Journal of Theoretical and Applied Information Technology
15th June 2021. Vol.99. No 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# DETECTION OF COVID-19 FROM CHEST CT IMAGES USING SELECTED FOURIER TRANSFORM FEATURES

**[1]FARID ALI MOUSA, [2] TAHA M. MOHAMED**

*[1]Information Technology Department, Faculty of Computers and Artificial Intelligence, Beni-Suef University, Egypt.*
*[2] Information Technology Department, Faculty of Computers and Artificial Intelligence, Helwan University, Egypt.*
*[2]MIS Dept., College of Business, University of Jeddah, Kingdom of Saudi Arabia (KSA)*

*Email:* [1] fared.ali@fcis.bsu.edu.eg, [2] tahamahdy3000@yahoo.com

**ABSTRACT**

In this paper, a novel method is proposed for COVID-19 detection from chest images. The proposed method uses some important features from both spatial and the Fourier transform of the input images. The binary particle swarm optimization is used to select the most relevant features. Two common classifiers are used for testing; support vector machine and k-nearest neighbor. Results show that the k-nearest neighbor outperforms support vector machine. The accuracy of the proposed method outperforms other algorithms in the literature. The accuracy of the proposed method approximately equals 91% when using the proposed features combined with the binary particle swarm optimization (BPSO). The sensitivity exceeds 89%, and also outperforms that proposed in previous work. Specificity is also maintained. These important findings may represent physicians' importance in decreasing diagnosis time and cost using automated systems. These systems may be useful for physicians in case of resource limitation.

**Keywords:** COVID-19, non-COVID, Classification, FT, Feature Selection, and BPSO.

## 1. INTRODUCTION

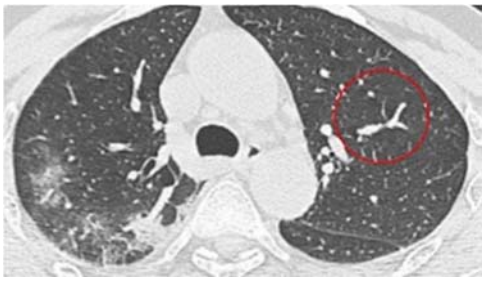
SARS CoV-2 infection has been reported first in December 2019 at Wuhan, China. The virus causes COVID-19 disease. It quickly spreads all over the world. COVID-19 has been declared by the World Health Organization as a public pandemic [1, 2]. Coronavirus causes respiratory diseases. The main signs of COVID-19 are cough, fever, muscle soreness, lack of respiration, and fatigue [3], [4]. About 20% of the patients progress to more severe diseases such as pneumonia and respiratory failure.

In some cases, the disease causes death [4]. No therapy for COVID-19 has been developed or accepted till now [5]. However, some vaccines have recently developed. The efficacy of these vaccines has not been confirmed yet. So, social distancing, mask usage, and lockdown has been the most important method for preventing the transmission of the disease to reduce clinic overload. [6]. Therefore, early detection for COVID-19 is essential [7].

Fast and accurate testing tools are required for pandemic control [6] . The available diagnostic tools are depending on the detection of the viral gene, antibody detection, and viral antigen detection. Detection of the viral gene is performed by using RT-PCR. It is the most reliable detection technique [5], [6] . Rapid antibody tests are used for extensive immune screening. However, the antibody tests don't show the virus's existence. The antigen tests till now are in progress [6]. Despite the general acceptance of RT-PCR, it suffers from false negatives [5]. Additional challenges related to RT-PCR requires processing delays, limiting resources, inconsistency in test procedures and poor sensitivity (60–70%) [8]. These RT-PCR limitations make CT chest testing more attractive for specific subsets of patients or in a resource-constrained environment. It is important to note that, screening large numbers of suspected cases is a priority at the time of the pandemic.

X-ray and Computed tomography (CT) take a major role in detection and treatment of COVID-19 [9]. CT scanning of pneumonia is helpful for equivocal chest X-Ray cases. CT is also useful for RT-PCR initial negative cases, and also for disease detection [5]. COVID-19 pneumonia has some essential characteristics. It includes opacities in the lower lobes. Also, air bubble signs appear on the patients [10] ,[11]. Unfortunately, CT characteristics differ according to the disease's stage and disease's

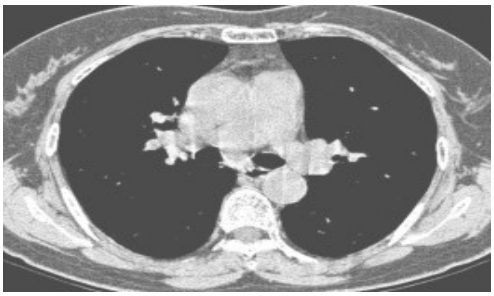
severity. Figure 1 shows examples of chest CT positive COVID. Figure 2 shows examples of chest CT negative COVID [12].



*(a) COVID image example*


*(b) COVID image example*

*Figure 1: Examples of COVID images adapted from[12]*



*(c) Non COVID image example*

*(d) Non COVID image example*

*Figure 2: Examples of non-COVID images adapted from [12]*

CT is significantly useful in classifying and evaluating COVID-19 [7]. That is, most of the hospitals have CT imaging equipment. Physicians can manually perform the detection task. However, manual analysis of chest CT images includes a considerable time and radiology expert. When COVID-19 infection occurs, time becomes very valuable. The main drawback of using manual chest CT classification is the overlap with other influenza pneumonia [8]. Besides, chest CT examinations could lead to false-positive results and false-negative results [13]. Therefore, automatic processing of chest CT images is ideal to save the time of specialists [7]. Besides, combining CT and RT-PCR is likely to produce more accurate classification results [8]. AI technologies can help medical specialists automate chest CT image analysis [2] , [9]. Deep learning methods could be used in COVID-19 detection from chest CT images. At the pandemic, artificial intelligence techniques will reduce workloads on doctors in epidemic regions. It could improve the accuracy of the diagnosis. The automatic detection reduce the physicians consumed detection time by about 30%-40% [14]. Moreover, it can help radiologists make clinical decisions, diagnosis, tracking, and prognosis [9].

In this paper we present a novel method to classify chest CT images into two classes either COVID or non-COVID. Fifteen features are computed from both the frequency and the spatial domain of the input CT image. The binary particle swarm optimization (BPSO) algorithm is applied to select the most important features. The proposed method shows the reliability of using chest CT images in the COVID pneumonia detection. This paper is structured as follows. Section 2 outlines the previous work. Overview of the classifiers used in this paper was discussed in Section 3. The suggested approach is outlined in Section 4. The findings and their discussion are discussed in Section 5. In section 6, the conclusion is finally addressed.

## 2. PREVIOUS WORK

In [9], the authors review the techniques of the medical imaging related to COVID-19. These techniques include diagnosis and image segmentation. In [11], the authors state that CT findings differ according to disease severity, and also according to disease stage. So, this information is important for successful classification and also for dataset selection. In [15], the authors outlined different diagnostic methods for COVID-19. In [16], the authors review the AI diagnostic methods by applying chest CT images of COVID-19. Their results show slight variations of the UK cases compared to European cases and Asian cases regarding chest CT images [5].

In [8], the authors use deep learning algorithms for COVID-19 pneumonia classification. They achieved 90.8% accuracy, 84% sensitivity, and 93% specificity. The false-positive rate was 10%. The authors conclude that AI-algorithms can identify COVID-19 associated pneumonia from input CT images. A deep learning model in [17] was developed for COVID-19 detection. The algorithm extracts some visual features from volumetric CT. Their findings show that profound learning models can reliably diagnose COVID-19 pneumonia. COVID-19 can also vary from other pulmonary diseases. In [14], the authors propose a system consists of classification and segmentation. They claim that, their system saves the detection time for the physicians by 30%-40%. Actually, this result is very important during the COVID pandemic age due to the shortage of the available medical resources.

In [18], the authors segment lung regions and vessels by deep learning algorithms. The results encourage using AI for COVID-19 region detection. In [12], the authors use convolutional neural networks to differentiate the suspected cases from the infected cases. Their model accurately classified the chest CT images.

In [2], the authors investigated the characteristics of chest CT imaging for infected preschool children. They conclude that the findings of the COVID-19 chest images are uncommon and different in these children [2]. Again, this conclusion affects the performance of the classifiers.

The authors in [3] discussed the COVID features of the non-enhanced CT chest imaging in Jordan. They conclude that the principle diagnostic tool of the COVID-19 is the CT. They noted some pulmonary change patterns such as Halo sign, ground glass, consolidation, and some other changes.

In [12], [17], the authors built an open source chest CT images of COVID-19. The dataset consists of 349 CT images of COVID-19. In which 216 patients were infected, and 463 were not. Based on self-supervised learning and multi-task learning, the authors developed their diagnosis methods with accuracy reaches 0.89.

Unfortunately, deep learning based methods suffer from some limitations as stated by [19] . One of these limitations is the requirement of large dataset for training to achieve perfect results. Obviously, it is difficult to avail a large COVID-19 dataset, especially in the pandemic age. Most of the previous overviewed methods use deep learning techniques. However, the methods still show a low accuracy, sensitivity and specificity due to poor feature extraction selection and/or poor classifiers. To enhance the classification performance, this paper proposes a novel method for the detection of COVID-19 from chest CT image. The proposed method uses the dataset presented in [12], [17]. The proposed method combines some important features from both the spatial and frequency domains of the images. The paper uses two popular classifiers: SVM and KNN. The obtained results show the superiority of the proposed model over other previous work.

## 3. OVERVIEW OF CLASSIFIERS

Support vector machine (SVM) is one of the most common and strongest classifiers. SVM uses a separating hyperplane *H* classification scheme. The maximum distance between two parallel hyperplanes is the margin of the SVM. These two margins have not any instances between them. The primary task of the SVM algorithm is to classify instances based on a linear feature function. However, non-linear classification can also be performed with a kernel function. The SVM classifier is supplied with pre-labelled instances. The SVM searches a hyperplane that maximises the distance by choosing points as support vectors as shown in figure 3. In our problem, the two instance classes are positive and negative COVID. The more the margin, the better the generalization of the misclassification [28]. This paper used three popular SVM kernel functions which are; polynomial kernel function, linear kernel function, and Radial Basis Function (RBF). The RBF kernel is also called Gaussian kernel function. More details on these kernel functions could be found in [29].

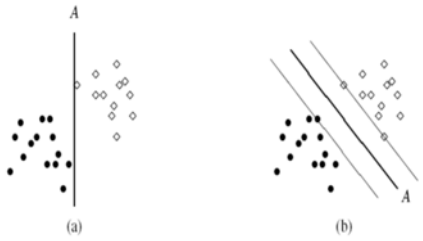



*Figure 3: SVM classifier*

The K nearest neighbour (knn) is a kind of supervised learning. This classifier attempts to categorize the set of input patterns (images in our case) based on data properties. Knn classifies a test image to the class having the maximum nearest K patterns. The majority voting technique is used in this case [30], [31]. Knn maintains training patterns and uses various distance functions to perform the

classification. Here, some of the distance metrics are detailed [30].

### 3.1 Distance Metrics

**a) City block distance**

City block distance calculates the distance from one point to the other by following a grid-like direction. The distance of block between two items is the Manhattan distance. City block distance is a vector-based approach so that in n-dimensional vector space 'q' and 'r' are defined. The distance L1 or block is determined from the summation of borders. This distance is illustrated by Equation 1.

$$\boldsymbol{L_1(q,r) = \sum_y |q(y) - r(y)|} \quad (1)$$

**b) Cosine similarity**

Cosine similarity is a measure between two vectors of an inner product. The attribute as a vector used by Cosine's similarity metric will find the normalization point product of the two vectors. Mathematically, this measure is shown by Equation 2 as:

$$\boldsymbol{\cos(\theta) = \frac{X.Y}{|X| * |Y|}} \quad (2)$$

Where (.) represent the dot product for x and y vectors.

**c) Euclidean distance**

The Euclidean distance is the easiest way to calculate similarities, since the axes on a graph are the standard attributes for data objects. This distance is mathematically expressed by Equation 3 for two vectors X and Y with n attributes:

$$\boldsymbol{E(X,Y) = \sqrt{\sum_{i=0}^{n}(X_i - Y_i)^2}}$$ Error! No text of specified style in document. (3)

**d) Jaccard Coefficient**

Jaccard coefficient used for measuring the scale and the proportions of the overlap between two samples A and B. It is defined for two sets as the cardinality of its crossroads separated by the cardinality of its union. This is expressed mathematically in Equation 4 as:

$$\boldsymbol{J(A,B) = \frac{|A \cap B|}{|A \cup B|}} \quad (4)$$

**e) Minkowski Distance**

The distance from Minkowski is a generality of the distance of Euclidean and City block, Equation 5 explains this distance.

$$\boldsymbol{dist = (\sum_{k=1}^{n} |p_k - q_k|^r)^{\frac{1}{r}}} \quad (5)$$

Where, *n* represents the dimensions numbers, *r* represents the parameters, $_k$ is the k$^{th}$ component *and* *p* and *q* represent the data objects.

**f) Mahalanobis Distance**

It's a multi-dimensional generalization of the concept of determining how many standard deviations P differs from q's mean. When P is at the mean of q, this distance is zero, and it increases as P moves away from the mean for each principal component axis. The formula shown in Equation 6:

$$\boldsymbol{dist(p,q) = (p-q)\sum^{-1}(p-q)^T} \quad (6)$$

Such that, $\sum^{-1}$ represents the matrix of the covariance among the variables.

### 3.2 Classifiers Evaluation Criteria

Different performance metrics could evaluate classifiers. The first metric is the Positive Predictive Value (PPV) [32], also called precision, which given by Equation (7) as:

$$PPV = TP / (TP + FP) \quad (7)$$

Where *TP* represents the true positive and *FP* represents the false positive. PPV tells how likely a positive test indicates that the person has the disease [32]. The Negative Predictive Value (NPV) is given by Equation 8 as:

$$NPV = TN / (TN + FN) \quad (8)$$

Where *TN* represents the true negative, and *FN* represents the false negative. NPV tells how likely a negative test indicates the disease is not present in the tested person. The accuracy of the classifier is one of the most important performance metrics. This metric measures the overall classifier performance by detecting all True cases from the test sample [33]. The accuracy is given by Equation 9 as:

$$\text{Accuracy} = \frac{(TP+TN)}{(P+N)} \quad (9)$$

Where *P* represents positive cases total number and *N* represents the total number of the negative cases. Both sensitivity and specificity are essential metrics of classifiers performance [32]. The test's sensitivity is the percentage of the persons who have a positive test result relative to the disease's overall persons. Sensitivity is vital as it considers the false negative (*FN*). At the age of the pandemic, *FN* is critical. Persons classified as FN may suffer from advanced infection and may transfer the infection to other persons. So, increasing sensitivity is essential. Sensitivity is given by Equation 10.

$$Sensitivity = \frac{TP}{TP+FN} \quad (10)$$

Equation 11 shows the specificity. For COVID, specificity is also important as it considers the number of false-positive (*FP*) cases. False-positive cases are the cases being classified as infected while actually, they are not. *FP* should be kept minimal due to the limitations of resources and medical care at the COVID pandemic [34].

$$Specificity = \frac{TN}{TN+FP} \quad (11)$$

## 4. THE PROPOSED METHOD

This paper presents a novel method to classify the COVID and non-COVID chest CT images. The proposed method uses some computed features of both spatial domains and frequency of the input images. The binary particle swarm optimization algorithm is used for feature selection. Additionally, two well-known classifiers are used. These classifiers are SVM and KNN. In Figure 4, the general block diagram of the proposed model is shown. The algorithm begins with a pre-processing of the input image. Then, it computes the FT of the pre-processed image. Therefore, fifteen features are computed. These features are fed to the two classifiers. Also, BPSO is used to select the most relevant features from these fifteen features. In the two cases, the performance of the classifiers is analysed. In the following subsections, we will detail the proposed algorithm.

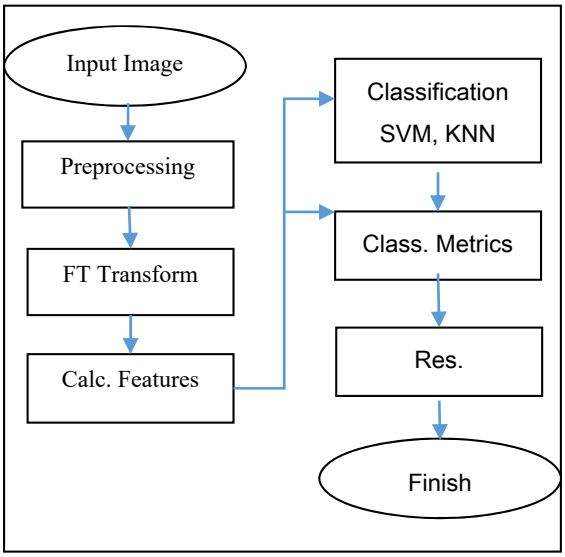



*Figure 4: The general block diagram of the proposed model*

### 4.1 Preprocessing

The inputs to the proposed algorithm are chest CT images. We use the dataset images collected in [12], [17]. These images have different dimensions. So, pre-processing is needed to apply the proposed algorithm. The first pre-processing is done to resize (zoom) images to be of the same size [20]. This process makes further processing simplified. Using a bilinear interpolation algorithm, zooming is done. The bilinear algorithm is used to map a pixel spot on the screen to a different point. To compute the weighted average, the four neighbouring or surrounding pixels are used. This average is applied to the pixel. This process takes into account the 2 by 2 identified pixels, and then, calculates their weighted average [21].

### 4.2 Feature Selection and Extraction

Feature extraction goal is to extract the needed features for classification from the input CT images. We use some frequency and spatial features. Frequency domain features could be obtained from the Fourier transform (FT) of the input image. FT analyses the input image to its sine and cosine components. FT could be computed using the Fast Fourier Transform algorithm (FFT). The result of FT analysis is a matrix of frequency coefficients [22]. Occasionally, FT was used for feature in many image and signal analysis domains such as in [23].

In this paper, we use fifteen computed features of the input image. These features are; Variance, Mean, Mean of Energy, Minimum Amplitude, Maximum Amplitude, Minimum Energy, Maximum Energy, Mid Frequency, Minimum frequency, Average Frequency, Maximum frequency, Half Point of the energy (HaPo), Entropy, Power Spectral Density (PSD) and Root Mean Square (RMS) [24, 25]. The statistical features of an image are commonly used in feature extraction. For example, both Entropy and RMS, are used for feature extraction in some applications [26].

Table 1 shows each of these features and types. It is clear from Table 1 that thirteen features are from the frequency domain, while the remaining two features are from the spatial domain. These features are detailed next.

*Table 1: Description of Spatial and FT features used*

| Feature# | Feature Name | Type |
|---|---|---|
| 1 | Mean | Frequency |
| 2 | Variance | Frequency |
| 3 | Mean of Energy | Frequency |
| 4 | Minimum Amplitude | Frequency |
| 5 | Maximum Amplitude | Frequency |
| 6 | Minimum Energy | Frequency |
| 7 | Maximum Energy | Frequency |
| 8 | Average Frequency | Frequency |
| 9 | Mid Frequency | Frequency |
| 10 | Maximum frequency | Frequency |
| 11 | Minimum frequency | Frequency |
| 12 | Half Point of the energy (HaPo) | Frequency |
| 13 | Power Spectral Density | Frequency |
| 14 | Entropy | Spatial |

| | | |
|---|---|---|
| 15 | Root Mean Square (RMS) | Spatial |

The Mean value, $m$, of the FT coefficients, is given by Equation 12 as:

$$m = \frac{1}{n}\sum_{i=1}^{n} X_i \quad (12)$$

Where, $n$ represents the frequency coefficients total number, $X_i$ is the input image's original intensity values.

The Variance of the FT is given by Equation (13) as:

$$Var = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - m)^2 \quad (13)$$

The third feature we used is the Mean of Energy. It is also considered as a type of average value of the frequency components. It is given by Equation 14 as:

$$Mean_{En} = \frac{1}{n}\sum_{i=1}^{n} {X_i}^2 \quad (14).$$

Where $X_i$ represents the frequency coefficients computed from the FT, we use the Maximum and the Minimum Amplitudes features of the FT, respectively. The Maximum and the Minimum Energies are the maximum and minimum energy values from the input image's frequency components, respectively. The Average Frequency is given by Equation 15 as:

$$Avg_{Fre} = \frac{\sum_{i=1}^{n} f_i x p_i}{\sum_{i=1}^{n} p_i} \quad (15)$$

Where $f$ is the frequency vector and $p$ is the power spectral density. The Mid Frequency is the frequency value derived from the highest power spectral density. The maximum and minimum frequencies are the spectrum's highest and lowest frequencies values of energy, respectively. The Half Point of the energy (HAPO) is the frequency that separates the spectrum into two parts. The Power Spectral Density (PSD) characterizes random signals. It is a measure of the signal's power versus frequency. PSD is given by Equation 16 as:

$$PSD = \frac{1}{2\pi n} \times |D_{FFT}|^2 \quad (16)$$

We also used two additional spatial domain features; Entropy and RMS. Entropy can calculate the complexity of items in information theory. It is a statistical measure of randomness used to model the texture of the input image. Thus, entropy could be used to extract the signal's internal features. The entropy is given by Equation 17 as:

$$E = -\sum_{i=1}^{n} P_i log_2 P_i \quad (17)$$

Where $P$ indicates the current probability associated with each grey level. The Root Mean Square (RMS) expresses the image intensities. It is given by Equation 18 as:

$$RMS = Sqrt(\sum_{i=1}^{n} \frac{1}{n}(x_i)^2) \quad (18)$$

Where $x_i$ represents the different image intensity levels.

In this paper, we also uses the binary particle swarm optimization (BPSO) algorithm. BPSO can select the most relevant features from the previous fifteen FT features. Reducing the number of features aims to simplify processing operations, and enhances classifiers performance. The particle swarm optimization (PSO) improves a candidate solution by optimizing a problem iteratively [27]. It is a non-deterministic heuristic. Particles in PSO is the population of candidate solutions. Space is searched to explore the most relevant features. Some enhanced BPSO is presented in the literature [27]. BPSO is the binary version of the PSO. Here, the problem which BPSO tries to solve is the choice of the most relevant feature. A candidate solution can represent a particular feature. In BPSO, when a particular feature is selected as a candidate, the binary value of BPSO equals 1; otherwise, the value equals 0.

## 5. EXPERIMENTAL RESULTS

This section shows the proposed results. All results are conducted using Matlab 2017. The proposed method uses the dataset presented in [12], [17] for performance comparison. In all experiments, k-fold cross-validation is used with k=10. We test two classifiers; SVM and KNN. For SVM, we test three kernel functions; Gaussian, Linear, and Polynomial. Mainly, we are interested in the accuracy of each classifier. However, both sensitivity and specificity are also critical. In the COVID-19 pandemic time, sensitivity may have more importance than specificity. That is, high sensitivity implies less FN, as previously illustrated. The results of applying SVM and KNN on the fifteen FT features are shown in Table 2 and Table 3. It is clear from the table that the Gaussian kernel achieves the highest accuracy among all SVM kernels; the accuracy of applying SVM Gaussian on FT features exceeds 89%. The second-ranked function is the polynomial function. The linear function has the

least accuracy. This result was expected as the separation between the classes is often not linear.

For sensitivity, the same analogy occurs for the three kernels. The highest sensitivity equals 86% in the case of the Gaussian kernel. However, the specificity of the polynomial function is the highest one among the three kernels. So, the SVM Gaussian kernel has the best performance for both accuracy and sensitivity. The SVM Gaussian slightly outperforms the accuracy described in [1], which is only 0.89. SVM polynomial achieves the maximum PPV in the table. For KNN, it is clear from the table that KNN accuracy exceeds 90%. KNN outperforms all SVM kernels.

Additionally, the sensitivity computed using KNN is higher than the sensitivity of all KNN kernels. However, the specificity of SVM-Polynomial is slightly larger than the specificity of KNN. So, KNN achieves the highest performance compared to all SVM kernels in both sensitivity and accuracy. Interestingly, both SVM Gaussian and KNN outperform the work presented in [17]. Also, the maximum NPV is achieved through using KNN.

*Table 2: SVM performance using computed features*

| Met. | SVM | | |
|---|---|---|---|
| | Gauss. | Linear | Poly. |
| Sens. | 0.8692 | 0.7238 | 0.8438 |
| Spec. | 0.9208 | 0.7734 | 0.9309 |
| PPV | 0.914 | 0.7507 | 0.9284 |
| NPV | 0.8791 | 0.7481 | 0.8489 |
| Accur. | 0.8954 | 0.7493 | 0.8861 |

*Table 3: KNN performance using computed features*

| Met. | KNN | | |
|---|---|---|---|
| | k=8 | k=10 | k=12 |
| Sens. | 0.826 | 0.8819 | 0.8349 |
| Spec. | 0.9165 | 0.9267 | 0.9236 |
| PPV | 0.9082 | 0.9198 | 0.9158 |
| NPV | 0.8404 | 0.8917 | 0.8489 |
| Accur. | 0.8712 | 0.9048 | 0.8794 |

In addition to original features, we perform feature selection using Binary Particle Swarm Optimization (BPSO). The target is to select the most relevant features from the fifteen FT features. Results show that BPSO chooses nine features from the fifteen features previously explained features. These features are; Minimum frequency, Average Frequency, Maximum frequency, Half Point of the energy (HaPo), Entropy, RMS, Power Spectral Density, Variance, and Mean of Energy. BPSO chooses the two spatial domain features due to their importance.

Table 4 and Table 5 shows the results of BPSO. Again, SVM Gaussian accuracy outperforms the two other kernels. The accuracy reaches 89%, which is slightly higher than FT results without feature selection. Moreover, the linear kernel function has the least performance compared to the other two kernels. So, using feature selection enhances the classification accuracy and chooses the most relevant features. In all cases, the sensitivity is increased compared to Table 2. However, the specificity of the SVM polynomial shown in Table 2 achieves the highest specificity. For KNN, the accuracy exceeds 90%. KNN accuracy in Table 3 is slightly higher than that of Table 2. So, using feature selection chooses the relevant features and always increases classifier accuracy. Sensitivity is also enhanced when being compared to all previous results. Specificity is almost equal to SVM-Polynomial without BPSO. The highest PPV is achieved using KNN. Also, the highest NPV is achieved using KNN.

*Table 4: SVM performance using BPSO*

| Met. | SVM- BPSO | | |
|---|---|---|---|
| | Gauss. | Linear | Poly. |
| Sens. | 0.8743 | 0.7315 | 0.8645 |
| Spec. | 0.9237 | 0.7848 | 0.9132 |
| PPV | 0.9169 | 0.7650 | 0.9063 |
| NPV | 0.8841 | 0.7531 | 0.8741 |
| Accur. | 0.8995 | 0.7587 | 0.8892 |

*Table 5: KNN performance using BPSO*

| Met. | KNN - BPSO | | |
|---|---|---|---|
| | k=8 | k=10 | k=12 |
| Sens. | 0.8372 | 0.8846 | 0.8306 |
| Spec. | 0.9259 | 0.9293 | 0.9211 |
| PPV | 0.9184 | 0.9226 | 0.9133 |
| NPV | 0.8511 | 0.8942 | 0.8447 |
| Accur. | 0.8817 | 0.9091 | 0.8759 |

The performance comparison of all tested classifiers is shown in Figures 4, 5, and 6. We can conclude general notes from the figures. Accuracy, sensitivity, and specificity obtained by applying KNN are higher than those of SVM. Also, using

BPSO enhances classification accuracy and sensitivity. SVM polynomial kernel outperforms other classifiers in specificity. Overall, the highest accuracy achieved exceeds 90%. It is higher than that proposed in [17] on the same dataset. This accuracy is achieved by applying BPSO on FT features and being classified with knn. The proposed method achieves larger accuracy compared to [8] which equals 0.908. Furthermore, the proposed method achieves a much larger sensitivity. The highest reported sensitivity in [8] equals 0.853. However, the proposed method achieves sensitivity exceeds 0.88. The specificity of both the proposed method and method described in [8] is almost the same.

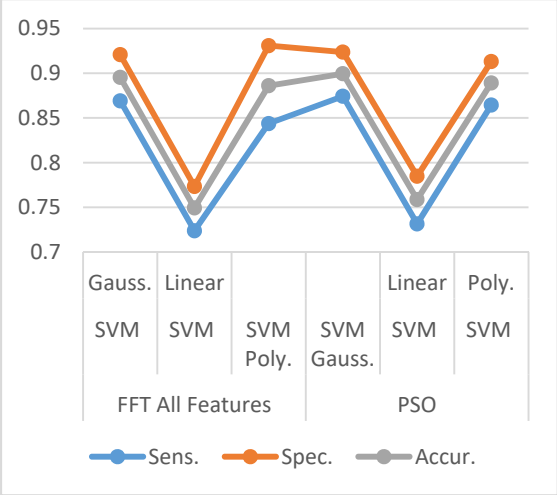



*Figure 4: Comparison of the tested classifiers (SVM)*

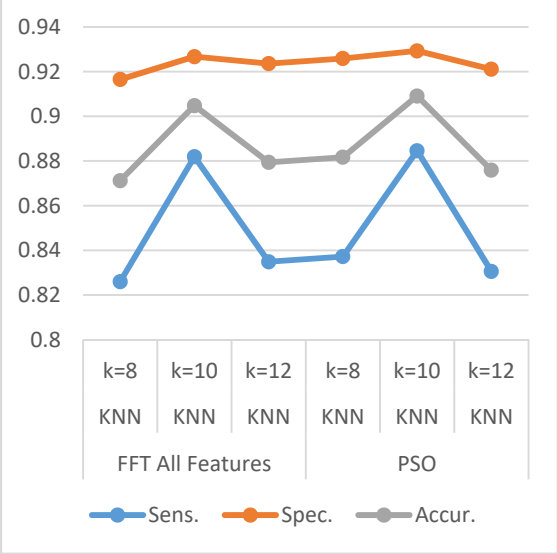
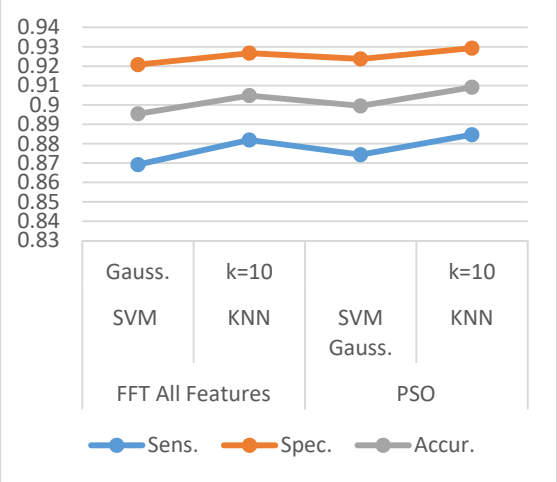



*Figure 5: Comparison of the tested classifiers (KNN)*

*Figure 6: Comparison of the tested classifiers and PSO*




## 6. CONCLUSIONS

We have proposed in this paper an efficient method for COVID-19 detection from chest CT images. The proposed method uses FT for feature extraction. Two spatial domain features are also added. Additionally, we have also used BPSO as a feature selection method. Two classifiers are tested for results comparison; SVM and KNN. Results show that selecting some features using BPSO achieves higher accuracy when compared to previous work. The KNN classifier outperforms SVM. The highest accuracy obtained approximately equals 91%. Both sensitivity and specificity have increased using the proposed method. These results are significant for rapid detection of COVID-19 infection, especially, in case of resource shortage. It represent a reliable tool for physicians for reliable, accurate and fast detection of COVID-19. Future work includes testing other frequency transforms such as the wavelet transform.

## REFERENCES

[1] Li, L., et al., *Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT.* Radiology, 2020.

[2] Li, Y., et al., *Chest CT imaging characteristics of COVID-19 pneumonia in preschool children: a retrospective study.* BMC pediatrics, 2020. 20: p. 1-8.

[3] Albtoush, O.M., R.B. Al-Shdefat, and A. Al-Akaileh, *Chest CT scan features from 302 patients with COVID-19 in Jordan.* European journal of radiology open, 2020. 7: p. 100295.
[4] Sheikhi, K., H. Shirzadfar, and M. Sheikhi, *A review on novel coronavirus (Covid-19): symptoms, transmission and diagnosis tests.* Research in Infectious Diseases and Tropical Medicine, 2020. 2(1): p. 1-8.
[5] Pakdemirli, E., U. Mandalia, and S. Monib, *Characteristics of chest CT images in patients with COVID-19 pneumonia in London, UK.* Cureus, 2020. 12(9).
[6] Yüce, M., E. Filiztekin, and K.G. Özkaya, *COVID-19 diagnosis—A review of current methods.* Biosensors and Bioelectronics, 2020: p. 112752.
[7] Singh, D., V. Kumar, and M. Kaur, *Classification of COVID-19 patients from chest CT images using multi-objective differential evolution–based convolutional neural networks.* European Journal of Clinical Microbiology & Infectious Diseases, 2020. 39(7): p. 1379-1389.
[8] Harmon, S.A., et al., *Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets.* Nature communications, 2020. 11(1): p. 1-7.
[9] Feng, S., *Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19.* arXiv preprint arXiv:2004.02731, 2020.
[10] Hani, C., et al., *COVID-19 pneumonia: a review of typical CT findings and differential diagnosis.* Diagnostic and interventional imaging, 2020. 101(5): p. 263-268.
[11] Hefeda, M.M., *CT chest findings in patients infected with COVID-19: review of literature.* Egyptian Journal of Radiology and Nuclear Medicine, 2020. 51(1): p. 1-15.
[12] C, C.-. *dataset of CT images, available at https://github.com/UCSD-AI4H/COVID-CT.* Computer Communications, 2021, 2020. last visited 7 Feb. 2021.
[13] Kwee, T.C. and R.M. Kwee, *Chest CT in COVID-19: what the radiologist needs to know.* RadioGraphics, 2020. 40(7): p. 1848-1865.
[14] Wang, B., et al., *AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system.* Applied Soft Computing, 2020. 98: p. 106897.
[15] Yuan, X., et al., *Current and perspective diagnostic techniques for COVID-19.* ACS infectious diseases, 2020. 6(8): p. 1998-2016.
[16] Ozsahin, I., et al., *Review on diagnosis of COVID-19 from chest CT images using artificial intelligence.* Computational and Mathematical Methods in Medicine, 2020. 2020.
[17] Zhao, J., et al., *Covid-ct-dataset: a ct scan dataset about covid-19.* arXiv preprint arXiv:2003.13865, 2020.
[18] Pu, J., et al., *Automated quantification of COVID-19 severity and progression using chest CT images.* European Radiology, 2021. 31(1): p. 436-446.
[19] Bahman Zohuri, M.M., *Deep Learning Limitations and Flaws.* Modern Approaches on Material Science, 2020(January 29, 2020): p. 241-250.
[20] Hashemi, M., *Enlarging smaller images before inputting into convolutional neural network: zero-padding vs. interpolation.* Journal of Big Data, 2019. 6(1): p. 1-13.
[21] Cao, L., et al., *Grayscale image colorization using an adaptive weighted average method.* Journal of Imaging Science and Technology, 2017. 61(6): p. 60502-1-60502-10.
[22] Wu, X., et al., *Fourier-based rotation-invariant feature boosting: An efficient framework for geospatial object detection.* IEEE Geoscience and Remote Sensing Letters, 2019. 17(2): p. 302-306.
[23] Hindarto, H., A. Muntasa, and S. Sumarno. *Fourier transform for feature extraction of Electro Encephalo Graph (EEG) signals.* in *Journal of Physics: Conference Series*. 2019. IOP Publishing.
[24] Boonyakitanont, P., et al., *A review of feature extraction and performance evaluation in epileptic seizure detection*

*using EEG*. Biomedical Signal Processing and Control, 2020. 57: p. 101702.

[25] Camilo, J.A., et al. *An improved frequency domain feature with partial least-squares dimensionality reduction for classifying buried threats in forward-looking ground-penetrating radar data*. in *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXII*. 2017. International Society for Optics and Photonics.

[26] Varga, D., *No-reference image quality assessment based on the fusion of statistical and perceptual features.* Journal of Imaging, 2020. 6(8): p. 75.

[27] El-Maleh, A.H., A.T. Sheikh, and S.M. Sait, *Binary particle swarm optimization (BPSO) based state assignment for area minimization of sequential circuits.* Applied soft computing, 2013. 13(12): p. 4832-4840.

[28] Chandra, M.A. and S. Bedi, *Survey on SVM and their application in image classification.* International Journal of Information Technology, 2018: p. 1-11.

[29] Achirul Nanda, M., et al., *A comparison study of kernel functions in the support vector machine and its application for termite detection.* Information, 2018. 9(1): p. 5.

[30] Bania, R.K. and A. Halder, *R-Ensembler: A greedy rough set based ensemble attribute selection algorithm with kNN imputation for classification of medical data.* Computer methods and programs in biomedicine, 2020. 184: p. 105122.

[31] Karanja, E.M., S. Masupe, and M.G. Jeffrey, *Analysis of internet of things malware using image texture features and machine learning techniques.* Internet of Things, 2020. 9: p. 100153.

[32] Tharwat, A., *Classification assessment methods.* Applied Computing and Informatics, 2020.

[33] Chicco, D. and G. Jurman, *The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation.* BMC genomics, 2020. 21(1): p. 1-13.

[34] Goldstein, N.D. and I. Burstyn, *On the importance of early testing even when imperfect in a pandemic such as COVID-19.* Global Epidemiology, 2020. 2: p. 100031.