begin to appear [11]. The effect of the COVID-19 on the patient’s lungs in the case of an X-ray and CT scan is illustrated in Table 1.

These studies have led to conducting researches on the ability of radiography; with both types of X-ray and CT imaging, to detect suspected cases of COVID-19 [12–14]. Although X-ray images and CT scans may help in the early examination of suspected cases, the images of pneumonia caused by COVID-19...
disease are very similar to inflammatory lung diseases [15], and the human eye falls short of that. This is the important reason that pushes us to use Artificial Intelligence and radiographic Image Analysis to address this shortcoming.

Deep learning is a part of Artificial Intelligence that simulates the work of the brain and neurons. It is used in many areas and has proven a great success in disease diagnosing through the analysis of X-rays, CT and magnetic resonance images. Because of these analyses, diseases such as breast, skin and brain tumors were discovered and diagnosed [16–21].

The main contributions in our work are summarized in the following:

- Building a diagnosis system that is characterized by:
  - Has the ability to detect COVID-19 disease for both types of medical chest images (X-ray or CT scan), i.e. the system detects the type of the medical image (X-ray or CT) then diagnoses it whether infected with COVID-19 or not.
  - Has the ability to retrieve the most similar images to the query (tested, scanned) image, which grants the system the following merits:
    - Evaluates the diagnosis of the scanned images by comparing the class of majority retrieve images with the diagnosis of the image.
    - Assists the specialist physician through using the similar treatment protocol of retrieved images.
  - Both diagnosing (classification) and retrieving tasks share the feature extraction step that considers the essential step in the whole work. For that, we proposed two novel methods for feature extraction as the following:
    - **Proposed1** method that relies on Visual Words Fusion of ResNet-50 and Histogram of Oriented Gradient (HOG) descriptors based on Bag of Visual Word (BoVW) methodology and used those features to train Adaptive Boosting classifier and retrieve the most similar images of the query images.
    - **Proposed2** method that relies on Features Fusion of ResNet-50 and HOG descriptors based on BoVW methodology and used those features for classification and retrieving tasks.

Furthermore, it is necessary to mention the following points:

- This study does not eliminate RT-PCR testing but rather complements it. In the cases where respiratory symptoms of the disease appear, it is better to use radiographic techniques based on Artificial Intelligence and Image Analysis. However, in cases where the respiratory symptoms not exist, the RT-PCR test remains the best way to detect the disease.
- The system trained to be able to diagnose COVID-19 disease using X-ray images or CT scans, and the choice of the imaging method used to detect COVID-19 depends on multiple factors, including:
  - Availability of medical imaging devices.
  - Determine the period of onset of symptoms.
  - Clinical examination of the patient and the final decision is up to the specialist physician.
- We implement our work on both types of images (X-ray and CT scan), knowing that X-ray is a faster, easier, cheaper, less harmful method than CT and detects the changes that occur to the COVID-19 patient’s lungs 4 or 5 days after the onset of symptoms. However, using CT scanning, we can examine the very soft structure of the active body part and show the soft tissues and internal organs more clearly and accurately than X-ray images. Moreover, a CT scan detects the changes in the infected lung with COVID-19 disease since the appearance of symptoms. Because of the pros and cons of each imaging method, we have developed COVID-19 detection methods for both medical images (X-ray and CT).

2. RELATED WORK

Many studies addressed the use of deep learning for diagnosing COVID-19 disease. We reviewed a few of those studies. Those studies are summarized as following:

Xu et al. [22] started their proposed methods by preprocessing the CT scan images, then used 3D CNN model to segment the images into multiple image cubes. After that, they used the ResNet-18 network for feature extraction and Bayesian function to classify all image patches into the following types: COVID-19, Influenza-A-viral-pneumonia and normal. They implemented their work on CT scan images and the database description was as follows: 219 images of COVID-19 category, 224 images of viral pneumonia category and 175 images of a normal category.

Training deep learning models on a small dataset can easily lead to overfitting; Yang et al. [2] addressed this problem using two approaches: transfer learning and data augmentation. The authors used a large collection of chest X-ray images to pretrain a DenseNet, and then fine-tune this pretrained network on the COVID-19 dataset. They implemented their work on CT scan images and the database description was as follows: 275 images of COVID-19 category and 195 images of a normal category.

Wang et al. [15] modified the typical Inception network in the last fully connected (FC) layers. They reduced the
dimension of features before sent them to the final classification layer. They used a modified Inception network (M-Inception) for feature extraction; then forward those features to the ensemble classifier based on the decision tree and Adaboost to improve the classification accuracy. They implemented their work on CT scan images and the database description was as follows: 195 images of COVID-19 category and 258 images of normal category.

Song et al. [23] proposed a method that consists of three main steps: first, they extracted the main regions of the lungs and filled the blank of lung segmentation with the lung itself to avoid noises caused by different lung contours. Second, they designed a Details Relation Extraction neural network (DRE-Net) to extract the top-K details in the CT images and obtain the image-level predictions. Third, the image-level predictions were aggregated to achieve patient-level diagnoses. The database description as follows: 88 images of COVID-19 category, 101 images of viral pneumonia category and 86 images of a normal category.

Zheng et al. [24] proposed a (DeCoVNet) that consists of three steps: the first one was network stem, and the second one composed of two 3D residual blocks (ResBlocks). The third step was a progressive classifier (ProClf), which contained three 3D convolution layers and an FC layer with the softmax activation function. They implemented their work on CT scan images and the database description was as follows: 313 images of COVID-19 category and 229 images of a normal category.

Li et al. [25] proposed COVNet model for COVID-19 detection that depends on RestNet-50 as the backbone, which takes a series of CT slices as input and generates features for the corresponding slices. Then, the extracted features from all slices are combined by a max-pooling operation. The final feature map is fed to an FC layer and the softmax activation function to generate a probability score. The database description was as follows: 1296 images of COVID-19 category, 1736 images of pneumonia category and 1325 images of a normal category.

Narin et al. [26] proposed COVID-19 detection models based on ResNet-50, InceptionV3 and Inception-ResNetV2 networks. They implemented their work on X-ray images and their database description was as follows: 50 images of the COVID-19 category and 50 images of a normal category.

At the end of the related works section, we reviewed a few works that discussed the COVID-19 detection using deep learning techniques. However, the importance of our work comparing to the literature works comes from the following factors; first, all the previous work used only one type of medical image. However, our work discusses and uses both types of medical images (X-ray and CT). Second, our work used the image retrieval task to evaluate the diagnostic outcomes and make the best decision. Third, our work proposed novel feature extraction methods that combined the deep learning (ResNet-50) with the handcraft features (HOG) and used BoVW to reduce the feature dimensions.

3. METHODOLOGY

Our proposed diagnosis system contains two main tasks, namely:

- Classification (Diagnosis) Task: This task is responsible for detecting the type of the medical image (X-ray or CT) and diagnosing it whether infected with COVID-19 or not.
- Retrieval Task: This task is responsible for retrieving the most similar images to the scanned image by comparing the query image with the images archive (database) of previous patients to assist the specialist physician in the treatment.

Both tasks intersect in the feature extraction step, which is the essential step in the whole work. Therefore, in this section, we discussed the proposed methods of feature extraction and the ideas related to them. The architecture scheme of the diagnosis system is illustrated in Figure 1.

3.1. Feature extraction method using visual word fusion of ResNet-50 and HOG descriptors based on BoVW (Proposed1)

The proposed method that relies on visual word fusion of ResNet-50 and HOG descriptors based on BoVW consists of several steps, summarized in Figure 2, as following:

**Step1**: Choose randomly 40% images from each class in the database to build the dictionary.

**Step2**: Preprocessing step for enhancing the image intensity where this type of images had a narrow range of intensity level.

**Step3**: Extract the feature vectors from the enhancement image after resizing it to $224 \times 224$ using:

- The layer ‘avg_pool’ of the ResNet-50 [27] and we refer to this feature vector as $f_{ResNet50}$.
- The HOG method [28] and we refer to this feature vector as $f_{HOG}$.

**Step4**: For compacting the feature vectors, we apply the K-Means++ clustering (which describes as a smart centroid initialization technique comparing to K-Means) [29] separately for each type of the feature vector ($f_{ResNet50}, f_{HOG}$), knowing that the centers of the clusters compose the visual words. The set of visual words constitute the vocabulary (sometimes called dictionary or bag) that forms by merging the visual words as the Equations (1) and (2):

$$VOC_{ResNet50} = \{ VW_{ResNet50_1}, VW_{ResNet50_2}, VW_{ResNet50_3}, \ldots, VW_{ResNet50_k} \}$$

(1)
FIGURE 1. Illustrates the main tasks of our work for (a) X-ray, (b) CT scan images.

\[ \text{VOC}_{\text{HOG}} = \{ VW_{\text{HOG}1}, VW_{\text{HOG}2}, VW_{\text{HOG}3}, \ldots, VW_{\text{HOG}M} \} \]  
(2)

where the \( VW_{\text{ResNet50}} \) refers to the first visual word (first cluster center) that resulted from applying k-means++ on the \( f_{\text{ResNet50}} \). \( N \) is the size of the vocabulary \( \text{VOC}_{\text{ResNet50}} \) and \( M \) is the size of \( \text{VOC}_{\text{HOG}} \).

**Step 5**: Fuse both vocabularies \( \{ \text{VOC}_{\text{ResNet50}}, \text{VOC}_{\text{HOG}} \} \) by concatenating \( \text{VOC}_{\text{ResNet50}} \) with \( \text{VOC}_{\text{HOG}} \) vertically as represented in Equation (3):

\[ \text{VOC}_{\text{fusion}} = \{ \text{VOC}_{\text{ResNet50}}, \text{VOC}_{\text{HOG}} \} \]  
(3)

After building the vocabulary (dictionary) using steps 1–5, the process of the feature extraction for a query image starts with step 2 (preprocessing), step 3 (feature extraction using ResNet-50 and HOG), then clustering the output of step 3 and builds two histograms using VOC-fusion where the first one related to \( f_{\text{ResNet50}} \) and the second one related to \( f_{\text{HOG}} \). Finally, concatenates the resulted histograms to form the feature vector in its final form that will use in the classification and retrieval tasks.

3.2. Feature extraction method using feature fusion of ResNet-50 and HOG descriptors based on BoVW (Proposed2)

The proposed method that relies on feature fusion of ResNet-50 and HOG descriptors based on BoVW consists of several steps, summarized in Figure 3, as following:

**Step 1**: Choose randomly 40% images from each class in the database to build the dictionary.

**Step 2**: Preprocessing step.

**Step 3**: Extract the feature vectors from the enhancement image after resizing it to \( 224 \times 224 \) using:

- The layer ‘avg_pool’ of the ResNet-50 to get \( f_{\text{ResNet50}} \).
- The HOG method to get \( f_{\text{HOG}} \).

**Step 4**: Fuse the features by concatenating the \( f_{\text{ResNet50}} \) and \( f_{\text{HOG}} \) to get \( f_{\text{fusion}} \).

**Step 5**: For compacting the feature vectors, we apply the K-means++ clustering algorithm on \( f_{\text{fusion}} \), knowing that the centers of the clusters compose the visual words. The set of visual words constitute the vocabulary, which formed by merging the visual words as the Equation (4):

\[ \text{VOC}_{\text{fusion}} = \{ VW_{\text{fusion}1}, VW_{\text{fusion}2}, VW_{\text{fusion}3}, \ldots, VW_{\text{fusion}N} \} \]  
(4)

where the \( VW_{\text{fusion}1} \) refers to the first visual word (first cluster center) that resulted from applying k-means++ on the \( f_{\text{fusion}} \) and \( N \) is the size of the vocabulary \( \text{VOC}_{\text{fusion}} \).

After building the vocabulary using steps 1–5, the process of the feature extraction for a query image starts with step 2 (preprocessing), step 3 (feature extraction using ResNet-50 and HOG) and step 4 (fusion features), then clusters the fused...
features and builds the histogram using $\text{VOC}_{\text{fusion}}$, which will be the feature vector in the final form.

3.3. Preprocessing step

The preprocessing step is an essential step in the medical system that aims to enhance the image in terms of contrast, brightness, removing noise and adjusting the size. In our proposed methods, the preprocessing step consists of two steps as illustrating in Figure 4:

- High-frequency emphasis filter (HFEF) for sharpening an image by emphasizing the edges.
- Histogram equalization for maximizing the image contrast and mapping the lowest and highest intensity pixels to 0 and 1, respectively.

In more details, the preprocessing step includes the following steps:

1. Convert the input image to a frequency domain by applying the Discrete Fourier Transform.
2. Apply the HFEF with Gaussian high-pass filter (GHPF) on the transformed image using the Equations (5) and (6) as the following:

$$H_{\text{GHPF}}(u,v) = 1 - e^{-D^2(u,v)/2D_0^2}$$  \hspace{1cm} (6)

where $k_1$ offsets the value of the transfer function $H_{\text{GHPF}}$, whereas $k_2$ controls the contribution of high frequencies, and $D_0$ is the cutoff frequency. In our work we used $k_1 = 0.5$ and $k_2 = 0.75$.

1. Computing the Inverse Discrete Fourier Transform to obtain the spatial domain representation of the processed image.
2. Performing the histogram equalization [30] to improve the image’s contrast by spreading out the histogram.

3.4. ResNet-50 network

ResNet-50 is a deep neural network [27] that consists of 50 layers, and more than 23 million neurons and takes $224 \times 224$ color image as input. This network trained on the famous ImageNet database [31], which includes more than 1 million and three hundred thousand images distributed on thousand classes. Therefore, it trained in a very rich group of images and visual features. Neural networks tend to be deeper and deeper so that they can extract and analyze the smallest details in the image and represent them, thereby increasing the network’s accuracy in the classification task.

However, increasing the depth of the network leads to the arrival of the network accuracy to saturation and often decreases the network accuracy. For solving those issues, it is necessary to use new technology in the structure of the network, which is Residual Learning. The Resent-50 network has gained its name from this Residual learning. We can achieve residual learning by utilizing skip connections [32] that define as extra connections between nodes in different layers of a neural network that skip one or more layers of nonlinear processing.

To train the network, we divided the image databases into 80% for training and 20% for testing. Matlab R2018b programming language used to train the model. All experiments performed on the desktop with core i7 and 8GB RAM of Windows 7 operating system. Regarding training options, we used the stochastic gradient descent with momentum optimizer. The batch size, learning rate and the number of epochs were experimentally set to 20, 1e-4, 10 respectively, and a shuffle of the training data achieved before each training epoch. We accomplished data augmentation using geometric transformation techniques applied to the original images, such as...
cropping, rotating, reflection around $X (Y)$ axis, horizontal (vertical) scaling, shearing and translation. The aim of data augmentation is to increase the number of training samples, preventing data overfitting in neural networks [33,34] and improving performance in imbalanced class problems [35].

3.5. HOG method

HOG feature extraction [28] involves the following steps:

- Convolve the image with two filters that are sensitive to horizontal and vertical rightness gradients. These filters capture edge, contour and texture information.
- Subdivide the image into cells of a predetermined size and compute a histogram of the orientations within each cell.
- Normalize the histogram in each cell by comparing it with the block of neighboring cells. This further suppresses the effect of illumination across the image.
- Construct a one-dimensional feature vector from the information in each cell.

4. CLASSIFICATION TASK

In the beginning, the patient (who wants to know whether has COVID-19 disease or not) clinically examines by the specialist doctor to determine the symptoms and the appropriate radiographic method. If a person does not have any respiratory symptoms, then the RT-PCR test remains the best method. When there are respiratory symptoms of the disease, it is helpful to use Artificial Intelligence and Image Analysis to detect COVID-19 disease rather than RT-PCR because it reduces the need it as well as it is faster, cheaper and more accurate than the RT-PCR test as will be proved in this study.

Regarding the image databases, we dealt with the two types of images (chest X-ray and CT) as the following:

- The chest X-ray image is classified into one of three types: NORMAL, PNEUMONIA or COVID-19, and we have two classification levels as follows:
  - The first level aims to classify the input image into COVID-19 or NOTCOVID-19 (i.e. NORMAL or PNEUMONIA).
  - We proceed to the second level of classification if the image classified as NOTCOVID-19 on the first level. The second level determines whether the image is NORMAL or PNEUMONIA.

- The chest CT scan image is classified as COVID-19 or NOTCOVID-19, i.e. we have here one classification level.
- If the image is classified as COVID-19, the person must undergo immediate treatment.
- It should note that regarding the COVID-19 detection using a CT scan, we used only two types (i.e. COVID-19 and NOTCOVID-19) because we did not find CT scan
resources of PNEUMONIA. However, and if it will be available, the proposed system can be expanded.

The conditional flow chart of the classification task is shown in Figure 5.

4.1. Ensemble classifier used in classification task

In this work, we used an Adaptive Boosting-based Ensemble classifier [36], which aims to build a strong classifier that learns from the mistakes of the weak learners in the ensemble, i.e. there are a number of iterations (weak classifiers) combined to create a strong classifier. Knowing that each iteration is used to update the weights by comparing the predicted and actual training data. Suppose we want to build an ensemble classifier that has three weak classifiers (WC1, WC2 and WC3) based on the decision trees algorithm. This classifier works as shown in Figure 6.

4.2. Databases used in the classification task

For the classification task, we used four different databases; knowing that, we divided each database into 80% for training and 20% for testing. The description of each database is as follows:

4.2.1. First database

We used this database to train the classifier to be able to classify the input image into an X-ray or CT scan classes. The database consists of the following classes:

- Chest CT scan class that consists of 300 images, selected randomly from [37].
- Chest X-ray class that consists of 300 images, selected randomly from [38].

4.2.2. Second database

Regarding the X-ray classification, we have two classification levels with two different databases. The aim is to classify the chest X-ray images into one of three classes (NORMAL, PNEUMONIA and COVID-19).

The first level of classification aims to classify the chest X-ray image to COVID-19 or NOTCOVID-19. Therefore, the image database here contains the following classes:

- COVID-19 class that contains 225 images available in [38].
- NOTCOVID-19 class that contains 1000 images selected from the Kaggle repository [39] and those images are as follows: 500 NORMAL images and 500 PNEUMONIA images.
FIGURE 6. Illustrating example of the Adaptive Boosting classifier.

| Database No. | Description                                      | No. of classes | Total no. images | Classes label   | No. of images in each class |
|--------------|--------------------------------------------------|----------------|------------------|-----------------|----------------------------|
| 1            | Used to classify images into X-ray or CT scan.   | 2              | 600              | X-ray           | 300                        |
|              |                                                  |                |                  | CT              | 300                        |
| 2            | Used for first level of X-ray classification.    | 2              | 1225             | COVID-19        | 225                        |
|              |                                                  |                |                  | NOTCOVID-19     | 1000                       |
| 3            | Used for second level of X-ray classification.   | 2              | 1000             | NORMAL          | 500                        |
|              |                                                  |                |                  | PNEUMONIA       | 500                        |
| 4            | Used for CT scan classification.                 | 2              | 700              | COVID-19        | 350                        |
|              |                                                  |                |                  | NOTCOVID-19     | 350                        |

4.2.3. Third database

If the chest X-ray image is classified as NOTCOVID-19, we proceed to the second level to classify the X-ray images to either NORMAL or PNEUMONIA. This type of classification called binary classification, and the image database consists of the following classes:

- NORMAL class that contains 500 images.
- PNEUMONIA class that contains 500 images.

Those images have chosen randomly from the open-source Kaggle repository [39].

4.2.4. Fourth database

We used this database to make the system able to classify the CT scan images into COVID-19 or not (i.e. binary classification). Therefore, it consists of two classes as the following:

- COVID-19 class that contains 350 images.
- NOTCOVID-19 class that contains 350 images.

Those images obtained from the open-source GitHub repository [37]. The details of the four databases illustrate in Table 2.

4.3. Performance measures of the classification task

The main performance measure [40–42] in the classification task is the confusion matrix that defined as a table with four different combinations of predicted and actual values as the following:

- True positive (TP), i.e. the image is classified as COVID-19, and it is actually COVID-19.
- True negative (TN), i.e. the image is classified as NOTCOVID-19, and it is actually NOTCOVID-19.
- False positive (FP), i.e. the image is classified as COVID-19; however, it is NOTCOVID-19 (known as type 1 error).
- False negative (FN), i.e. the image is classified as NOTCOVID-19, however, it is COVID-19 (known as type 2 error).

From the confusion matrix, we can derive five important measures that are, Sensitivity (Recall), Specificity, Precision, F-measure and Accuracy. For all metrics, the higher are the better. Those metrics calculated as follows:
• Sensitivity (also known as Recall) calculates using Equation (7):

\[
Recall = \frac{TP}{TP + FN}
\]  

(7)

• Specificity calculates using Equation (8):

\[
Specificity = \frac{TN}{TN + FP}
\]  

(8)

• Precision calculates using Equation (9):

\[
Precision = \frac{TP}{TP + FP}
\]  

(9)

• F-measure calculates using Equation (10):

\[
F\text{−measure} = \frac{2 \times Recall \times Precision}{Recall + Precision}
\]  

(10)

• Accuracy calculates using Equation (11):

\[
Accuracy = \frac{TP + TN}{TP + TN + FP + FN}
\]  

(11)

4.4. Effect of vocabulary size on the classification performance

The vocabulary (dictionary) size is an important parameter in our proposed methods. As we increase the size of the vocabulary at a certain level, the performance of the system increases. However, the larger sizes of the vocabulary lead to overfitting that decreases the performance. In this experiment, we studied the effect of the different vocabulary sizes on the performance analysis in terms of the classification accuracy of the proposed methods. Regarding the X-ray (first and second classification level) and CT databases, we have used different vocabulary sizes (i.e. 50, 200, 400, 600, 800, 1000 and 1200) for measuring the performance of the proposed methods to find out where the proposed techniques are generating the best results.

According to the experimental results shown in Figure 7a–c, we conclude the following:

• The classification accuracy of the Proposed1 method start increasing from the 50-vocabulary size until 600 then start decreasing, because of the overfitting. Knowing that in the Proposed1 method, we fused two vocabularies (the first vocabulary regards the ResNet-50 and the second one regards the HOG). Moreover, the vocabulary sizes of the Proposed1 method present in Figure 7a–c are the size of each vocabularies. Therefore, the feature vector length generated with vocabulary size \( x \) is \( 2x \).

• As well as, the classification accuracy of the Proposed2 method starts increasing from the 50-vocabulary size until 1000 then starts decreasing. As illustrated in Figure 7a–c, the accuracy of the Proposed2 method performs better than Proposed1 method on 1000-vocabulary size due to more compact representation of the visual contents of the images because the size of the vocabulary is twice less compared with visual word fusion technique (Proposed1).

• The Proposed1 method performs better comparing to Proposed2 method for almost all vocabulary sizes in the X-ray (first and second level) and CT scan databases. Moreover, we got the best accuracies of the Proposed1 method on 600-vocabulary size, and the results are as the following:
  . 99.18% for the X-ray database (first classification level).
  . 98.50% for the X-ray database (second classification level).
  . 97.86% for CT scan database.

• The best accuracies of the Proposed2 method obtained on 1000-vocabulary size, and the results are as the following:
  . 98.78% for the X-ray database (first classification level).
  . 96.5% for the X-ray database (second classification level).
  . 96.43% for CT scan database.

4.5. Results of the classification task

For all classification tasks, we used our proposed methods for feature extraction and the Adaptive Boosting Ensemble classifier for the classification. As we concluded in the previous experiment, the best accuracy of the visual words fusion of ResNet-50 and HOG descriptors based on BoVW methodology (Proposed1) was on 600-vocabulary, whereas the best classification accuracy of the features fusion of ResNet-50 and HOG descriptors based on BoVW methodology (Proposed2) obtained on 1000-vocabulary size.

In this experiment, and regarding the X-ray (first and second classification level) and CT databases, we will compare our
FIGURE 7. (a) Comparison of the classification accuracy between our methods: Proposed1 and Proposed2, for the X-ray DB (first level). (b) Comparison of the classification accuracy between our methods: Proposed1 and Proposed2, for the X-ray DB (second level). (c) Comparison of the classification accuracy between our methods: Proposed1 and Proposed2, for the CT scan database.

**Proposed1**: Visual Words Fusion of ResNet-50 and HOG descriptors based on BoVW methodology.

**Proposed2**: Features Fusion of ResNet-50 and HOG descriptors based on BoVW methodology.
proposed methods (with 600-vocabulary size for Proposed1 method and 1000-vocabulary size for Proposed2 method) in terms of sensitivity, specificity, precision, F-measure and accuracy, with the following methods:

- Standalone ResNet-50 feature technique based on the BoVW model (we referred to this method as ResNet-50 + BoVW) that depends on the deep learning technique.
- Standalone HOG feature technique based on the BoVW (HOG + BoVW) that is handcraft feature extraction method.
- Recent methods referred with the reference [2,15,22], which focused on the deep learning-based COVID-19 detection.

We used 600-vocabulary size for implementing both ResNet-50 + BoVW and HOG+BoVW because it seems to be an intermediate value between enhancing the system’s performance and the overfitting. The experimental results shown in Figure 8a–c proved that our proposed methods that used both deep learning and handcraft features (Proposed1 and Proposed2) improved the performance of classification as compared with standalone ResNet-50 based on BoVW, standalone HOG descriptors based on BoVW and other recent methods. Knowing that the Proposed1 method gives the best results for the aforementioned databases, as the following:

- The accuracy of classifying the X-ray image to COVID-19 or NOTCOVID-19 is 99.18%; the first level of X-ray classification (second database).
- The accuracy of classifying the X-ray image to NORMAL or PNEUMONIA is 98.5%; the second level of X-ray classification (third database).
- The accuracy of classifying the CT scan image to COVID-19 or NOTCOVID-19 is 97.86%; the fourth database.

The classification accuracies using the Proposed2 method for the second, third and fourth databases are 98.78%, 96.50% and 96.43%, respectively. The sensitivity, specificity, precision, F-measure and accuracy for second, third and fourth databases illustrated in Figure 8a–c.

4.6. Notes

- In the first database, we used only 600 chest images; 300 images for X-ray and other 300 images for CT scan. Knowing that, we got 100% accuracy for both Proposed1 (on 600-vocabulary size) and Proposed2 (on 1000-vocabulary size), and the accuracy remains constant while increasing the database size. Moreover, we got the same accuracy (100%) for other comparing methods.
- This system will store in a central computer (server); thus, differentiating between a chest X-ray and a chest CT image is an important primary task. Therefore, after the radiographic image is taken, it will be sent to the server, which in turn performs its work and sends the results back to the sender.
- It should note that all performance criteria used in the classification tasks are important because our X-ray database is imbalanced. Therefore, we did not rely only on the accuracy, but also on other criteria to demonstrate the importance of the work.
- It should note that the COVID-19 images used in our work related to COVID-19 disease, without a doubt, do not include all cases of the disease due to the novelty of the disease. However, our work can consider as a smart prototype that tested it on the available images and we obtained promising results. Thus, in the event of a large image database that includes all cases of COVID-19 diseases, we will obtain satisfactory results.
- This system provides the diagnosis result in a few seconds (depending on system specifications). Moreover, all we need is a radiograph scan that is also cheaper than the RT-PCR test, as well as the accuracy of this system is better than the RT-PCR test.

5. RETRIEVAL TASK

The COVID-19 detection is considered a critical system, and because diagnosing the COVID-19 using Artificial Intelligence and Image Analysis is not 100% accurate, it is necessary to use a supporting mechanism to detect the wrong diagnosis (wrong classification). This is the reason of needed the second task that aims to retrieve the most similar images from the database based on measuring the similarity. For example, if we retrieve 10 similar images from the database, then the majority of the retrieved images will be from the class of the query image, and this will guide us to verify the work of the classifier. Moreover, we assume that each image stored in the database has a medical report, and this facilitates diagnosing of the studied case.

The image retrieval task consists of off-line and on-line phases. The off-line phase [43] defined as extracting the features from all images in the database to build a features database. Whereas in the on-line phase, we extract the features from the query image and measure the similarity distance between the query feature and features database. In our work, we used Euclidean distance as a similarity measure [44], which defines as the Equation (12):

\[
D_{\text{Euclid}}(Q, DB) = \sqrt{\sum_{i=1}^{L} (f_{DB_i} - f_Q)^2} \tag{12}
\]
FIGURE 8. (a) Comparison of our proposed methods with other methods in terms of sensitivity, specificity, precision, F-measure and accuracy, for the X-ray database (first level). (b) Comparison of our proposed methods with other methods in terms of sensitivity, specificity, precision, F-measure and accuracy, for the X-ray database (second level). (c) Comparison of our proposed methods with other methods in terms of sensitivity, specificity, precision, F-measure and accuracy, for the CT scan database.
where \( f_{DB_i} \) is the \( i^{th} \) feature of the \( j^{th} \) image in the database, and \( f_Q \) is the feature vector of image query.

## 5.1. Performance measures of image retrieval task

The performance measures used to evaluate the image retrieval task [44] are Precision (number of similar images when we retrieved top 10 images as output), Recall (number of similar images when retrieved top 100 images) and mean Average Precision (mAP) that computed as the Equations (13–15):

\[
\text{Precision} = \frac{\text{No. of retrieved relevant images}}{\text{No. of relevant images}} \quad (13)
\]

\[
\text{Recall} = \frac{\text{No. of retrieved relevant images}}{\text{No. of retrieved images}} \quad (14)
\]

\[
\text{mAP} = \frac{1}{N} \sum_{i=1}^{N} \text{Average precision}_i \quad (15)
\]

We got the mAP measure by calculating the top 10, top 20, ... top 100, then take the average and repeats this procedure for all images in the database.

## 5.2. Databases used in the image retrieval task

We used the following databases to evaluate the performance of the retrieval task:

- **X-ray database**, which consists of three classes as the following:
  - COVID-19 class: It contains 225 images taken from the second classification database.
  - NORMAL class: It contains 500 images taken from the third classification database.
  - PNEUMONIA class: It contains 500 images taken also from the third classification database.

- **CT scan database**, which consists of two classes as the following:
  - COVID-19 class: It contains 350 images taken from the fourth classification database.
  - NOTCOVID-19 class: It contains 350 images taken also from the fourth classification database.

## 5.3. Effect of vocabulary sizes in the image retrieval task

We have evaluated the performance of the image retrieval task in terms of mAP on different vocabulary sizes (i.e. 50, 200, 400, 600, 800, 1000 and 1200) for both databases, using the following techniques:

- Visual words of ResNet-50 and HOG descriptors (Proposed1) based on the BoVW methodology.
- Features fusion of ResNet-50 and HOG descriptors (Proposed2) based on the BoVW methodology.
- Standalone ResNet-50 descriptor based on the BoVW methodology.
- Standalone HOG descriptor based on the BoVW methodology.

According to the experimental results shown in Figure 9a and b, the experimental analysis as the following:

- The proposed technique based on visual words fusion (Proposed1) performs better than the features fusion technique (Proposed2) on the vocabulary of sizes (50, 200, 400, 600 and 800), whereas it decreased on 1000 and 1200 due to the large sizes of the vocabulary that caused overfitting.
- Both proposed methods perform better than standalone ResNet-50 and standalone HOG techniques based on the BoVW in term of mAP on different sizes of the vocabulary.
- The best accuracy of the Proposed1 method in term of mAP (86.02% for X-ray database and 88.23% for CT scan database) was on 600-vocabulary size.
- The best accuracy of the Proposed2 method in term of mAP (85.35% for X-ray database and 86.144% for CT scan database) was on 1000-vocabulary size.

## 5.4. Results of the image retrieval task

We conclude from the previous experiment that the best results in terms of mAP of Proposed1 (Proposed2) method obtained on 600-vocabulary size (1000-vocabulary size) for both databases. In this experiment, we will measure the performance of our proposed methods on specific vocabulary sizes (600 for Proposed1 and 1000 for Proposed2) in terms of precision, recall and mAP.

For both databases, we calculated precision (top 10 images), recall (top 100) and mAP, as well as we compared our proposed methods with the following:

- Deep learning based methods referred in [2,15,22].
- Standalone ResNet-50 feature techniques based on BoVW methodology with 600-vocabulary size.
- Standalone HOG feature techniques based on BoVW methodology with 600-vocabulary size.

The comparisons of our proposed methods with the aforementioned methods in terms of precision, recall and mAP regarding both databases are illustrated in Table 3. Moreover, Figure 10 studies the relationships between precision and recall by calculating the Precision–Recall (P–R) curves for both databases. The first point of each curve refers to the precision.
FIGURE 9. Comparison of our proposed methods with ResNet-50 + BoVW and HOG + BoVW in terms of mAP using different vocabulary sizes, for (a) X-ray, (b) CT scan databases.

TABLE 3. Comparison of our proposed methods with other methods in terms of Precision, Recall and mAP using Euclidean similarity measures for (a) X-ray and (b) CT scan databases.

|          | Performance measure for X-Ray | Performance measure for CT |          |
|----------|--------------------------------|-----------------------------|----------|
|          | Methods                        | Performance measures %      | Methods |
|          |                                | Precision | Recall | mAP  |                                | Precision | Recall | mAP  |
| Xu et al. [22] | 82.90 | 75.10 | 78.42  | Xu et al. [22] | 83.80 | 77.20 | 80.15  |
| Yang et al. [2] | 83.65 | 76.90 | 79.69  | Yang et al. [2] | 84.00 | 78.00 | 80.97  |
| Wang et al. [15] | 84.40 | 77.80 | 81.33  | Wang et al. [15] | 86.60 | 79.00 | 82.94  |
| HOG + BoVW | 79.45 | 71.60 | 75.17  | HOG + BoVW | 78.60 | 72.60 | 75.69  |
| ResNet50 + BoVW | 85.20 | 78.60 | 81.74  | ResNet50 + BoVW | 87.80 | 79.30 | 84.24  |
| Proposed2 | 89.75 | 81.30 | 85.35  | Proposed2 | 90.44 | 82.96 | 86.14  |
| Proposed1 | 90.74 | 82.40 | 86.02  | Proposed1 | 92.50 | 84.70 | 88.23  |

(top 10), and the last point refers to the recall (top 100), and between them, we have the PR values for 20, 30 . . . 90 retrieved images.

From Table 3 and Figure 10, we can conclude the following:

- Our proposed method relies on visual words fusion of ResNet-50, and HOG descriptors based on BoVW methodology performs better than the features fusion of ResNet-50 and HOG descriptors based on BoVW method in terms of precision, recall and mAP.
- Both proposed methods perform better than standalone ResNet-50 and standalone HOG based on the BoVW methodology.
- In addition, both proposed methods outperform the recent works that depend only on deep neural networks.
- The results of our Proposed1 method for X-ray (CT scan) databases are 90.74%, 82.40% and 86.02% (92.50%, 84.70% and 88.23%), respectively, in terms of precision, recall and mAP.
- The results of our Proposed2 method for X-ray (CT scan) databases are 89.75%, 81.30% and 85.35% (90.44%, 82.96% and 86.14%), respectively, in terms of precision, recall and mAP.

6. CONCLUSION AND FUTURE WORK

In our study, we proposed a diagnosis system that detects the COVID-19 disease using chest X-ray or CT scan images. This system performs two major tasks that are classifying the image and retrieving similar images. We proposed two novel
methods for the feature extraction, the first one (Proposed1) relies on Visual Words Fusion of ResNet-50 and HOG descriptors based on BoVW methodology. The second one (Proposed2) relies on Features Fusion of ResNet-50 and HOG descriptors based on BoVW methodology. We used those features to train the Adaptive Boosting classifier and retrieve the most similar images of the scanned images. The results obtained in our work are very promising, and the system can use in the hospital as a complementary method to the RT-PCR test.

In the future, we can develop the system to be able to determine the period of the injury, i.e. the system will diagnose the pathological condition, and when it is COVID-19, the system will determine whether it is in the first week of the injury or the second week and so on. Finally, making the system able to locate the severity degree of COVID-19 disease and suggesting a therapeutic protocol appropriate to the degree of specific risk.

DATA AVAILABILITY

The data underlying this article are available in [Mendeley Data], at http://dx.doi.org/10.17632/h4jh5khgzy.1.

REFERENCES

[1] Li, Q. et al. (2020) Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. N. Engl. J. Med., 382, 1199–1207.
[2] Yang, X., He, X., Zhao, J., Zhang, Y., Zhang, S. and Xie, P. (2020) COVID-CT-dataset: a CT image dataset about COVID-19. arxiv:2003.13865V3. 19 June 2020, preprint: not peer reviewed.
[3] Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang, P. and Ji, W. (2020) Sensitivity of chest CT for COVID-19: comparison to RT-PCR. Radiology, 296, E115–E117.
[4] Ai, T. et al. (2020) Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology, 296, E32–E40.
[5] Mitra, S. et al. (2020) Development and evaluation of a multiplex conventional reverse-transcription polymerase chain reaction assay for detection of common viral pathogens causing acute gastroenteritis. Diagn. Microbiol. Infect. Dis., 97, 115061.
[6] Samsami, M., Bagherpour, J., Nemathicon, B. and Tahmasbi, H. (2020) COVID-19 pneumonia in asymptomatic trauma patients; report of 8 cases. Arch. Acad. Emerg. Med., 8, e46.
[7] Wu, Z. and McGoogan, J. (2020) Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China. JAMA, 323, 1239.
[8] European Society of Cardiology (2020) Diagnostic Imaging in COVID-19 Chest X-ray and CT. [Online] https://www.escardio.org/static-file/Escardio/Education-General/Topic%20pages/Covid-19/Infographic-Diagnostic%20Imaging%20in%20COVID-19.pdf (accessed February 20, 2021).
[9] Wang, Y. et al. (2020) Dynamic evolution of COVID-19 on chest computed tomography: experience from Jiangsu Province of China. Eur. Radiol., 30, 6194–6203.
[10] Xu, X. et al. (2020) Imaging and clinical features of patients with 2019 novel coronavirus SARS-CoV-2. Eur. J. Nucl. Med. Mol. Imaging, 47, 1275–1280.
[11] Lin, C., Ding, Y., Xie, B., Sun, Z., Li, X., Chen, Z. and Niu, M. (2020) Asymptomatic novel coronavirus pneumonia patient outside Wuhan: the value of CT images in the course of the disease. Clin. Imaging, 63, 7–9.
[12] Yang, W., Sirajuddin, A., Zhang, X., Liu, G., Teng, Z., Zhao, S. and Lu, M. (2020) The role of imaging in 2019 novel coronavirus pneumonia (COVID-19). Eur. Radiol., 30, 4874–4882.
[13] Cozzi, D. et al. (2020) Chest X-ray in new coronavirus disease 2019 (COVID-19) infection: findings and correlation with clinical outcome. Radiol. Med., 125, 730–737.
[14] BD, S., Osei-Poku, K. and Brakohiapa, E. (2020) Diagnosing COVID-19 from chest X-ray in resource limited environment-case report. Inmedpub J., 6, 135.
[15] Wang, S. et al. (2020) A deep learning algorithm using CT images to screen for coronavirus disease (COVID-19). MedRxiv 2020.02.14.20023028. 17 February 2020, doi:10.1101/2020.02.14.20023028, arxiv medrxiv:2020.02.14.20023028v5, preprint: not peer reviewed.
[16] Yildirim, O., Talo, M., Ay, B., Baloglu, U., Aydin, G. and Acharya, U. (2019) Automated detection of diabetic subject

FIGURE 10. P–R curves of our proposed methods and other methods using Euclidean distance for (a) X-ray, (b) CT scan databases.
using pre-trained 2D-CNN models with frequency spectrum images extracted from heart rate signals. *Comput. Biol. Med.*, 113, 103387.

[17] Saba, T., Sameh Mohamed, A., El-Affendi, M., Amin, J. and Sharif, M. (2020) Brain tumor detection using fusion of hand crafted and deep learning features. *Cogn. Syst. Res.*, 59, 221–230.

[18] Dorj, U., Lee, K., Choi, J. and Lee, M. (2018) The skin cancer classification using deep convolutional neural network. *Multimed. Tools Appl.*, 77, 9909–9924.

[19] Hosseinzadeh Kassani, S. and Hosseinzadeh Kassani, P. (2019) A comparative study of deep learning architectures on melanoma detection. *Tissue Cell*, 58, 76–83.

[20] Ribili, D., Horváth, A., Unger, Z., Pollner, P. and Csabai, I. (2018) Detecting and classifying lesions in mammograms with deep learning. *Sci. Rep.*, 8, 1–7.

[21] Celik, Y., Talo, M., Yildirim, O., Karabatak, M. and Acharya, U. (2020) Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images. *Pattern Recognit. Lett.*, 133, 232–239.

[22] Xu, X. et al. (2020) A deep learning system to screen novel coronavirus disease 2019 pneumonia. *Engineering*, 6, 1122–1129.

[23] Song, Y. et al. (2020) Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images. In *MedRxiv*. 2020.02.23.20026930.

[24] Zheng, C. et al. (2020) A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT. *IEEE Trans. Med. Imaging*, 39, 2615–2625.

[25] Li, L. et al. (2020) Using Artificial Intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. *Radiology*, 296, E65–E71.

[26] Narin, A., Kaya, C. and Pamuk, Z. (2020) Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. arXiv: 2003.10849. October 06, 2020, preprint: not peer reviewed.

[27] He, K., Zhang, X., Ren, S., and Sun, J. (2016) Deep residual learning for image recognition. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, 27–30 June, pp. 770–778. IEEE, USA.

[28] Dalal, N. and Triggs, B. (2005) Histograms of oriented gradients for human detection. *Proceedings of IEEE Computer Society Conference On Computer Vision And Pattern Recognition (CVPR’05)*, San Diego, CA, USA, 20–25 June, IEEE, USA.

[29] Arthur, D. and Vassilvitskii, S. (2007) k-means++: the advantages of careful seeding. In *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, 7–9 January, pp. 1027–1035. Society for Industrial and Applied Mathematics, 3600 University City Science Center Philadelphia, PA, Philadelphia, PA.

[30] Gonzalez, R. and Woods, R. (2008) *Digital image processing*. Parson, New Jersey, US.

[31] ImageNet (2020) [Online] http://www.image-net.org/ (accessed February 20, 2021).

[32] Orhan, E. and Pitkow, X. (2018) Skip connections eliminate singularities. *Proceedings of 6th International Conference on Learning Representations*, Vancouver Convention Center, Vancouver, BC, Canada, 30 April–3 May, pp. 1–22.

[33] Simard, P., Steinkraus, D., and Platt, J. (2003) Best practices for convolutional neural networks applied to visual document analysis. *Proceedings of Seventh International Conference on Document Analysis and Recognition*, Edinburgh, UK, 6–6 August, IEEE, Edinburgh, UK.

[34] Cireşan, D., Meier, U., Gambardella, L. and Schmidhuber, J. (2010) Deep, big, simple neural nets for handwritten digit recognition. *Neural Comput.*, 22, 3207–3220.

[35] Chawla, N., Bowyer, K., Hall, L. and Kegelmeyer, W. (2002) SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.*, 16, 321–357.

[36] Wang, R. (2012) AdaBoost for feature selection, classification and its relationship with SVM, a review. *Phys. Procedia*, 25, 800–807.

[37] UCSD-AI4H/COVID-CT (2020) [GitHub] [Online] https://github.com/UCSD-AI4H/COVID-CT (accessed February 20, 2021).

[38] COVID-19 Radiography Database (2020) [Kaggle.com] [Online] https://www.kaggle.com/tawsifurrahman/covid19-radiography-database (accessed February 20, 2021).

[39] Chest X-Ray Images (Pneumonia) (2020) [Kaggle.com] [Online] https://www.kaggle.com/paulimothymooney/chest-xray-pneumonia (accessed February 20, 2021).

[40] Powers, D.M.W. (2011) Evaluation: from precision, recall and F-measure to ROC, informedness, markedness & correlation. *J. Mach. Learn. Technol.*, 2, 37–63.

[41] Altman, D. and Bland, J. (1994) Statistics notes: diagnostic tests 1: sensitivity and specificity. *BMJ*, 308, 1552–1552.

[42] Fawcett, T. (2006) An introduction to ROC analysis. *Pattern Recognit. Lett.*, 27, 861–874.

[43] Alrahhal, M. and Supreethi, K. (2019) Content-based image retrieval system combining local patterns and supervised machine learning techniques. *Proceedings of Amity International Conference on Artificial Intelligence (AICAI)*, Dubai, United Arab Emirates, 4–6 February, pp. 118–124. IEEE, Amity University, Dubai, United Arab Emirates.

[44] Alrahhal, M. and Supreethi, K.P. (2020) Multimedia image retrieval system by combining CNN with handcraft features in three different similarity measures. *Int. J. Comput. Vis. Image Process.*, 10, 1–23.