Covid-19 diagnosis by WE-SAJ

Wei Wang\textsuperscript{a,*}, Xin Zhang\textsuperscript{b,*}, Shui-Hua Wang\textsuperscript{a} and Yu-Dong Zhang\textsuperscript{a}

\textsuperscript{a}School of Computing and Mathematical Sciences, University of Leicester, Leicester, UK; \textsuperscript{b}Department of Medical Imaging, The Fourth People’s Hospital of Huai’an, Huai’an, People’s Republic of China

**ABSTRACT**
With a global COVID-19 pandemic, the number of confirmed patients increases rapidly, leaving the world with very few medical resources. Therefore, the fast diagnosis and monitoring of COVID-19 are one of the world’s most critical challenges today. Artificial intelligence-based CT image classification models can quickly and accurately distinguish infected patients from healthy populations. Our research proposes a deep learning model (WE-SAJ) using wavelet entropy for feature extraction, two-layer FNNs for classification and the adaptive Jaya algorithm as a training algorithm. It achieves superior performance compared to the Jaya-based model. The model has a sensitivity of $85.47 \pm 1.84$, specificity of $87.23 \pm 1.67$ precision of $87.03 \pm 1.34$, an accuracy of $86.35 \pm 0.70$, and F1 score of $86.23 \pm 0.77$, Matthews correlation coefficient of $72.75 \pm 1.38$, and feature mutual information of $86.24 \pm 0.76$. Our experiments demonstrate the potential of artificial intelligence techniques for COVID-19 diagnosis and the effectiveness of the Self-adaptive Jaya algorithm compared to the Jaya algorithm for medical image classification tasks.

**ARTICLE HISTORY**
Received 16 December 2021
Accepted 19 February 2022

**KEYWORDS**
COVID-19; diagnosis; deep learning; Wavelet Entropy; self-adaptive Jaya; Jaya

1. Introduction
Covid-19 is a respiratory disease caused by the novel coronavirus SARS-CoV-2. Since the first cases appeared in 2019, Covid-19 has spread to most countries and territories worldwide (Hotez et al., 2021). Although there are several effective vaccines against the disease, they have not been able to definitively stop the spread of Covid-19 due to its high variability. Depending on the severity of the disease, patients with Covid-19 may suffer from mild respiratory symptoms such as cough and fever, severe pneumonia, multi-organ failure, or even death (Struyf et al., 2020). As the global pandemic progresses, the number of severe cases and deaths from Covid-19 continues to rise, a significant blow to human life and the global economy. Scholars in various fields have become highly concerned about the potentially severe consequences of Covid-19 and have maintained a constant interest in possible solutions to Covid-19. An important issue that continues to be addressed is the lack of medical resources due to the rapid increase in patients. The current standard method of Covid-19 detection is Reverse Transcription Polymerase Chain Reaction (RT-PCR) (Fang et al., 2020), which has the disadvantage of a high proportion of false negatives and often requires multiple tests to produce reliable results so that the process is highly time-consuming. Diagnostic methods using CT and X-Ray images of the chest also have the advantage of assessing the extent of a patient’s disease (Fang et al., 2020). However, this method requires a large number of medical experts to do the diagnosis task.

Meanwhile, the number of confirmed cases of Covid-19 is increasing rapidly. It is difficult to diagnose and monitor the vast number of Covid-19 patients promptly manually. Therefore, finding a quick and accurate diagnosis method is becoming one of the most critical tasks for stopping the spread of COVID-19.

Artificial intelligence has been a popular field of research in recent years, attracting many researchers to solve complex problems in various areas such as medicine, economics, and the cyber security field (Chen et al., 2005; Du et al., 2007; Han et al., 2008; Han & Huang, 2006; Wang et al., 2010; Wang & Huang, 2009). A significant advantage of AI is that it can be trained to replace humans with machines that perform repetitive and complex tasks (Du et al., 2006; Huang, 1999; Huang & Du, 2008; Meskó et al., 2018). The advantage of artificial intelligence is that it can solve the diagnostic difficulties associated with the rapid increase in patients. Many scholars believe that machine learning techniques on medical images can effectively diagnose COVID-19 patients (Wehbe et al., 2021). Many studies on machine learning techniques for chest X-ray and CT images of COVID-19...
patients have emerged, and some of them have achieved relatively good performance. Also, a significant number have implemented innovative and inspiring image processing methods on COVID-19.

(Szegedy et al., 2015) proposed a 22 layers deep network to do the image classification and detection task. Their model is suitable for COVID-19 diagnosis. Their research did not mention what optimisation algorithm they used, which could be a direction to further increase their model performance. (Lu, 2016) proposed a radial basis function (RBF) based model for brain disease diagnosis. Their model can be generalised for the diagnostic task of COVID-19, but did not achieve a stable and promising performance. (Chen, 2020) combined the greyscale co-occurrence matrix (GLCM) and support vector machine (SVM) to classify COVID-19 chest CT images and demonstrated the effectiveness of this method. (Yao & Han, 2020) worked on the COVID-19 Chest CT image classification task by using a wavelet entropy (WE) and biogeography-based optimisation (BBO) based method (WE-BBO). They have discovered the possible performance improvement of the combination of optimisation algorithm and Wavelet Entropy. Based on the WE-BBO, (Wang, 2021) proposed a Wavelet and Jaya combined method to do the same task and has improved, but their method needs to set the population size manually, which may cause the local optimal solution. The Jaya algorithm they used was proposed by (Rao, 2016), which can solve constrained and unconstrained optimisation problems as close to the optimal solution as possible while avoiding the worst solution. Moreover, this algorithm is parameter-free and straightforward to use.

(Rao & More, 2017) proposed a modified version of the Jaya algorithm, called the self-adaptive Jaya algorithm, which solves the problem of the Jaya algorithm requiring a set population size. Furthermore, this algorithm can automatically adjust the population size based on the current and previous population size and the current solution. Because of these advantages, self-adaptive Jaya can bring higher performance and application value to the model. We constructed this experiment on research (Wang, 2021) by replacing one of the Jaya algorithms and proposing a WE-SAJ model that combines the self-adaptive Jaya algorithm and Wavelet Entropy and made considerable progress compared to the previous research.

Our contributions are as follows: (i) We discovered a self-adaptive Jaya algorithm and Wavelet entropy combined method (WE-SAJ) for Covid-19 diagnosis; (ii) We demonstrated the performance improvement of the self-adaptive Jaya algorithm over the Jaya algorithm for medical image classification models; (iii) We have further demonstrated the value of AI technology for the COVID-19 diagnostic task.

The second part of the paper presents the dataset used for the experiments. The third part describes the main methods involved in the experiments. Finally, the fourth part presents and discusses the experimental results.

### 2. Dataset

We used a chest CT image dataset for the experiment. The dataset consisted of chest CT from 77 men and 55 women, for a total of 132 subjects. Each sample consisted of one complete chest CT image and the corresponding nucleic acid test results. The data set was divided into two groups, the COVID-19-infected group and the healthy group, and each group included 148 chest CT slices. The COVID-19-infected group consisted of chest CT images from 66 COVID-19 patients from the Fourth People’s Hospital in Huai’an, China, while the healthy group consisted of chest CT images from 66 healthy subjects. Table 1 shows the statistic of the dataset. Figure 1 shows the two samples from this dataset (Wang, 2021).

### Table 1. Statistic of the dataset.

| Class               | Ratio | No. of Images |
|---------------------|-------|---------------|
| Covid-19            | 50%   | 148           |
| Health Control      | 50%   | 148           |

### 3. Methodology

#### 3.1. Wavelet entropy

##### 3.1.1. Wavelet transform

The Fourier transform is widely used in many areas of signal analysis as a method that can transform a signal from the time domain to the frequency domain, and the form of the transform is shown in Equation (1).

\[
F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt,
\]

where \(\omega\) refers to the frequency, \(t\) refers to time.

Although the Fourier transform can analyse the spectrum of a signal and has a high application value, it has certain limitations when dealing with non-stationary signals. The Fourier transform can only capture which frequencies a section of the signal generally consists of and cannot reflect the moments when these frequencies occur, making it possible for two non-stationary signals that differ in the time domain to appear identical in the frequency domain (Saravanan & Ramachandran, 2010). Many signals in nature are non-stationary, and biology and medical signal analysis problems can rarely be solved using the straightforward Fourier transform as a solution.
A simple and feasible way to solve such problems is to decompose the entire time-domain signal into an infinite number of short-time signals. Then, making each short-time signal approximately smooth and using the Fourier transform on this basis to know the moment at which each frequency occurs. This decomposition process is known as adding windows, and this Fourier transform process based on the signal decomposition is known as the short-time Fourier transform (Allen & Rabiner, 1977). However, this method is limited by the width of the window. A too wide window will result in a low temporal resolution and a lack of refinement in the time domain. At the same time, a window that is too narrow will result in a poor frequency resolution and a lack of precision in frequency analysis. Furthermore, the window’s width does not transform during a short-time Fourier transform, so the short-time Fourier transform is also not the best solution for non-stationary signal analysis.

The wavelet transform replaces the infinitely long trigonometric basis of the Fourier transform with a finite decaying wavelet basis to locate the moment when the frequency occurs while obtaining the frequency (Quian Quiroga et al., 2001). The transformation equation is shown in Equation (2).

\[
W(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi \left( \frac{t - \tau}{a} \right) dt,
\]

where \(a\) represents scale, \(\tau\) represents translation, \(\psi\) represents parent wavelet function, and \(t\) represents time.

A wavelet is a wave that is more concentrated in the time domain than a sine wave in the Fourier transform, where the energy is finite and concentrated at a point. Wavelets can be used to efficiently extract information from a signal and analyse functions or signals at multiple scales of refinement through operational functions such as scaling and translation. The essence of the wavelet transform is similar to the Fourier transform in that a carefully selected basis represents the signal function. Each wavelet transform has a mother wavelet and a scaling function. The basis function of any wavelet transform is the set of scaling and translation of the mother wavelet and scaling function. This scaling and translation correspond to two variables in the wavelet transform function, the scale and the translation. The scale controls the scaling of the wavelet function, which corresponds to frequency, and the translation controls the translation of the wavelet function, which corresponds to time (Saritha et al., 2013). The wavelet transform can therefore capture which frequency components the signal contains at different moments, thus solving the shortcomings of the Fourier transform when it comes to analysing unstable signals.

### 3.1.2. Wavelet entropy

Wavelet entropy is a novel tool for analysing the instantaneous characteristics of non-stationary signals based on the wavelet transform combining wavelet decomposition and entropy. The original Shannon entropy was proposed to provide a valuable criterion for quantifying and comparing the energy distribution in wavelet subbands. Thus, it can be defined as Equation (3) (Yildiz et al., 2009).

\[
S(g) = -\sum_{i} P(g_i) \log_b P(g_i),
\]

where \(g\) refers to grey levels, and \(P\) refers to the probabilities of grey levels. Thus, \(S(g)\) can represent the energy distribution in wavelet subbands according to the probabilities of grey levels occurring.

However, most of the research on Shannon’s entropy has been on engineering applications, and its physical meaning and principles have not been discussed in depth. Moreover, the shortcomings of Shannon entropy
make it prone to wavelet mixing and energy leakage when dealing with non-stationary signals, which may lead to inaccurate or even incorrect results. Given this, many new solutions to these problems have emerged, such as relative wavelet entropy (Rosso et al., 2001) and Tsallis Wavelet Entropy (Chen & Li, 2014). Our research uses a 4-level decomposition of biorthogonal wavelets. Compared to orthogonal wavelet bases, biorthogonal wavelet bases resolve the incompatibility of symmetry and exact signal reconstruction. Biorthogonal wavelets consist of two wavelets called dyads, which decompose and reconstruct the signal separately. Bi-orthogonal wavelet bases resolve the incompatibility of symmetry and orthogonality requirements and are widely used in signal and image reconstruction. In this research, wavelet entropy is used for feature extraction. And then, the extracted features are fed into a two-layer Feedforward Neural Network for classification.

3.2. Feedforward neural network

The Feedforward Neural Network (FNN) (Jansen-Winkeln et al., 2021) is a typical deep learning model consisting of a multi-layer logistic regression model (continuous nonlinear function), also known as Single/Multi-Layer Perceptron, depending on the number of network layers (Han & Huang, 2008). Each network layer contains different numbers of neurones (perceptron). The structure of a perceptron can be defined as in Equation (4) (Venkata, 2017).

\[ f(x) = \sum_{i=1}^{m} w_i x_i + b, \]  

where \( w_i \) are weights, \( x \) is a set of inputs \( x_i \), \( m \) is the number of inputs to the perceptron, and \( b \) is the bias added after the sum of weighted inputs. Learning the value of \( w \) and \( b \) can help the model cope with different tasks.

A FNN usually consists of an input layer, several hidden layers, and an output layer (Han et al., 2010). In a FNN, the preceding network layer is the input layer. The last layer is the output layer, and the network layer between the input and output layers are hidden layers. Although a wide variety of classifiers already exist, when faced with classification tasks of more-approximate classes, neural networks can add non-linearity and change the representation of the data through hidden layers to better generalise the model (Rudolph, 1997).

In this study, a two-layer FNN is used with Mean Squared Error (MSE) as the loss function. A sample structure of FNN is shown in Figure 2, \( m \) is a hyperparameter representing the number of hidden layers. The number of perceptrons in each hidden layer is also a hyperparameter. \( c \) represents the number of input \( x_c \), \( d \) represents the number of outputs \( y_d \), which depends on the number of output classes. For example, for a 5-classes classification task, \( d \) can be equal to 5.

3.3. Self-adaptive Jaya algorithm

3.3.1. Jaya algorithm

The Jaya algorithm is a population-based heuristic algorithm (Cheng, 2018). In the Jaya algorithm, the search agent updates the current value based on the optimal known value and the worst known value, continuously approaching the optimal solution while avoiding the worst one. Thus, this approach helps converge to the optimal global solution (Rao, 2016), (Degertekin et al., 2021). Compared to traditional evolutionary algorithms, the Jaya algorithm is parameter-free, effectively avoiding the situation where traditional evolutionary algorithms produce locally optimal solutions after incorrect algorithmic tuning (Han, 2018). Moreover, it has more excellent reliability and generalisation than other heuristics and is widely used in optimisation applications and research in several fields. The basic equations of the Jaya algorithm are shown in Equation (5) (Zhao, 2018).

\[ Y_{q,r,i} = Y_{q,r,i} + r_1(Y_{q,\text{best},i} - |Y_{q,r,i}|) - r_2(Y_{q,\text{worst},i} - |Y_{q,r,i}|), \]

where \( r_1 \) and \( r_2 \) are two subjective random numbers taken randomly from \([0,1]\), \( Y_{q,\text{best},i} \) refers to the best solution, \( Y_{q,\text{worst},i} \) refers to the worst solution, \( q = 1, 2, \ldots, n \), and \( i = 1, 2, \ldots, d \).

3.3.2. Self-adaptive Jaya algorithm

This study used a modified Jaya algorithm called Self-adaptive Jaya algorithm (Rao & More, 2017). This method divides the solution into groups based on quality distinctions, distributed in the search space, to obtain the best solution. The most important feature of this algorithm is that the population size can be determined automatically, and its population development is shown in Equation (6) (Ravipudi & Neebha, 2018).

\[ n_{\text{new}} = \text{round}(n_{\text{old}} + r n_{\text{old}}), \]

where \( n_{\text{new}} \) represents the new population, \( n_{\text{old}} \) represents the old population, and \( r \) is the relative population development rate, a random value taken from \([-0.5, 0.5]\). Since \( r \) is random, the size of the new population may be larger or smaller than the old population.

- Only the current best population can enter the next generation when the new population is smaller than the old population.
- When the new population equals the old population, no change occurs.
When the new population is larger than the old population, all current populations enter the next generation.

The flow chart of the Self-adaptive Jaya algorithm is shown in Figure 3. This research uses the Self-adaptive Jaya algorithm as the training algorithm.

3.4. K-fold cross-validation

In machine learning, researchers usually divide the dataset into a training set, which is used for model training, and a test set, which is used to test model performance and thus improve the generalisation of the model. However, machine learning is a data-driven science. The size of the dataset has a significant impact on the model’s performance, with larger amounts of data tending to train higher performance models. However, many studies face difficulties with data scarcity, and the division of datasets into training and testing sets severely reduces the amount of data available for training, thus affecting model performance. The core theory of Cross-Validation is to reuse data to increase the amount of data available for training, thus affecting model performance. The K-fold Cross-Validation (Rajasekaran & Rajwade, 2021) used in this study is a widely used Cross-Validation method. This method divides the dataset into pre-specified K groups, takes one of them without repeating as the test set, uses all the other data as the training set, and calculates the model performance using the training set. The training is repeated K times while ensuring that a different data set is switched as the test set. Finally, get the final performance ($P_{final}$) from the model performances over the K tests. Figure 4 illustrates a concrete form of K-fold Cross-Validation. To obtain a more reliable and robust result, we used a 10-fold Cross-Validation to divide the dataset.

3.5. Evaluation

In this study, we evaluated the model performance by multiple methods. To be specific, they are Accuracy, Sensitivity, Precision, Specificity, F-score, Area Under the ROC Curve (AUC) and Confusion Matrix.

3.5.1. Confusion matrix

Confusion Matrix is a visualisation tool used to see how the model performs on each class and is represented as a matrix of n rows and n columns, where n is the number of classes in the classification task. The confusion matrix consists of four central values: True Positive (TP), the number of positive samples that are correctly predicted as positive), True Negative (TN), the number of negative samples that are correctly predicted as negative), False Positive (FP), the number of negative samples that are incorrectly predicted as positive) and False Negative (FN), the number of positive samples incorrectly predicted as negative), as shown in Figure 5.

3.5.2. Accuracy

Accuracy is a widely used indicator to assess the performance of machine learning models, and it represents the proportion of correctly determined samples to the total number of samples. In general, the higher the accuracy, the better the classifier. Equation (7) is the formula of accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

3.5.3. Precision

Precision is the proportion of correctly classified samples among all data samples in the dataset. The calculation of
Figure 3. The flow chart of the self-adaptive Jaya algorithm.

Figure 4. Illustration of dataset divide and performance calculation using K-fold cross-validation.

Figure 5. Confusion matrix sample.

Precision is shown as Equation (8).

\[
\text{Precision} = \frac{TP}{TP + FP}. \quad (8)
\]

3.5.4. Specificity

Specificity, also known as true negative rate, represents the ability of the model to distinguish between negative samples. The calculation of specificity is shown as Equation (9).

\[
\text{Specificity} = \frac{TN}{TN + FP}. \quad (9)
\]
3.5.5. Sensitivity
Sensitivity, also known as true positive rate, represents the ability of the model to differentiate between positive samples. The calculation sensitivity is shown as Equation (10).

\[
\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}.
\]  

(10)

3.5.6. F-Score
F-Score is the benchmark generated by the combination of Precision and Sensitivity. The higher the F1-score value, the higher the stability of the classification model. The calculation of precision is shown as Equation (11).

\[
\text{F1 score} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}}.
\]  

(11)

3.5.7. AUC
AUC, a nonparametric statistic that is not affected by category distribution (Wu & Flach, 2005), is one of the most common indicators used to evaluate binary classification models. AUC evaluates the model performance by calculating the area under the curve (ROC curve) of the model’s classification results on an axis with a false positive rate on the horizontal axis and a true positive rate on the vertical axis. Since the AUC considers the classification ability of the model for both positive and negative cases, it is still possible to evaluate the classifier reasonably well, even in the case of sample imbalance. Assume the ROC is constructed by \( m \) points, \((x_1, y_1), (x_2, y_2), \cdots , (x_m, y_m)\), the estimation calculation of AUC can be defined as Equation (12) (Lee, 2019):

\[
\text{AUC} = \frac{1}{2} \sum_{i=1}^{m-1} (x_{i+1} - x_i) \cdot (y_i + y_{i+1}).
\]  

(12)

4. Experiment results and discussions
4.1. WE results
Figure 6 illustrates a sample of 4-level decomposition of biorthogonal. Figure 6(a) is the result of the first-level wavelet transform, shows the input map’s low-frequency sub-bands (upper left corner) and high-frequency sub-bands (upper right corner, lower-left corner, and lower right corner). Since the wavelet transform introduces downsampling, the edge lengths of the four self-band maps are general to the input. In the second-level wavelet transform Figure 6(b), the low-frequency self-bands and the high-frequency sub-bands are obtained after a similar transformation of the low-frequency sub-bands obtained from the first-level wavelet transform. The results of the third-level wavelet transform – as shown in Figure 6(c) – and the fourth-level wavelet transform – as shown in Figure 6(d) – are obtained after a recursive operation.

| Table 2. Results of 10 runs of 10-fold cross-validation. |
| Run | Sen | Spc | Prc | Acc | F1 | MCC | FMI |
|-----|-----|-----|-----|-----|----|-----|-----|
| 1   | 85.14 | 86.49 | 86.30 | 85.81 | 85.71 | 71.63 | 85.72 |
| 2   | 83.11 | 90.54 | 89.78 | 86.82 | 86.32 | 73.85 | 86.38 |
| 3   | 85.81 | 85.14 | 85.23 | 85.47 | 85.52 | 70.95 | 85.52 |
| 4   | 81.76 | 88.51 | 87.68 | 85.14 | 84.62 | 70.43 | 84.67 |
| 5   | 85.81 | 87.84 | 87.59 | 86.82 | 86.69 | 73.66 | 86.69 |
| 6   | 86.49 | 87.84 | 87.67 | 87.16 | 87.07 | 74.33 | 87.08 |
| 7   | 87.84 | 85.14 | 85.53 | 86.49 | 86.67 | 73.00 | 86.67 |
| 8   | 87.16 | 85.81 | 86.00 | 86.49 | 86.58 | 72.98 | 86.58 |
| 9   | 85.14 | 87.16 | 86.90 | 86.15 | 86.01 | 72.31 | 86.01 |
| 10  | 86.49 | 87.84 | 87.67 | 87.16 | 87.07 | 74.33 | 87.08 |
| MSD | 84.21 | 86.49 | 86.03 | 86.23 | 86.23 | 72.75 | 86.24 |
| ±1.84 | ±1.67 | ±1.34 | ±0.70 | ±0.77 | ±1.38 | ±0.76 |

(Sen = Sensitivity; Spc = Specificity; Prc = Precision; Acc = Accuracy; F1 = F1 Score; MCC = Matthews correlation coefficient; FMI = Feature Mutual Information)

| Table 3. Performance comparison between WE-Jaya and WE-SAJ. |
| Method | Sen | Spc | Prc | Acc | F1 | MCC | FMI |
|--------|-----|-----|-----|-----|----|-----|-----|
| WE-Jaya (Wang, 2021) | 78.31 | 78.11 | 77.03 | 75.71 | 75.10 | 51.51 | 75.14 |
| WE-SAJ (Ours) | 85.47 | 87.23 | 87.03 | 86.35 | 86.23 | 72.75 | 86.24 |
| ±1.84 | ±1.67 | ±1.34 | ±0.70 | ±0.77 | ±1.38 | ±0.76 |

(Sen = Sensitivity; Spc = Specificity; Prc = Precision; Acc = Accuracy; F1 = F1 Score; MCC = Matthews correlation coefficient; FMI = Feature Mutual Information)

Note here that the image in the sample is painted with pseudo-colour, which is still actually represented as greyscale.

4.2. Statistical results
WE-SAJ used Wavelet Entropy as the feature extraction method, 2-layer FNN as the classifier, the Self-adaptive Jaya algorithm as the training algorithm, and K-fold cross-validation to report unbiased performance. As a result, the experiments achieved great performance (shown in Table 2) with an average sensitivity of 85.47±1.84, specificity of 87.23±1.67 precision of 87.03±1.34, the accuracy of 86.35±0.70, and F1 score of 86.23±0.77, Matthews correlation coefficient of 72.75±1.38, and feature mutual information of 86.24±0.76.

4.3. Self-adaptive Jaya compared to Jaya
Table 3 shows the specific numerical performance comparison between the Self-adaptive Jaya algorithm-based model (WE-SAJ) and the previous research (WE-Jaya) based on the Jaya algorithm. Compared to WE-Jaya, WE-SAJ has achieved significant improvements in all performance metrics.

The WE-SAJ has improved accuracy by more than ten percentage points, which means our method has a higher
practical value. More detailed performance improvements can be seen in the other performance indicators. WE-SAJ improves sensitivity by more than 12 percentage points and specificity by more than 11 percentage points. It suggests that WE-SAJ can reduce unnecessary healthcare resources by misdiagnosing as few healthy people as possible while ensuring that more infected people are correctly identified to identify COVID-19 patients effectively. It shows that WE-SAJ can ensure that more infected patients are correctly identified to effectively identify COVID-19 patients while minimising misdiagnosis of healthy people and reducing unnecessary wastage of healthcare resources. The improvement in F1-score also demonstrates that the model achieves better overall performance with equal weighting of precision and sensitivity. The increase of over 11 percentage points in the FMI index indicates the higher relevance of the data features extracted by the model to the data labels, which means that the model has improved its ability to extract useful features.

This series of improvements can be attributed to that the automatic population sizing feature of Adaptive Jaya can effectively set the most appropriate population size, thus enhancing the tracking of the optimal solution and ultimately helping the model achieve better performance.

Figure 7 illustrates the ROC curve and AUC value of WE-Jaya (a) and WE-SAJ (b), respectively. Each point in the graph corresponds to a threshold. When the threshold is maximum, True Positive Rate (TPR) = False Positive Rate (FPR) = 0, which corresponds to the origin (0,0) of the graph. TPR = FPR = 1 corresponds to the point (1,1) in the upper right corner when the threshold is minimal. As the Threshold increases, both TPR and FPR increase. The ROC curve shows that the WE-SAJ can achieve a lower False Positive Rate and higher True Positive Rate than the Jaya-based model at most thresholds, which leads to a lower AUC for WE-SAJ than the Jaya-based model. It indicates that the WE-SAJ has more diagnostic value than WE-Jaya.

Figure 6. A sample of 4-level decomposition of biorthogonal wavelets result.
Figure 7. ROC curve and AUC comparison between WE-Jaya and WE-SAJ.

Table 4. Performance comparison to State-of-the-art Approaches.

| Method          | Sen  | Spc  | Prc  | Acc  | F1   | MCC  | FMI  |
|-----------------|------|------|------|------|------|------|------|
| RBFNN (Lu, 2016) | 66.89| 75.47| 73.23| 71.18| 69.88| 42.56| 69.97|
| ±2.43 ±2.53 ±1.48 ±1.80 ±1.08 ±1.61 ±1.04 |
| WE-BBO (Yao & Han, 2020) | 72.97| 74.93| 74.48| 73.95| 73.66| 47.99| 73.66|
| ±2.96 ±2.39 ±1.34 ±0.98 ±0.98 ±2.00 ±1.33 |
| GLCM-SVM (Chen, 2020) | 72.03| 78.04| 76.66| 75.03| 74.24| 50.20| 74.29|
| ±3.94 ±1.72 ±1.07 ±1.12 ±1.57 ±2.17 ±1.53 |
| WE-JAYA (Wang, 2021) | 73.31| 78.11| 77.63| 75.71| 75.10| 51.51| 75.14|
| ±2.26 ±1.92 ±1.35 ±1.04 ±1.23 ±2.07 ±1.22 |
| GoogLeNet (Szegedy et al., 2015) | 77.64| 83.85| 82.82| 80.74| 80.12| 61.65| 80.17|
| ±2.22 ±2.00 ±1.54 ±0.91 ±1.07 ±1.81 ±1.05 |
| WE-SAJ | 85.47| 87.23| 87.03| 86.35| 86.23| 72.75| 86.24|
| ±1.84 ±1.67 ±1.34 ±0.70 ±0.77 ±1.38 ±0.76 |

(Sen = Sensitivity; Spc = Specificity; Prc = Precision; Acc = Accuracy; F1 = F1 Score; MCC = Matthews correlation coefficient; FMI = Feature Mutual Information)

4.4. Comparison to state-of-the-art approaches

Compared to the other state-of-art approaches in CT image classification of COVID-19, WE-SAJ shows significant improvement in all aspects, as shown in Table 4 for numerical comparison. The model can obtain the overall improvement to these SOTA models because of the advantages of the Self-adaptive Jaya algorithm. It can automatically adjust the population size based on the current and the previous population size. The current solution avoids the local optimal solution and increases the possibility of finding the global optimal solution. The results also demonstrate that our method is promising for the CT image classification task of COVID-19 and has great scope for improvement.

5. Conclusions

Artificial intelligence-based medical image analysis techniques have a significant application in the fight against COVID-19. It can help in the diagnosis of COVID-19 challenges caused by the lack of medical resources. This experiment validates the feasibility of the Wavelet Entropy and Self-adaptive algorithm-based model for the COVID-19 chest CT image classification task and achieved auspicious performance. It is highly generalisable and theoretically applicable to all types of medical image classification tasks, which needs to be further validated in future studies. Based on the experimentally obtained performance, we have reasons to believe that with more optimisations and improvements in the near future, we can obtain models with better performance for diagnosis and recognition of COVID-19 and even more diseases and solve more medical challenges.

Acknowledgement

The paper was partially supported by: Royal Society International Exchanges Cost Share Award, UK (RP202G0230); Medical Research Council Confidence in Concept Award, UK (MC_PC_17171); Hope Foundation for Cancer Research, UK (RM60G0680); British Heart Foundation Accelerator Award, UK AA/18/3/34220; Sino-UK Industrial Fund, UK (RP202G0289); Global Challenges Research Fund (GCRF), UK (P202PF11).
Disclosure statement
No potential conflict of interest was reported by the author(s).

Funding
This work was supported by Royal Society International Exchanges Cost Share Award, UK [Grant Number RP202G0230]; Medical Research Council [Grant Number MC_PC_17171]; Sino-UK Industrial Fund, UK [Grant Number RP202G0289]; British Heart Foundation [Grant Number AA/18/3/34220].

References
Allen, J. B., & Rabiner, L. R. (1977). A unified approach to short-time Fourier analysis and synthesis. Proceedings of the IEEE, 65(11), 1558–1564. https://doi.org/10.1109/PROC.1977.10770
Chen, J., & Li, G. (2014). Tsallis wavelet entropy and its application in power signal analysis. Entropy, 16(6), 3009–3025. https://doi.org/10.3390/e16063009
Chen, W.-S., Yuen, P. C., Huang, J., & Dai, D.-Q. (2005). Kernel machine-based one-parameter regularized fisher discriminant method for face recognition. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 35(4), 659–669. https://doi.org/10.1109/TSMCB.2005.844596
Chen, Y. (2020). COVID-19 classification based on gray-level co-occurrence matrix and support vector machine. In K. C. Santosh, & A. Joshi (Eds.), COVID-19: Prediction, Decision-Making, and Its Impacts (pp. 47–55). Springer Singapore.
Cheng, H. (2018). Multiple sclerosis identification based on fractional Fourier entropy and a modified Jaya algorithm. Entropy, 20(4), Article ID: 254. https://doi.org/10.3390/e18040130
Degertekin, S. O., Bayar, G. Y., & Lamberti, L. (2021). Parameter free Jaya algorithm for truss sizing-layout optimization under natural frequency constraints. Computers & Structures, 245, 29. https://doi.org/10.1016/j.compstruc.2020.106461
Du, J.-X., Huang, D.-S., Wang, X.-F., & Gu, X. (2007). Shape recognition based on neural networks trained by differential evolution algorithm. Neurocomputing, 70(4-6), 896–903. https://doi.org/10.1016/j.neucom.2006.10.026
Du, J.-X., Huang, D.-S., Zhang, G.-J., & Wang, Z.-F. (2006). A novel full structure optimization algorithm for radial basis probabilistic neural networks. Neurocomputing, 70(1-3), 592–596. https://doi.org/10.1016/j.neucom.2005.05.003
Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang, P., & Ji, W. (2020). Sensitivity of chest CT for COVID-19: comparison to RT-PCR. Radiology, 296(2), E115–E117. https://doi.org/10.1148/radiol.2020200432
Han, F., & Huang, D.-S. (2006). Improved extreme learning machine for function approximation by encoding a priori information. Neurocomputing, 69(16-18), 2369–2373. https://doi.org/10.1016/j.neucom.2006.02.013
Han, F., & Huang, D.-S. (2008). A new constrained learning algorithm for function approximation by encoding a priori information into feedforward neural networks. Neural Computing and Applications, 17(5), 433–439. https://doi.org/10.1007/s00521-007-0135-5
Han, F., Ling, Q.-H., & Huang, D.-S. (2008). Modified constrained learning algorithms incorporating additional functional constraints into neural networks. Information Sciences, 178(3), 907–919. https://doi.org/10.1016/j.ins.2007.09.008
Han, F., Ling, Q.-H., & Huang, D.-S. (2010). An improved approximation approach incorporating particle swarm optimization and a priori information into neural networks. Neural Computing and Applications, 19(2), 255–261. https://doi.org/10.1007/s00521-009-0274-y
Han, L. (2018). Identification of alcoholism based on wavelet renyi entropy and three-segment encoded Jaya algorithm. Complexity, 2018(2018), Article ID 3198184.
Hotez, P. J., Fenwick, A., & Molyneux, D. (2021). The new COVID-19 poor and the neglected tropical diseases resurgence. Infectious Diseases of Poverty, 10(1), 3. https://doi.org/10.1186/s40249-020-00784-2
Huang, D.-s. (1999). Radial basis probabilistic neural networks: Model and application. International Journal of Pattern Recognition and Artificial Intelligence, 13(07), 1083–1101. https://doi.org/10.1142/S0218001499000604
Huang, D.-S., & Du, J.-X. (2008). A constructive hybrid structure optimization methodology for radial basis probabilistic neural networks. IEEE Transactions on Neural Networks, 19(12), 2099–2115. https://doi.org/10.1109/TNN.2008.2004370
Jansen-Winkeln, B., Barberio, M., Chalopin, C., Schierke, K., Diana, M., Köhler, H., Gockel, I., & Maktabi, M. (2021). Feedforward artificial neural network-based colorectal cancer detection using hyperspectral imaging: A step towards automatic optical biopsy. Cancers, 13(5), 13. https://doi.org/10.3390/cancers13050967
Lee, J.-S. (2019). AUC4: S. AUC-based C4.5 decision tree algorithm for imbalanced data classification. IEEE Access, 7, 106034–106042. https://doi.org/10.1109/ACCESS.2019.2931865
Lu, Z. (2016). A pathological brain detection system based on radial basis function neural network. Journal of Medical Imaging and Health Informatics, 6(5), 1218–1222. https://doi.org/10.1166/jmih.2016.1901
Meskø, B., Hetényi, G., & Györrfy, Z. (2018). Will artificial intelligence solve the human resource crisis in healthcare? BMC Health Services Research, 18(1), 1–4. https://doi.org/10.1186/s12913-018-3359-4
Quián Quiroga, R., Rosso, O. A., Basar, E., & Schürmann, M. (2001). Wavelet entropy in event-related potentials: A new method shows ordering of EEG oscillations. Biological Cybernetics, 84(4), 291–299. https://doi.org/10.1007/s004220000212
Rajasekaran, S., & Rajwade, A. (2021). Analyzing cross-validation of compressed sensing with poison noise. Signal Processing, 182, 9. https://doi.org/10.1016/j.sigpro.2020.107947
Rao, R. (2016). Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems. International Journal of Industrial Engineering Computations, 7(1), 19–34.
Rao, R., & More, K. (2017). Design optimization and analysis of selected thermal devices using self-adaptive Jaya algorithm. Energy Conversion and Management, 140, 24–35. https://doi.org/10.1016/j.enconman.2017.02.068
Ravipudi, J. L., & Neebha, M. (2018). Synthesis of linear antenna arrays using Jaya, self-adaptive jaya and chaotic jaya algorithms. AEU-International Journal of Electronics and Communications, 92, 54–63. https://doi.org/10.1016/j.aeue.2018.05.022
Rosso, O. A., Blanco, S., Yordanova, J., Kolev, V., Figliola, A., Schürmann, M., & Başar, E. (2001). Wavelet entropy: An improved approximation approach incorporating particle swarm optimization and a priori information into neural networks. Neural Computing and Applications, 19(2), 255–261. https://doi.org/10.1007/s00521-009-0274-y
Rudolph, S. (1997). On topology, size and generalization of non-linear feed-forward neural networks. *Neurocomputing*, 16(1), 1–22. https://doi.org/10.1016/S0925-2312(96)00059-8

Saravanan, N., & Ramachandran, K. (2010). Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN). *Expert Systems with Applications*, 37(6), 4168–4181. https://doi.org/10.1016/j.eswa.2009.11.006

Saritha, M., Joseph, K. P., & Mathew, A. T. (2013). Classification of MRI brain images using combined wavelet entropy based spider web plots and probabilistic neural network. *Pattern Recognition Letters*, 34(16), 2151–2156. https://doi.org/10.1016/j.patrec.2013.08.017

Struyf, T., et al. (2020). Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19 disease. *Cochrane Database of Systematic Reviews*, 7(7), Article ID: CD013665.

Szegedy, C., et al. (2015). Going deeper with convolutions. IEEE conference on computer vision and pattern recognition (CVPR). Boston, MA: IEEE.

Venkata, R. R. (2017). Abnormal breast detection in mammogram images by feed-forward neural network trained by Jaya algorithm. *Fundamenta Informaticae*, 151(1–4), 191–211. https://doi.org/10.3233/FI-2017-1487

Wang, W. (2021). COVID-19 detection by wavelet entropy and Jaya. *Lecture Notes in Computer Science*, 12836, 499–508. https://doi.org/10.1007/978-3-030-84522-3_41

Wang, X.-F., & Huang, D.-S. (2009). A novel density-based clustering framework by using level set method. *IEEE Transactions on Knowledge and Data Engineering*, 21(11), 1515–1531. https://doi.org/10.1109/TKDE.2009.21

Wang, X.-F., Huang, D.-S., & Xu, H. (2010). An efficient local chain–vese model for image segmentation. *Pattern Recognition*, 43(3), 603–618. https://doi.org/10.1016/j.patcog.2009.08.002

Wehbe, R. M., Sheng, J., Dutta, S., Chai, S., Dravid, A., Barutcu, S., Wu, Y., Cantrell, D. R., Xiao, N., Allen, B. D., MacNealy, G. A., Savas, H., Agrawal, R., Parekh, N., & Katsaggelos, A. K. (2021). DeepCOVID-XR: An artificial intelligence algorithm to detect COVID-19 on chest radiographs trained and tested on a large US clinical data set. *Radiology*, 299(1), E167–E176. https://doi.org/10.1148/radiol.2020203511

Wu, S., & Flach, P. (2005). A scored AUC metric for classifier evaluation and selection. In *Second workshop on ROC analysis in ML, Bonn, Germany* (pp. 1–6). ACM.

Yao, X., & Han, J. (2020). COVID-19 detection via wavelet entropy and biogeography-based optimization. In K. C. Santosh & A. Joshi (Eds.), *COVID-19: Prediction, decision-making, and its impacts* (pp. 69–76). Springer.

Yıldız, A., Akin, M., Poyraz, M., & Kirbas, G. (2009). Application of adaptive neuro-fuzzy inference system for vigilance level estimation by using wavelet-entropy feature extraction. *Expert Systems with Applications*, 36(4), 7390–7399. https://doi.org/10.1016/j.eswa.2008.09.003

Zhao, G. (2018). Smart Pathological brain detection by synthetic minority oversampling technique, extreme learning machine, and Jaya algorithm. *Multimedia Tools and Applications*, 77(17), 22629–22648. https://doi.org/10.1007/s11042-018-6359-9