frontiers in Neuroscience

ORIGINAL RESEARCH
published: 13 September 2021
doi: 10.3389/fnins.2021.737785

Multiple Sclerosis Recognition by Biorthogonal Wavelet Features and Fitness-Scaled Adaptive Genetic Algorithm

Shui-Hua Wang[1†], Xianwei Jiang[2†] and Yu-Dong Zhang[3]*

[1] School of Mathematics and Actuarial Science, University of Leicester, Leicester, United Kingdom, [2] Nanjing Normal University of Special Education, Nanjing, China, [3] School of Informatics, University of Leicester, Leicester, United Kingdom

Aim: Multiple sclerosis (MS) is a disease, which can affect the brain and/or spinal cord, leading to a wide range of potential symptoms. This method aims to propose a novel MS recognition method.

Methods: First, the bior4.4 wavelet is used to extract multiscale coefficients. Second, three types of biorthogonal wavelet features are proposed and calculated. Third, fitness-scaled adaptive genetic algorithm (FAGA)—a combination of standard genetic algorithm, adaptive mechanism, and power-rank fitness scaling—is harnessed as the optimization algorithm. Fourth, multiple-way data augmentation is utilized on the training set under the setting of 10 runs of 10-fold cross-validation. Our method is abbreviated as BWF-FAGA.

Results: Our method achieves a sensitivity of 98.00 ± 0.95%, a specificity of 97.78 ± 0.95%, and an accuracy of 97.89 ± 0.94%. The area under the curve of our method is 0.9876.

Conclusion: The results show that the proposed BWF-FAGA method is better than 10 state-of-the-art MS recognition methods, including eight artificial intelligence-based methods, and two deep learning-based methods.

Keywords: multiple sclerosis, recognition, biorthogonal wavelet transform, fitness scaling, genetic algorithm, multiple-way data augmentation

OPEN ACCESS

Edited by:
Yuankai Huo,
Vanderbilt University, United States

Reviewed by:
Pengpeng Pi,
Henan Polytechnic University, China
Mingzhou Lu,
Nanjing Agricultural University, China

****Correspondence:***
Yu-Dong Zhang
yudong.zhang@le.ac.uk

[†] These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *07 July 2021*
Accepted: *05 August 2021*
Published: *13 September 2021*

Citation:
Wang S-H, Jiang X and Zhang Y-D (2021) Multiple Sclerosis Recognition by Biorthogonal Wavelet Features and Fitness-Scaled Adaptive Genetic Algorithm. Front. Neurosci. 15:737785. doi: 10.3389/fnins.2021.737785

INTRODUCTION

Multiple sclerosis (MS) is a disease, which can affect the brain and/or spinal cord, leading to a wide range of potential symptoms, including problems with balance control (Allum et al., 2021), binocular vision (Gil-Casas et al., 2021), finger movement (Viatkin et al., 2021), etc. MS is a lifelong condition, which occasionally causes moderate to severe disabilities (Livne-Margolin et al., 2021).

In practice, MS patients may be wrongly mixed with other white matter disorders, e.g., neuromyelitis optica (Sousa et al., 2021), acute disseminated encephalomyelitis, acute cerebral infarction (Moreno-Andrade et al., 2021), etc. Consequently, error-free MS diagnosis is of extreme significance to patients, as they can be allowed more time to decide on the subsequent treatments to help control the condition, such as treating relapses (Hartung et al., 2021) with short courses of steroid medicine (Blechinger et al., 2021).

In the last decade, artificial intelligence (AI)-based techniques are widely harnessed to develop novel calculation models for identifying MS. Loizou et al. (2011) presented a method called "multiscale amplitude modulation and frequency modulation (MAMFM)" for detecting MS. Nayak et al. (2016) used two-dimensional discrete wavelet transform (2D-DWT) to identify brain diseases. Their method is adapted to the MS recognition task in this study. Zhan and Chen (2016) combined biorthogonal wavelet transform (BWT) and logistic regression (LR) for MS recognition. Lopez (2017) used a Haar wavelet transform (HWT). Their experiment showed three-level HWT (L3-HWT) could achieve the best performance. Zhou and Shen (2018) combined gray-level co-occurrence matrix (GLCM) and biogeography-based optimization (BBO) for MS identification. Yahia et al. (2018) presented a new method—decimal descriptor pattern (DDP)—to assess MS lesions. Han and Hou (2019) combined wavelet entropy (WE) and adaptive genetic algorithm (AGA) to detect MS. The authors chose a three-level decomposition using the db2 wavelet. The shortcoming of their method is that WE may remove some important information. Besides, the AGA algorithm can still be improved. Han and Hou (2020) combined Hu moment invariant (HMI) and particle swarm optimization (PSO) for MS recognition.

Recently, deep learning (DL) attracts research attention from a diversity of academic fields. Tang (2021) presented a five-layer convolutional neural network (5L-CNN) for MS detection. Salem et al. (2020) presented a fully convolutional neural network (FCNN) to detect lesions in MS. All the AI-based methods and DL-based methods mentioned earlier are adapted and used as comparison baseline methods in this study.

This study proposes an improved algorithm based on Han and Hou (2019) instead of DL, as our dataset is small. The experiments show that our method is better than the two DL methods. Four improvements are proposed based on Han and Hou (2019). First, we use bior4.4 to replace the db2 wavelet in Han and Hou (2019). Second, we extract three different types of features, whereas Han and Hou (2019) only uses the entropy feature. Third, a fitness-scaled adaptive genetic algorithm (FAGA) is introduced to replace the AGA in Han and Hou (2019). Fourth, multiple-way data augmentation (MDA) is used. In all, the contributions of this study are fivefold:

(i) The bior4.4 wavelet is used to extract multiscale coefficients.
(ii) Three types of biorthogonal wavelet features (BWFs) are proposed and calculated.
(iii) FAGA is utilized as the optimization algorithm.
(iv) Multiple-way DA is harnessed on the training set to avoid overfitting when training classifiers.
(v) Our method is superior to 10 state-of-the-art methods, including two DL methods.

DATA PREPROCESSING

Sources of This Dataset

The dataset in this study is obtained from Han and Hou (2019). Its demographic description is itemized in **Table 1**, where two categories exist: (i) MS and (ii) healthy control (HC). Within the dataset, MS images are obtained from Ref (eHealth Lab, 2021) and HC images from Pan (2018). **Figure 1A** presents a raw MS brain slice I, and **Figure 1B** provides the corresponding annotated result with two MS plaques. **Figure 1C** presents another raw MS brain slice II, and **Figure 1D** gives the corresponding annotated result with four plaques.

TABLE 1 | Demographic characteristics.

Category	NSu	NSl	Sex (m/f)	Age
MS (eHealth Lab, 2021)	38	676	17/21	34.1 ± 10.5
HC (Pan, 2018)	26	681	12/14	33.5 ± 8.3

NSu, Number of Subjects; NSl, Number of slices; m, male; and f, female.

Data Harmonization

The brain images in this dataset are obtained from two different sources. Hence, data harmonization (DH; Westenbrink et al., 2021) is necessary so that the AI models will not learn those deciding factors of different scanning machines. There are loads of DH techniques, such as histogram equalization (Misra et al., 2021), color adjustment (Hatayama et al., 2020), Gamma correction, histogram stretching (HS), etc.

Histogram stretching (Abdullah et al., 2021) is utilized due to its ease of implementation. Suppose α denotes the original brain slice, and φ stands for the processed slice. Let (w, v) denote the coordinates, HS operation HS : $\alpha \mapsto \varphi$ is defined via:

$$\varphi(w, v) = \frac{\alpha(w, v) - \alpha_l}{\alpha_h - \alpha_l}, \tag{1}$$

where (α_l, α_h) stand for the lowest and highest grayscale intensity values of the original brain slice α, respectively.

$$\alpha_l = \min_w \min_v \alpha(w, v), \tag{2}$$

$$\alpha_h = \max_w \max_v \alpha(w, v). \tag{3}$$

METHODOLOGY

Wavelet Decomposition

Table 2 lists all abbreviations used in this paper and their cognate meanings. To decompose signals at different scales, the wavelet decomposition technique (Zucatelli et al., 2021) is employed. For a given signal $s(t)$, two subbands will be generated: (i) the approximate subband stands for the low frequency information of $s(t)$, whilst (ii) the detail subband stands for the high frequency information of $s(t)$.

In practice, discrete wavelet transform (DWT) is used to transform $s(t)$ into the wavelet domain (Ren et al., 2021). DWT achieves multistage and multiscale transformation by transmitting the previous approximation subband C^A to the quadrature mirror filters (Singh et al., 2021). Compared with the standard Fourier transform with several shortcomings, such as complex signal spectrum caused by noise and unintuitive

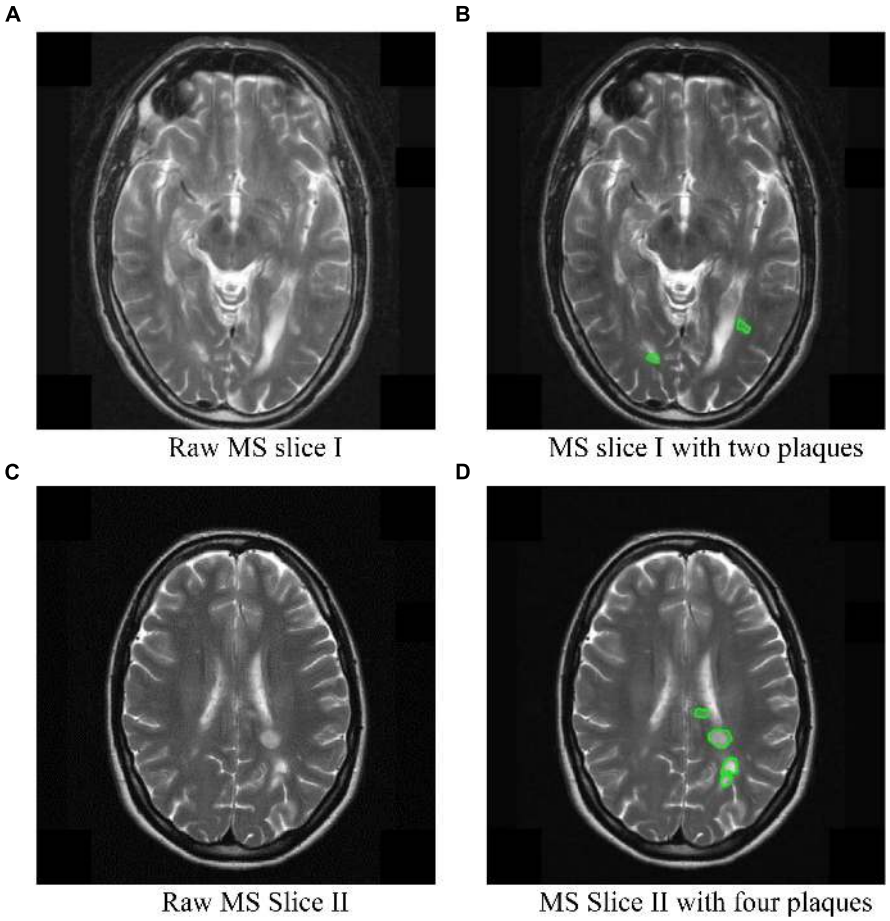

FIGURE 1 | MS samples. **(A)** Raw MS slice I. **(B)** MS slice I with two plaques. **(C)** Raw MS Slice II. **(D)** MS Slice II with four plaques.

representation, DWT is advantageous in multiscale time/space resolution (Rahim et al., 2021).

Let $s(t)$ be a particular one-dimension signal, then the continuous wavelet transform of $s(t)$ is shown below.

$$C_\gamma\left(h_s, h_t\right) = \int_{-\infty}^{\infty} s(t) \times \gamma\left(t \mid h_s, h_t\right) \mathrm{d}t, \tag{4}$$

where the C is the wavelet coefficient, γ is a mother wavelet. $\gamma\left(t \mid h_s, h_t\right)$ is defined as:

$$\gamma\left(t \mid h_s, h_t\right) = \frac{1}{\sqrt{h_s}} \gamma\left(\frac{t - h_t}{h_s}\right), h_s > 0, h_t > 0, \tag{5}$$

where the h_s stands for the scale factor and h_t the translation factor.

In DWT, Eq. (4) is discretized by replacing h_s and h_t with two discrete variables c and u, where c represents the discrete value of scale factor (Kshatriya and Prasanna, 2021) and u the discrete value of translation factor.

$$h_s = 2^c, \tag{6}$$

$$h_t = u \times 2^c. \tag{7}$$

Also, the raw signal $s(t)$ is discretized to $s(n)$, where n is the discrete form of variable t. In this way, the approximation subband $C^A(n|c, u)$ is obtained by

$$C^A(n \mid c, u) = f^{DS}\left[\sum_n s(n) \times f_A^*\left(\frac{n - 2^c u}{2^c}\right)\right]. \tag{8}$$

The detail subband $C^D(n|c, u)$ is obtained as:

$$C^D(n \mid c, u) = f^{DS}\left[\sum_n s(n) \times f_D^*\left(\frac{n - 2^c u}{2^c}\right)\right], \tag{9}$$

where f^{DS} is the down-sampling, $f_A(n)$ and $f_D(n)$ are the low-pass filter and the high-pass filter, respectively.

Two-Dimensional Discrete Wavelet Transform

Assuming a particular image is symbolized by S, then the 2D-DWT can be implemented by running row-wise and column-wise one-dimensional DWT in sequence. First, 2D-DWT decomposition runs on the raw image S. Afterward, four subbands (J_1, W_1, F_1, K_1) are generated, in which subband J is

TABLE 2 | Abbreviation list.

Abbreviation	Meaning
2D-DWT	Two-dimensional DWT
AI	Artificial intelligence
ANN	Artificial neural network
AUC	The area under the curve
BWF	Biorthogonal wavelet feature
BWT	Biorthogonal wavelet transform
CV	Computer vision
CWT	Continuous wavelet transform
DA	Data augmentation
DH	Data harmonization
DL	Deep learning
DWT	Discrete wavelet transform
FAGA	Fitness-scaled adaptive genetic algorithm
FMI	Fowlkes–Mallows index
FN	False negative
FNN	Feedforward neural network
FP	False positive
HC	Healthy control
HMI	Horizontally mirrored image
LR	Logistic regression
MCC	Matthews correlation coefficient
MDA	Multiple-way data augmentation
MS	Multiple sclerosis
MSD	Mean and standard deviation
ROC	Receiver operating characteristic
SGA	Standard genetic algorithm
TN	True negative
TP	True positive
WM	White matter

the horizontal quadrant, subband W is the vertical quadrant, subband F is the diagonal quadrant, and subband K denotes the approximate component quadrant (Sorkhabi et al., 2021). The subscript "1" means 1-level decomposition (Motlagh et al., 2021).

Assuming g^{D2} denotes a 2D-DWT decomposition operation, we get

$$\begin{pmatrix} K_1 & J_1 \\ W_1 & F_1 \end{pmatrix} = g^{D2}(S). \tag{10}$$

The subsequent decompositions run as:

$$\begin{pmatrix} K_{i+1} & J_{i+1} \\ W_{i+1} & F_{i+1} \end{pmatrix} = g^{D2}(K_i), i = 1, 2, \ldots L-1, \tag{11}$$

where L is the maximum decomposition level.

The subband K_1 can be further decomposed into four subbands (K_2, J_2, W_2, F_2) at the second level. The subband K_2 is then decomposed into (K_3, J_3, W_3, F_3). **Figure 2** presents a schematic of 4-level 2D-DWT decomposition. In our study, we will choose a L-level decomposition. The value of L will be investigated via trial-and-error method and reported in the following sections. The pseudocode of 2D-DWT is shown in Algorithm 1.

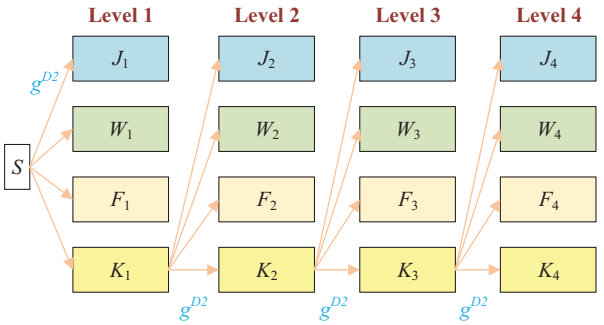

FIGURE 2 | A schematic of a four-level 2D-DWT.

Algorithm 1 Pseudocode of 2D-DWT

Input	Image S
Step 1	Decompose the image S into four subbands $S \mapsto \begin{pmatrix} K_1 & J_1 \\ W_1 & F_1 \end{pmatrix}$.
Step 2	for $i = 1 : L-1$ Select the current approximation subband K_i. K_i is decomposed into four subbands. $K_i \mapsto \begin{pmatrix} K_{i+1} & J_{i+1} \\ W_{i+1} & F_{i+1} \end{pmatrix}$. end
Output	Output the subbands $(K_L, J_L, W_L, F_L, J_{L-1}, W_{L-1}, F_{L-1}, \ldots, J_1, W_1, F_1)$.

Biorthogonal Wavelet Transform

There are loads of various wavelet families, such as Haar (Ganesan and Rajarajeswari, 2021), Daubechies (Kasnazani and Alipanah, 2021), etc. While we choose the BWT in this study. The advantage of BWT is that it allows more degrees of freedom compared to orthogonal wavelets. Besides, BWTs are compactly supported biorthogonal spline wavelets for which symmetry and exact reconstruction are possible with FIR filters (Han and Michelle, 2021).

The nomination of BWT is usually written as

$$\text{BWT} = \text{bior}(N_r).(N_d), \tag{12}$$

where N_r stands for the reconstruction order and N_d for the decomposition order. In this study, we choose bior4.4 wavelet, which indicates $N_r = N_d = 4$. The filters of low-pass and high-pass decomposition filters are shown in **Figures 3A,B**, and while the low-pass and high-pass reconstruction filters are displayed in **Figures 3C,D**.

Biorthogonal Wavelet Features

The $(3L+1)$ subbands $(K_L, J_L, W_L, F_L, J_{L-1}, W_{L-1}, F_{L-1}, \ldots, J_1, W_1, F_1)$ of BWT contain too many features. In order to reduce the number of features, scholar proposes various methods. de Sena et al. (2021) combined wavelet subband coefficients with entropy. Machorro-Lopez et al. (2021) presented a wavelet energy method for damage identification. Radi et al. (2019)

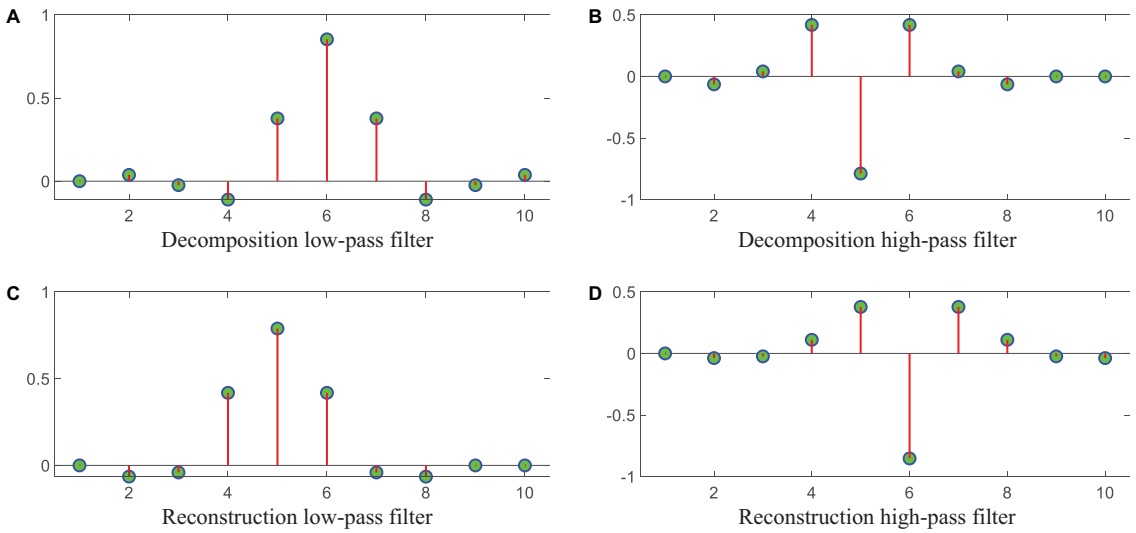

FIGURE 3 | Filters for bior4.4. **(A)** Decomposition low-pass filter. **(B)** Decomposition high-pass filter. **(C)** Reconstruction low-pass filter. **(D)** Reconstruction high-pass filter.

developed a wavelet variance-based method for modeling sensor error. Inspired by those three papers, we propose the extraction of triple features based on wavelet subbands.

For each subband $s\left(i,j\right)$ in the $(3L+1)$ subbands, we extract the energy feature f_1 as

$$f_1\left(s\right)=\sum_{i}^{I(s)}\sum_{j}^{J(s)}s^2\left(i,j\right), \tag{13}$$

where $\left(i,j\right)$ is the row and column index of subband s, and $I\left(s\right)$ is the height, $J\left(s\right)$ is the width.

The variance feature $f_2\left(s\right)$ is calculated as

$$f_2\left(s\right)=\frac{1}{I\left(s\right)\times J\left(s\right)-1}\sum_{i}^{I(s)}\sum_{j}^{J(s)}\left[s\left(i,j\right)-\mu\left(s\right)\right]^2, \tag{14}$$

where $\mu\left(s\right)$ is the mean value of subband s

$$\mu\left(s\right)=\frac{1}{I\left(s\right)\times J\left(s\right)}\sum_{i}^{I(s)}\sum_{j}^{J(s)}s\left(i,j\right). \tag{15}$$

The entropy feature $f_3\left(s\right)$ is computed in a more complicated way. We assume the coefficients in subband s is a discrete random variable $\mathbb{S}$ with quantization values $(s_1, s_2, \ldots, s_H)$, then we calculate the corresponding probability mass function $p=\left\{p_h\right\}$.

$$p_h=\Pr\left(\mathbb{S}=s_h\right), \tag{16}$$

where Pr is the probability function.

Finally, the entropy $f_3\left(s\right)$ is defined as:

$$f_3\left(s\right)=-\sum_{h=1}^{H}p_h\times\log p_h. \tag{17}$$

In all, the triple features of all subbands are concatenated to form a feature vector $F=\left\{\left(f_1,f_2,f_3\right)\right\}$.

$$F=\begin{bmatrix}\left(f_1,f_2,f_3\right)\left(K_L\right) & \left(f_1,f_2,f_3\right)\left(J_L\right) & \left(f_1,f_2,f_3\right)\left(W_L\right) & \left(f_1,f_2,f_3\right)\left(F_L\right)\\ & \left(f_1,f_2,f_3\right)\left(J_{L-1}\right) & \left(f_1,f_2,f_3\right)\left(W_{L-1}\right) & \left(f_1,f_2,f_3\right)\left(F_{L-1}\right)\\ & \cdots & \cdots & \cdots\\ & \left(f_1,f_2,f_3\right)\left(J_1\right) & \left(f_1,f_2,f_3\right)\left(W_1\right) & \left(f_1,f_2,f_3\right)\left(F_1\right)\end{bmatrix}. \tag{18}$$

The number of features in F is $N_F=3\times(3L+1)=9L+3$. The pseudocode of BWFs is depicted in Algorithm 2.

Algorithm 2 Pseudocode of calculating BWFs

Input: $(3L+1)$ Subbands:
$(K_L, J_L, W_L, F_L, J_{L-1}, W_{L-1}, F_{L-1}, \ldots, J_1, W_1, F_1)$
for $m=1:3L+1$
Select m-th subband s
Calculate energy $f_1\left(s\right)$. See Eq. (13).
Calculate mean value. See Eq. (15).
Calculate variance $f_2\left(s\right)$. See Eq. (14).
Calculate probability mass function. See Eq. (16).
Calculate entropy $f_3\left(s\right)$. See Eq. (17).
end
Output: The concatenated feature vector F with N_F number of features. See Eq. (18).

One-Hidden-Layer Feedforward Neural Network

The N_F features are sent to a feedforward neural network (FNN), a type of artificial neural network wherein the connections do not generate a loop (Vanchurin, 2021). Within FNN, the information moves along one direction—forward—from input layers through hidden layers to output layer (Valizadeh et al., 2021). The advantage of FNN is that it does not need any

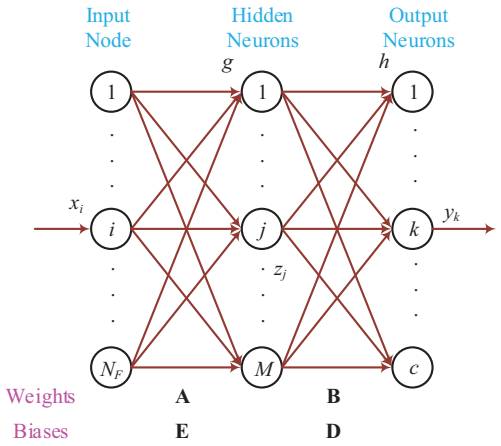

FIGURE 4 | Diagram of one-hidden-layer FNN.

information related to probability distribution and the *a priori* probabilities of different classes (Koo and Kim, 2021).

One-hidden-layer FNN is chosen in this study due to the statements of the "universal approximation theory." **Figure 4** illustrates its diagram. Suppose (x, t) denotes a training sample, where

$$x = \left[x_1, x_2, \ldots, x_i, \ldots, x_{N_F}\right]^T, \tag{19}$$

denotes the input vector with N_F-dimension, and i is the neuron index at the input layer.

Let t be the target label

$$t = [t_1, t_2, \ldots, t_k, \ldots, t_c]^T, \tag{20}$$

where c is the number of categories and k is the neuron index at the output layer.

Let n be the sample index and N the number of total training samples; we can describe the training sample (x, t) as $\{x(n), t(n) | n = 1, \ldots, N\}$. The training of one-hidden-layer FNN (Ng et al., 2021) is an optimization problem of minimizing the sum of mean-squared error E between the target t and realistic output y as

$$E = \sum_{n=1}^{N} \sum_{k=1}^{c} \left[y_k(n) - t_k(n)\right]^2. \tag{21}$$

Zoom in on the details of the above formula. Assume g is the activation function in the hidden layer, h the activation function in the output layer, $\mathbf{A} = \left\{a\left(i, j\right)\right\}, i = 1, \ldots, N_F, j = 1, \ldots, M$ and $\mathbf{E} = \left\{e\left(j\right)\right\}, j = 1, \ldots, M$ the weights and biases of neurons that connect input layer with hidden layer, and $\mathbf{B} = \left\{b\left(j, k\right)\right\}, j = 1, \ldots, M, k = 1, \ldots c$, and $\mathbf{D} = \left\{d\left(k\right)\right\}, k = 1, \ldots, c$ the weights and biases of neurons that connect hidden layer to output layer. We can calculate the output y_k as

$$y_k(n) = h\left[\sum_{j=1}^{M} b\left(j, k\right) z_j(n) + d\left(k\right)\right], \tag{22}$$

where $z_j(n), j = 1, \ldots, M$ represents the output of j-th neuron in the hidden layer (Kiran and Naik, 2021). Its definition is

$$z_j(n) = g\left[\sum_{i=1}^{N_F} a\left(i, j\right) x_i(n) + e\left(j\right)\right]. \tag{23}$$

Encoding Scheme

The parameter-training procedure can be regarded as an optimization function by which we seek the optimal parameter value for the set (**A**, **B**, **E**, **D**). Suppose an i-th candidate solution is written as a chromosome θ_i which can be decomposed into four gene blocks:

$$\theta_i = [\mathbf{A}_i, \mathbf{B}_i, \mathbf{E}_i, \mathbf{D}_i]. \tag{24}$$

In the preceding section, the variables **A** and **B** are in matrix forms. In this section, we rephrase them in vector forms to a better description of the encoding scheme. Hence, the four variables' range is shown as:

$$\begin{cases} A = \left\{a\left(k\right)\right\}, k = 1, \ldots, N_F \times M \\ B = \left\{b\left(k\right)\right\}, k = 1, \ldots, M \times c \\ E = \left\{e\left(k\right)\right\}, k = 1, \ldots, M \\ D = \left\{d\left(k\right)\right\}, k = 1, \ldots, c \end{cases}, \tag{25}$$

where k is the gene index in each chromosome. Thus, it means the length of the chromosome is

$$l = N_F \times M + M \times c + M + c. \tag{26}$$

Figure 5 gives an example, suppose $N_F = 30$, $M = 10$, and $c = 2$, we can see now each chromosome contains $l = 332$ genes, namely, size $(\theta_i) = 332$. The whole population $\overrightarrow{\theta}$ is written as

$$\overrightarrow{\theta} = \left(\theta_1, \theta_2, \ldots \theta_i, \ldots, \theta_p\right), \tag{27}$$

where $i = 1, \ldots, p$, and p stands for the size of the population, within which each chromosome corresponds to a candidate solution. The whole population will be updated iteratively to approximate the optimal value via an optimization algorithm described in the following section.

Fitness-Scaled Adaptive Genetic Algorithm

The FAGA is introduced from Li (2017), which combines standard genetic algorithm (SGA), adaptive mechanism (So et al., 2021), and fitness scaling methods. **Figure 6** shows the flowchart of SGA, of which the basic principles can be found in references (Leow et al., 2021).

After analyzing the SGA, two issues are observed: (i) The fixed crossover and mutation rates make SGA susceptible to get stuck at local minima. (ii) the wide span favors the best individual, driving them reproduce too rapidly and thus take over the whole population (Li, 2017).

To overcome the first issue, an adaptive crossover rate (p_c) and an adaptive mutation rate (p_m) are adopted to heighten the convergence performance of SGA. Those two adaptive rates $\left(p_c, p_m\right)$ aim at the trade-off between exploration and

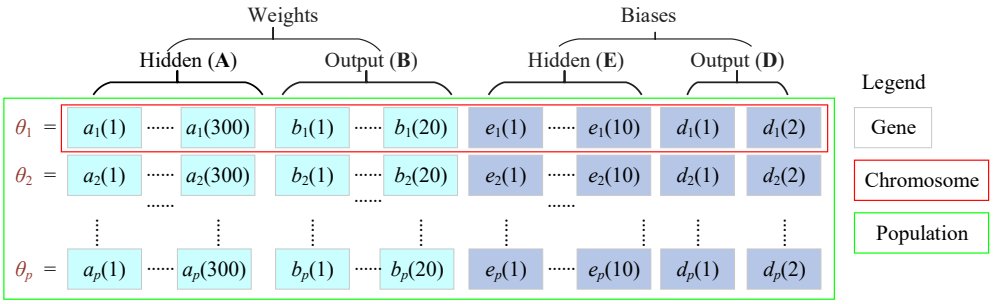

FIGURE 5 | Encoding scheme.

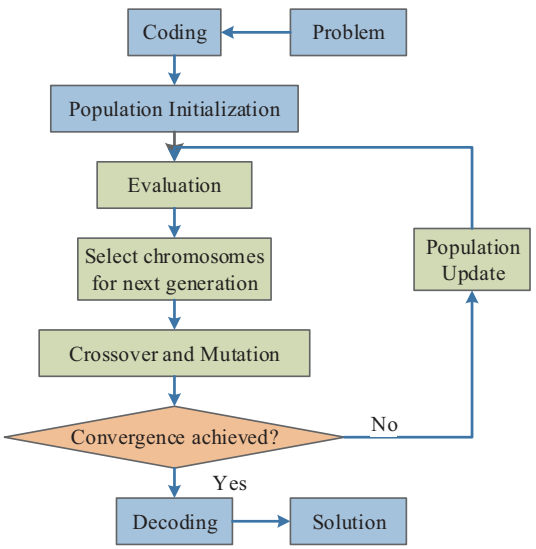

FIGURE 6 | Flowchart of SGA.

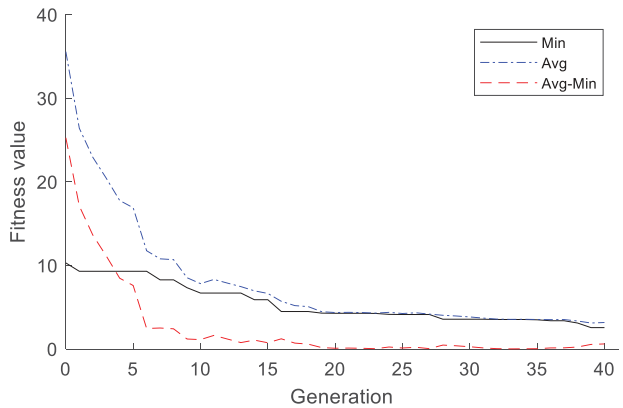

FIGURE 7 | The iteration of SGA of a convergent case (Note $f_a - f_{\min}$ decreases).

exploitation. p_c controls the rate at which candidates are to crossover. The higher the p_c is, the quicker the new solution is introduced into the population. Typical values of p_c are in the range of [0.5, 1.0]. p_m controls the mutation rate. Although it is a secondary operation, the choice of p_m is essential to the performance of SGA. Larger value of p_m transforms the GA into a purely random search algorithm (Srinivas and Patnaik, 1994). Typical values of p_m are in the range of [0.005, 0.05].

Suppose we are coping with a minimum problem, and let f' represent the better of the fitness values of the two candidates to crossover, f_a the average fitness value, and $f_{\min}$ the minimum fitness value. Then, the definition of p_c is

$$p_c = \begin{cases} k_c \times \frac{f' - f_{\min}}{f_a - f_{\min}} & f' \leq f_a \\ k_c & \text{otherwise} \end{cases}, \tag{28}$$

and the definition of p_m is

$$p_m = \begin{cases} k_m \times \frac{f - f_{\min}}{f_a - f_{\min}} & f \leq f_a \\ k_m & \text{otherwise} \end{cases}, \tag{29}$$

where $0 = k_c, k_m = 1$.

In **Figure 7**, the blue dash-dotted curve corresponds to the average fitness value f_a, the black solid curve corresponds to the minimum fitness value $f_{\min}$, and finally the red dashed curve corresponds to $f_a - f_{\min}$. **Figure 7** illustrates that the diversity (measured by $f_a - f_{\min}$) decreases when SGA converges; thus, p_c and p_m should increase at early stage so as to increase the exploration capability (Fasel et al., 2021). Nevertheless, extremely large values will counteract the algorithm converging to the global optimal (Le Guisquet and Amabili, 2021). Therefore, p_c and p_m should shrink for the good chromosome, since the better chromosome (measured by $f - f_{\min}$) should be preserved.

To overcome the second issue, power-rank fitness scaling (Li, 2017) is harnessed. There are many other fitness scaling methods, such as shift linear scaling (To and Liew, 2014), top scaling, etc. In our study, we find power-rank fitness scaling works the best among all fitness scaling methods. Suppose the original fitness function of i-th chromosome is f_i, and the fitness-scaled result is g_i. We first calculate the rank sequence $\{r_i\}$ of all chromosomes as

$$r_i = \text{rank}\left(f_i, \{f_i\}\right), \tag{30}$$

where the "rank" stands for the order of the chromosome f_i in the vector $\{f_i\}$.

Second, the fitness-scaled g_i is calculated via taking the power function and followed by normalization function:

$$g_i = \frac{r_i^k}{\sum r_i^k}, \tag{31}$$

where r^k denotes the value of r is raised to the power of k.

Implementation and Multiple-Way Data Augmentation

T-fold cross-validation is employed to run our algorithm that is named BWF-FAGA. The whole dataset (676 MS images and 681 HC images) is divided into T folds. At t-th trial, $1 \leq t \leq T$, the t-th fold is chosen as the test, set and the rest $T-1$ folds: $[1, \ldots, t-1, t+1, \ldots, T]$ are picked up as the training set. **Figure 8** shows the diagram of T-fold cross validation. Here we let $T = 10$, viz., a 10-fold cross validation is run in the experiments. Besides, we run this 10-fold cross-validation R times.

Multiple sclerosis images are hard to collect, which leads to the small-size dataset problem that may bring in the overfitting to the classifier. To solve the small-size dataset problem, Zhou (2021) presented an 18-way data augmentation (DA), as displayed in **Figure 9**. The differences between ordinary DA and MDA are two points: (i) an MDA combines different single ordinary DA methods; (ii) MDA is a modular design so the users are easygoing to add or remove any DA methods from an MDA framework.

Suppose there is a raw training image $r(w)$, in which w represents the image index in the training set. First, O_1 different DA methods displayed in are applied to $r(w)$. Let $\mathbf{Z}_o, o = 1, \ldots, O_1$ be each DA operation, we get O_1 augmented datasets on raw image $r(w)$ as:

$$\mathbf{Z}_o[r(w)], o = 1, \ldots, O_1. \tag{32}$$

Let O_2 mean the size of generated new images for each DA method, thus,

$$|\mathbf{Z}_o[r(w)]| = O_2. \tag{33}$$

Second, HMI is generated by:

$$r^{(h)}(w) = \beta_1[r(w)], \tag{34}$$

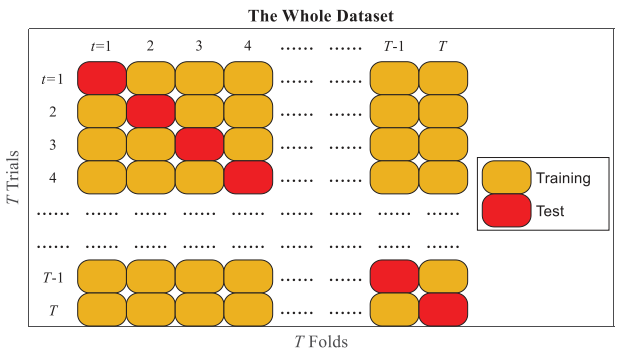

FIGURE 8 | Diagram of T-fold cross-validation.

where β_1 stands for horizontal mirror function.

Third, all O_1 different DA methods run on the HMI $r^{(h)}(w)$, and produce O_1 new datasets as.

$$\begin{cases} \mathbf{Z}_o\left[r^{(h)}(w)\right], o = 1, \ldots, O_1 \\ \left|\mathbf{Z}_o\left[r^{(h)}(w)\right]\right| = O_2, o = 1, \ldots, O_1 \end{cases}. \tag{35}$$

Fourth, the raw image $r(w)$, the HMI $r^{(h)}(w)$, all O_1-way DA results $\mathbf{Z}_o[r(w)]$ of the raw image, and all O_1-way DA results $\mathbf{Z}_o\left[r^{(h)}(w)\right]$ of HMI are combined. The final generated dataset from $r(w)$ is defined as $\mathbf{P}(w)$:

$$r(w) \mapsto \mathbf{P}(w) = \beta_2 \left\{ \begin{array}{cc} r(w) & r^{(h)}(w) \\ \underbrace{\mathbf{Z}_1[r(w)]}_{O_2} & \underbrace{\mathbf{Z}_1\left[r^{(h)}(w)\right]}_{O_2} \\ \cdots & \cdots \\ \underbrace{\mathbf{Z}_{O_1}[r(w)]}_{O_2} & \underbrace{\mathbf{Z}_{O_1}\left[r^{(h)}(w)\right]}_{O_2} \end{array} \right\}, \tag{36}$$

where β_2 stands for the combination function.

Let augmentation factor be O_3, which means the number of images in $\mathbf{P}(w)$, which is calculated as

$$O_3 = \frac{|\mathbf{P}(w)|}{|r(w)|} = \frac{(1 + O_1 \times O_2) \times 2}{1} = 2 \times O_1 \times O_2 + 2. \tag{37}$$

Algorithm 3 recapitulates the pseudocode of 18-way DA. We set $O_1 = 9$ to yield an 18-way DA.

Algorithm 3 Pseudocode of 18-way DA on w-th training image

Input	Import a raw preprocessed w-th training image $r(w)$.
Step I	O_1 geometric or photometric or noise-injection DA transforms $\mathbf{Z}_o$ are utilized on $r(w)$. We obtain $\mathbf{Z}_o[r(w)], o = 1, \ldots, O_1$. See Eq. (32). Each enhanced dataset contains O_2 new images. See Eq. (33).
Step II	An HMI is produced as $r^{(h)}(w) = \beta_1[r(w)]$. See Eq. (34).
Step III	O_1-way data augmentation methods run on $r^{(h)}(w)$, we obtain $\mathbf{Z}_o\left[r^{(h)}(w)\right], o = 1, \cdots, O_1$. See Eq. (35).
Step IV	$r(w)$, $r^{(h)}(w)$, $\mathbf{Z}_o[r(w)], o = 1, \ldots, O_1$, and $\mathbf{Z}_o\left[r^{(h)}(w)\right], o = 1, \cdots, O_1$ are combined via β_2. See Eq. (36).
Output	A new dataset $\mathbf{P}(w)$ is generated. Its image number is $O_3 = 2 \times O_1 \times O_2 + 2$. See Eq. (37).

Measures

The T-fold cross-validation run R times. Within every run, the dataset split is reset randomly. The confusion matrix of each run is recorded. Let G denotes the confusion matrix

$$G = \begin{bmatrix} G(1,1) & G(1,2) \\ G(2,1) & G(2,2) \end{bmatrix}. \tag{38}$$

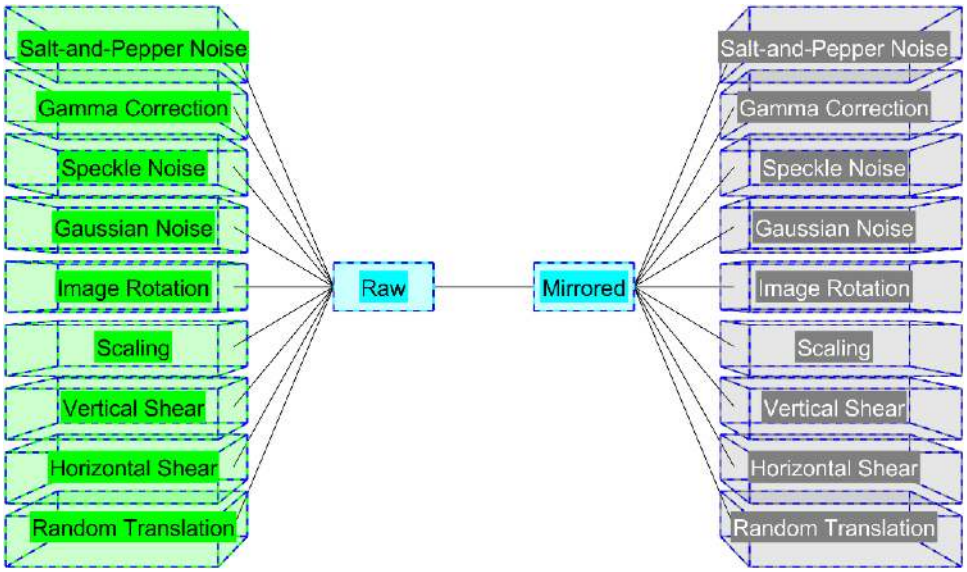

FIGURE 9 | Diagram of multiple-way data augmentation.

TABLE 3 | Definitions in the confusion matrix.

Abbreviation	Symbol	Meaning
P		Positive class: Multiple sclerosis
N		Negative class: Healthy control
TP	$G(1,1)$	MS is correctly classified into MS.
FN	$G(1,2)$	MS is wrongly classified into HC.
FP	$G(2,1)$	HC is wrongly classified into MS.
TN	$G(2,2)$	HC is correctly classified into HC.

Suppose MS is the positive class, and HC is the negative class. The true positive (TP), false negative, false positive, and true negative measures are defined in **Table 3**.

Sensitivity, specificity, and precision are defined below

$$\begin{cases} \text{Sen} = \frac{G(1,1)}{G(1,1)+G(1,2)} \\ \text{Spc} = \frac{G(2,2)}{G(2,1)+G(2,2)} \\ \text{Prc} = \frac{G(1,1)}{G(1,1)+G(2,1)} \end{cases}. \tag{39}$$

Accuracy is defined as:

$$\text{Acc} = \frac{G(1,1)+G(2,2)}{G(1,1)+G(1,2)+G(2,1)+G(2,2)}. \tag{40}$$

F1 score considers both the precision and the sensitivity. It is the harmonic mean of the previous two measures: precision and sensitivity (van der Goot and van Noord, 2017; Duan et al., 2021). F1 score is defined as

$$\text{F1} = \left(\frac{\text{Sen}^{-1}+\text{Prc}^{-1}}{2}\right)^{-1} = \frac{2\times G(1,1)}{2\times G(1,1)+G(1,2)+G(2,1)}. \tag{41}$$

TABLE 4 | Parameter setting.

Parameter	Value
L	3
N_r	4
N_d	4
N_F	30
M	10
c	2
I	332
p	20
k_c	1.0
k_m	0.5
T	10
R	10
O_1	9
O_2	30
O_3	542

Two other measures: Matthews correlation coefficient (MCC) and Fowlkes–Mallows index (FMI) are defined as:

$$\text{MCC} = \frac{G(1,1)\times G(2,2) - G(2,1)\times G(1,2)}{\sqrt{\begin{array}{l}[G(1,1)+G(2,1)]\times[G(1,1)+G(1,2)] \\ \times[G(2,2)+G(2,1)]\times[G(2,2)+G(1,2)]\end{array}}}, \tag{42}$$

$$\text{FMI} = \sqrt{\frac{G(1,1)}{G(1,1)+G(2,1)}\times\frac{G(1,1)}{G(1,1)+G(1,2)}}. \tag{43}$$

The above seven measures are calculated for each run. Then, the mean and standard deviation (MSD) is computed across all the R runs. Furthermore, both the receiver operating characteristic (ROC) curve and the area under the curve (AUC;

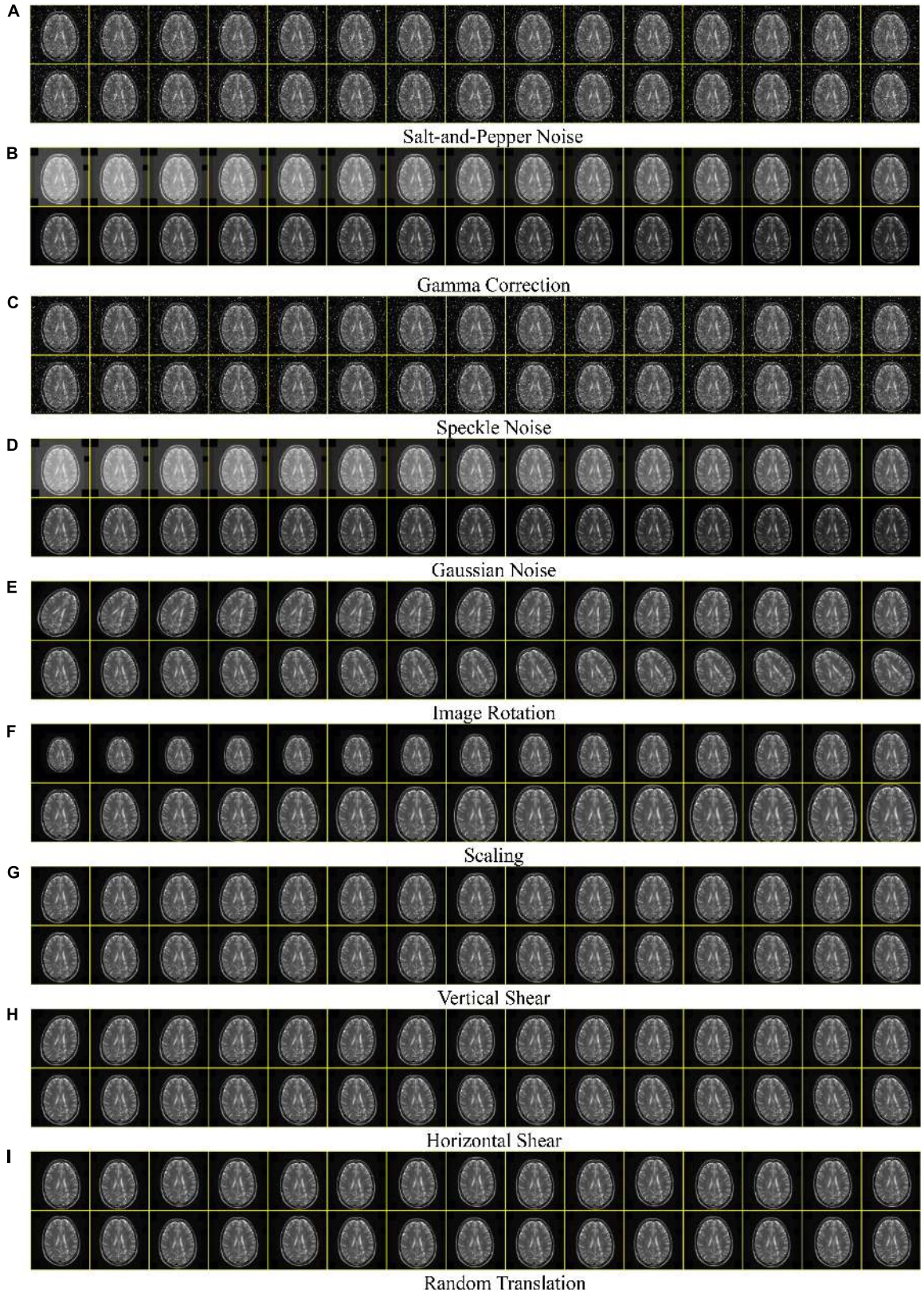

FIGURE 10 | Results of Multiple-way DA. **(A)** Salt-and-Pepper Noise. **(B)** Gamma Correction. **(C)** Speckle Noise. **(D)** Gaussian Noise. **(E)** Image Rotation. **(F)** Scaling. **(G)** Vertical Shear. **(H)** Horizontal Shear. **(I)** Random Translation.

TABLE 5 | Results of 10 runs of proposed BWF-FAGA algorithm.

Run	Sen	Spc	Prc	Acc	F1	MCC	FMI
1	97.93	97.65	97.64	97.79	97.78	95.58	97.78
2	96.45	96.04	96.02	96.24	96.24	92.48	96.24
3	96.75	97.06	97.03	96.90	96.89	93.81	96.89
4	98.52	98.38	98.38	98.45	98.45	96.91	98.45
5	99.11	98.97	98.97	99.04	99.04	98.08	99.04
6	98.52	98.09	98.09	98.31	98.30	96.61	98.30
7	98.37	98.24	98.23	98.31	98.30	96.61	98.30
8	98.22	97.94	97.94	98.08	98.08	96.17	98.08
9	97.04	96.62	96.61	96.83	96.83	93.66	96.83
10	99.11	98.83	98.82	98.97	98.97	97.94	98.97
MSD	98.00 ± 0.95	97.78 ± 0.95	97.77 ± 0.95	97.89 ± 0.94	97.89 ± 0.95	95.79 ± 1.89	97.89 ± 0.95

Chappuis et al., 2021) indicators are harnessed to provide a graphical plot of measuring our proposed algorithm. ROC and AUC are calculated via the following way: (i) ROC curve is produced by plotting the TP rate against the false-positive rate at various threshold levels. (ii) AUC is calculated by measuring the entire 2D area underneath the ROC curve from (0, 0) to (1, 1).

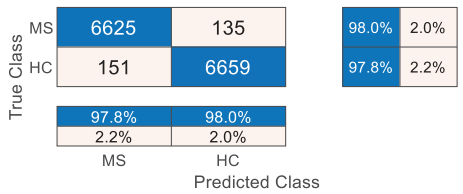

FIGURE 11 | Confusion matrix of BWF-FAGA algorithm.

EXPERIMENTS, RESULTS, AND DISCUSSION

Parameter Setting

The parameters in this study are itemized in **Table 4**. The maximum decomposition level is chosen as $L = 3$. The reconstruction order of BWT equals the decomposition order of BWT as 4. The number of features extracted is $N_F = 30$. The number of hidden neurons is set to $M = 10$, and the number of output neurons is set to $c = 2$. The length of chromosome is $l = 332$. The size of population is $p = 20$. Two parameters in FAGA are set as $k_c = 1.0$ and $k_m = 0.5$. A 10-fold cross validation is performed, which is further repeated ten runs. We use $O_1 = 9$ different DA methods on both raw image and HMI; thus, we harness an 18-way DA. Each DA generates $O_2 = 30$ images. The augmentation factor is $O_3 = 542$.

Results of Multiple-Way Data Augmentation

Use the MS image in **Figure 1C** as an example; **Figure 10** shows the newly generated images. The HMI and its corresponding DA results are not displayed due to the page limit. In **Figures 10A–I**, we can see the MDA indeed help increase the variety of the training image set.

Results of Proposed Method

Table 5 shows the ten-run results using the parameter shown in **Table 4**. The final MSDs of 10 runs are shown at the last row of **Table 5**, where we can observe this BWF-FAGA achieves a sensitivity of 98.00 ± 0.95%, a specificity of 97.78 ± 0.95%, a precision of 97.77 ± 0.95%, an accuracy of 97.89 ± 0.94%, an F1 score of 97.89 ± 0.95%, an MCC of 95.79 ± 1.89%, and an FMI of 97.89 ± 0.95%.

If we combine the confusion matrix across ten runs, we get the following chart shown in **Figure 11**. It shows that among ten runs, 135 MS are wrongly classified into HC, while 151 HC are wrongly classified into MS. **Figure 12C** shows the ROC curve, from which we can observe the AUC is 0.9876.

Effect of Decomposition Level

We now change the decomposition level $L = 1, 2, 3, 4$, and check how the performance changes with L. The resultant results are displayed in **Table 6**, which shows the $L = 3$ setting achieves the best performance among all four settings $L = 1, 2, 3, 4$. The ROC curves and AUCs corresponding to L with values of 1, 2, 3, and 4 are shown in **Figure 12**, which indicates $L = 3$ yields the maximum AUC value of 0.9876.

Comparison With State-of-the-Art Methods

The proposed BWF-FAGA algorithm is compared with 10 state-of-the-art MS recognition approaches, including MAMFM (Loizou et al., 2011), 2D-DWT (Nayak et al., 2016), BWT-LR (Zhan and Chen, 2016), L3-HWT (Lopez, 2017), GLCM-BBO (Zhou and Shen, 2018), DDP (Yahia et al., 2018), WE-AGA (Han and Hou, 2019), HMI-PSO (Han and Hou, 2020), 5L-CNN (Tang, 2021), and FCNN (Salem et al., 2020). The last is one DL method. The comparison results on the same dataset using the same 10 runs of 10-fold CV are shown in **Tables 7**, **8**.

For ease of visual comparison, **Figure 13** draws the 3D bar plot of performances of eleven methods altogether. Since MCC has lower values than other measures, we move MCC to the leftmost and rank all the algorithms in terms of MCC. The summer

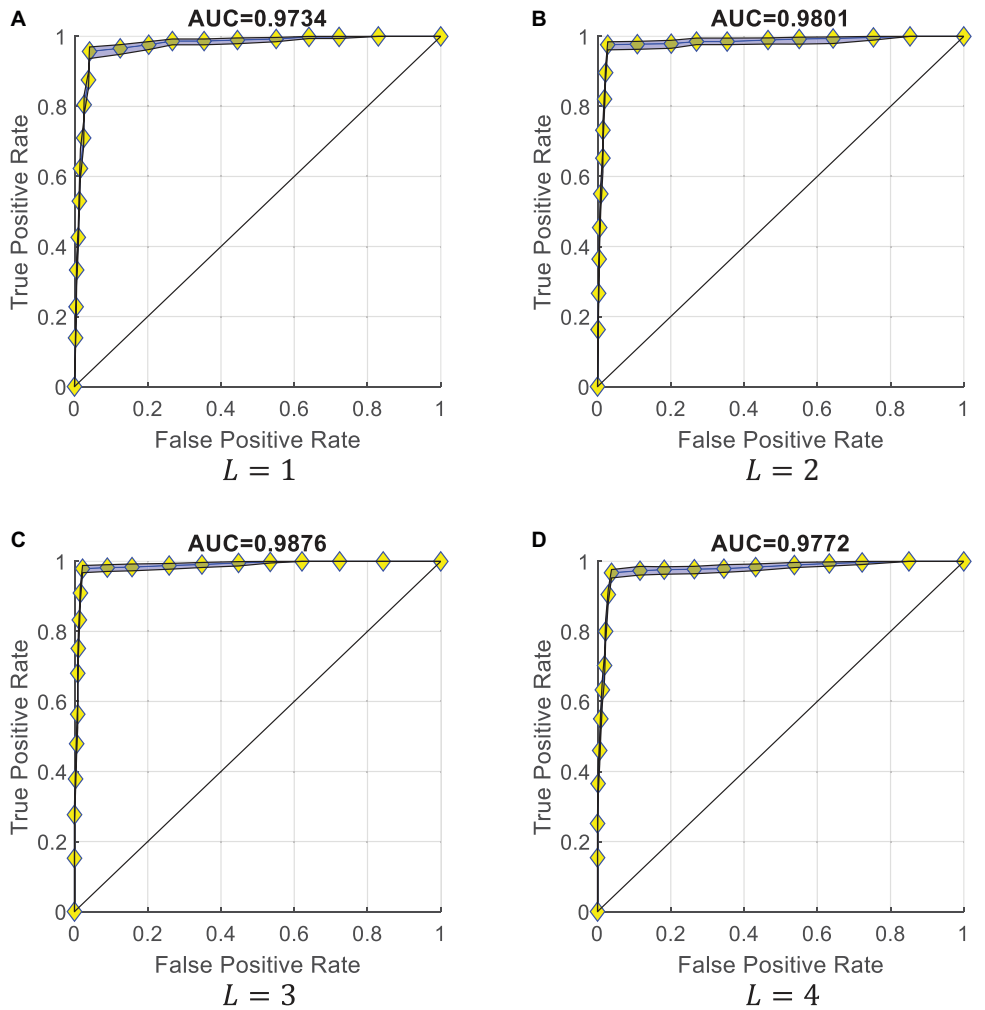

FIGURE 12 | Comparisons of different decomposition levels in ROC and AUC. **(A)** $L = 1$. **(B)** $L = 2$. **(C)** $L = 3$. **(D)** $L = 4$.

TABLE 6 | Comparison of different decomposition levels.

Decomposition level	Sen	Spc	Prc	Acc	F1	MCC	FMI
$L = 1$	95.78 ± 1.26	95.95 ± 1.25	95.91 ± 1.25	95.87 ± 1.18	95.85 ± 1.18	91.73 ± 2.35	95.85 ± 1.18
$L = 2$	97.69 ± 1.01	97.21 ± 1.26	97.21 ± 1.24	97.45 ± 1.07	97.45 ± 1.07	94.90 ± 2.15	97.45 ± 1.07
$L = 3$ (Ours)	**98.00 ± 0.95**	**97.78 ± 0.95**	**97.77 ± 0.95**	**97.89 ± 0.94**	**97.89 ± 0.95**	**95.79 ± 1.89**	**97.89 ± 0.95**
$L = 4$	96.76 ± 1.07	96.15 ± 1.37	96.16 ± 1.33	96.46 ± 1.08	96.46 ± 1.08	92.92 ± 2.16	96.46 ± 1.07

Bold means the best.

pseudocolor is used where green stands for the lowest value and yellow for the greatest value.

From **Figure 13**, we can observe the proposed BWF-FAGA method yields the best performances among all the other methods, including eight AI-based methods and two DL-based methods. The reasons are as follows: (a) We use bior4.4 wavelet to extract multiscale coefficients. (b) Three BWFs are proposed to extract from wavelet coefficients. (c) FAGA is harnessed as the optimization algorithm. (d) MDA is used on the training set.

There are some other advanced technologies that we may use in the future: (i) Several novel mechanisms can help the classification performances, e.g., generative adversarial networks (Jin et al., 2018); (ii) Fusion techniques: 3D CNN fusion (Jin et al., 2019) and label fusion (Huo et al., 2017). (iii) Segmentation techniques: 3D whole-brain segmentation (Huo et al., 2019) and multi-atlas brain segmentation (Huo et al., 2016). (iv) Curve extraction technique, such as: curve skeletonization (Jin et al., 2016). Furthermore, some feature extraction methods (Lu et al., 2016, 2018) may help.

TABLE 7 | Comparison with AI-based MS recognition methods (unit: %).

Method	Sen	Spc	Prc	Acc	F1	MCC	FMI
MAMFM (Loizou et al., 2011)	94.25 ± 0.58	93.51 ± 1.43	93.53 ± 1.34	93.88 ± 0.87	93.88 ± 0.84	87.76 ± 1.73	93.88 ± 0.83
2D-DWT (Nayak et al., 2016)	96.12 ± 0.82	96.27 ± 0.76	96.24 ± 0.76	96.20 ± 0.75	96.18 ± 0.75	92.40 ± 1.50	96.18 ± 0.75
BWT-LR (Zhan and Chen, 2016)	97.01 ± 0.91	97.52 ± 0.99	97.49 ± 0.99	97.27 ± 0.90	97.25 ± 0.90	94.53 ± 1.79	97.25 ± 0.90
L3-HWT (Lopez, 2017)	89.63 ± 1.75	90.48 ± 1.45	90.34 ± 1.43	90.06 ± 1.44	89.98 ± 1.47	80.13 ± 2.87	89.98 ± 1.47
GLCM-BBO (Zhou and Shen, 2018)	92.97 ± 0.58	92.82 ± 0.84	92.78 ± 0.81	92.90 ± 0.63	92.88 ± 0.63	85.79 ± 1.27	92.88 ± 0.63
DDP (Yahia et al., 2018)	95.19 ± 1.76	95.43 ± 1.45	95.39 ± 1.45	95.31 ± 1.48	95.29 ± 1.49	90.63 ± 2.96	95.29 ± 1.49
WE-AGA (Han and Hou, 2019)	91.91 ± 1.24	91.98 ± 1.36	91.97 ± 1.32	91.95 ± 0.19	91.92 ± 1.20	83.89 ± 2.41	91.92 ± 1.19
HMI-PSO (Han and Hou, 2020)	91.67 ± 1.41	91.73 ± 0.77	91.70 ± 0.78	91.70 ± 0.97	91.67 ± 1.00	83.40 ± 1.98	91.67 ± 0.99
BWF-FAGA (Ours)	98.00 ± 0.95	97.78 ± 0.95	97.77 ± 0.95	97.89 ± 0.94	97.89 ± 0.95	95.79 ± 1.89	97.89 ± 0.95

TABLE 8 | Comparison with DL-based MS recognition methods (unit: %).

Method	Sen	Spc	Prc	Acc	F1	MCC	FMI
5L-CNN (Tang, 2021)	94.96 ± 1.43	94.82 ± 1.67	94.85 ± 1.62	94.89 ± 1.37	94.87 ± 1.37	89.83 ± 2.74	94.87 ± 1.37
FCNN (Salem et al., 2020)	97.56 ± 1.08	97.33 ± 1.33	97.32 ± 1.32	97.44 ± 1.16	97.44 ± 1.15	94.89 ± 2.31	97.44 ± 1.15
BWF-FAGA (Ours)	98.00 ± 0.95	97.78 ± 0.95	97.77 ± 0.95	97.89 ± 0.94	97.89 ± 0.95	95.79 ± 1.89	97.89 ± 0.95

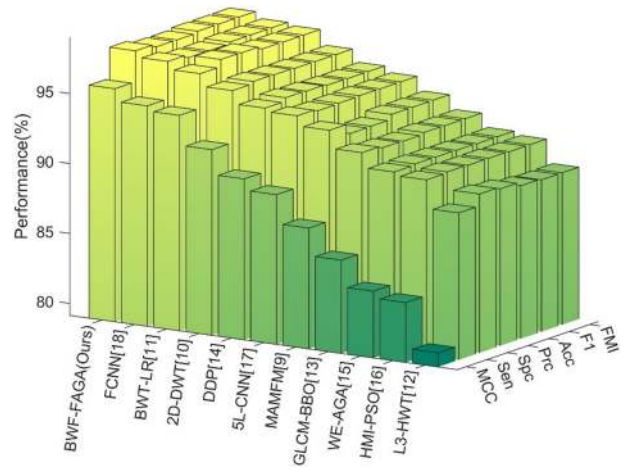

FIGURE 13 | 3D bar plot of method comparison (ranked in the descending order of MCC).

CONCLUSION

A novel BWF-FAGA method is proposed. First, the bior4.4 wavelet is used to extract multiscale coefficients. Second, three types of BWFs are proposed and calculated. Third, FAGA is harnessed as the optimization algorithm. Fourth, MDA is utilized on the training set under the setting of 10 runs of 10-fold cross-validation.

The results express that the proposed BWF-FAGA method achieves better performances than 10 state-of-the-art MS recognition methods, including 8 AI-based methods and 2 DL-based methods. Our method achieves a sensitivity of 98.00 ± 0.95%, a specificity of 97.78 ± 0.95%, and an accuracy of 97.89 ± 0.94%.

The shortcomings of our BWF-FAGA method are threefolds: (i) It still needs to extract features manually. (ii) It does not go through strict clinical validation. (iii) Our dataset is small. To overcome the shortcomings above, we shall try to combine biorthogonal wavelets intrinsically with deep neural networks. On the other hand, we shall release our method to the online cloud computing environment and invite clinicians in the hospital to test its effectiveness. Finally, we shall try to collect more MS images.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

S-HW: methodology, software, investigation, writing – original draft, writing – review and editing, visualization, and funding acquisition. XJ: software, validation, data curation, writing – review and editing, and project administration. Y-DZ: conceptualization, methodology, investigation, resources, writing – review and editing, supervision, and funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This manuscript is partially supported by Medical Research Council Confidence in Concept Award, United Kingdom (MC_PC_17171); Royal Society International Exchanges

Cost Share Award, United Kingdom (RP202G0230); British Heart Foundation Accelerator Award, United Kingdom (AA/18/3/34220); Hope Foundation for Cancer Research, United Kingdom (RM60G0680); Global Challenges Research Fund (GCRF), United Kingdom (P202PF11); and Sino-UK Industrial Fund, United Kingdom (RP202G0289).

REFERENCES

Abdullah, S., Ismail, S. M., Hasan, M. K., and Shivakumara, P. (2021). Novel Adaptive Binarization Method for Degraded Document Images. *CMC* 67, 3815–3832. doi: 10.32604/cmc.2021.014610

Allum, J. H. J., Rust, H. M., Lutz, N., Schouenborg, C., Fischer-Barnicol, B., Haller, V., et al. (2021). Characteristics of improvements in balance control using vibrotactile biofeedback of trunk sway for multiple sclerosis patients. *J. Neurol. Sci.* 425:11. doi: 10.1016/j.jns.2021.117432

Blechinger, S., Ehler, J., Bsteh, G., Winkelmann, A., Leutmezer, F., Meister, S., et al. (2021). Therapeutic plasma exchange in steroid-refractory multiple sclerosis relapses. A retrospective two-center study. *Ther. Adv. Neurol. Dis.* 14:1756286420975642. doi: 10.1177/1756286420975642

Chappuis, V. N., Deham, H., Cottet, P., Gartner, B. A., Sarasin, F. P., Niquille, M., et al. (2021). Emergency physician's dispatch by a paramedic-staffed emergency medical communication centre: sensitivity, specificity and search for a reference standard. *Scand. J. Trauma Resusc. Emerg. Med.* 29:31. doi: 10.1186/s13049-021-00844-y

de Sena, A. P. C., de Freitas, I. S., Lima, A. C., and Sobrinho, C. A. N. (2021). Fuzzy diagnostics for gearbox failures based on induction motor current and wavelet entropy. *J. Brazil. Soc. Mech. Sci Eng.* 43:265. doi: 10.1007/s40430-021-02964-z

Duan, G. H., Zhang, S. F., Lu, M. Z., Okinda, C., Shen, M. X., and Norton, T. (2021). Short-term feeding behaviour sound classification method for sheep using LSTM networks. *Int. J. Agric. Biol. Eng.* 14, 43–54. doi: 10.25165/j.ijabe.20211402.6081

eHealth. *MRI Lesion Segmentation in Multiple Sclerosis Database*. in eHealth laboratory, University of Cyprus. Available Online from: http://www.medinfo.cs.ucy.ac.cy/index.php/facilities/32-software/218-datasets (accesseed January 15, 2021).

Fasel, U., Tiso, P., Keidel, D., and Ermanni, P. (2021). Concurrent Design and Flight Mission Optimization of Morphing Airborne Wind Energy Wings. *AIAA J.* 59, 1254–1268. doi: 10.2514/1.J059621

Ganesan, T., and Rajarajeswari, P. (2021). Hybrid Genetic Algorithm With Haar Wavelet for Maximum Target Coverage Node Deployment in Wireless Sensor Networks. *J. Cases Inform. Technol.* 23, 78–95. doi: 10.4018/JCIT.20210701.oa6

Gil-Casas, A., Pinero-Llorens, D. P., and Molina-Martin, A. (2021). Binocular Vision in Patients with Multiple Sclerosis. *Clin. Optomet.* 13, 39–49. doi: 10.2147/OPTO.S286862

Han, B., and Michelle, M. (2021). Wavelets on intervals derived from arbitrary compactly supported biorthogonal multiwavelets. *Appl. Comp. Harmonic Anal.* 53, 270–331. doi: 10.1016/j.acha.2021.02.006

Han, J., and Hou, S.-M. (2019). Multiple Sclerosis Detection via Wavelet Entropy and Feedforward Neural Network Trained by Adaptive Genetic Algorithm. *Lecture Notes Comp. Sci.* 11507, 87–97. doi: 10.1007/978-3-030-20518-8_8

Han, J., and Hou, S.-M. (2020). "A Multiple Sclerosis Recognition via Hu Moment Invariant and Artificial Neural Network Trained by Particle Swarm Optimization" in *Yu-Dong Zhang 0001, Shui-Hua Wang, Shuai Liu 0009, editors, Multimedia Technology and Enhanced Learning, Second EAI International Conference, ICMTEL 2020, Leicester, UK, April 10-11, 2020, Proceedings, Part II*. (United States: Springer). 327, 254–264. doi: 10.1007/978-3-030-51103-6_22

Hartung, H. P., Meuth, S. G., Miller, D. M., and Comi, G. (2021). Stopping disease-modifying therapy in relapsing and progressive multiple sclerosis. *Curr.Opinion Neurol.* 34, 598–603. doi: 10.1097/WCO.0000000000000960

Hatayama, T., Kano, Y., Aida, A., Chiba, A., Sato, K., Seki, N., et al. (2020). The combined effect of light-illuminating direction and enamel rod orientation on color adjustment at the enamel borders of composite restorations. *Clin. Oral Investig.* 24, 2305–2313. doi: 10.1007/s00784-019-03085-7

Huo, Y., Plassard, A. J., Carass, A., Resnick, S. M., Pham, D. L., Prince, J. L., et al. (2016). Consistent cortical reconstruction and multi-atlas brain segmentation. *Neuroimage* 138, 197–210. doi: 10.1016/j.neuroimage.2016.05.030

Huo, Y. K., Asman, A. J., Plassard, A. J., and Landman, B. A. (2017). Simultaneous total intracranial volume and posterior fossa volume estimation using multi-atlas label fusion. *Hum. Brain Map.* 38, 599–616. doi: 10.1002/hbm.23432

Huo, Y. K., Xu, Z. B., Xiong, Y. X., Aboud, K., Parvathaneni, P., Bao, S. X., et al. (2019). 3D whole brain segmentation using spatially localized atlas network tiles. *Neuroimage* 194, 105–119. doi: 10.1016/j.neuroimage.2019.03.041

Jin, D., Iyer, K. S., Chen, C., Hoffman, E. A., and Saha, P. K. (2016). A robust and efficient curve skeletonization algorithm for tree-like objects using minimum cost paths. *Pattern Recognit. Lett.* 76, 32–40. doi: 10.1016/j.patrec.2015.04.002

Jin, D. K., Guo, D. Z., Ho, T. Y., Harrison, A. P., Xiao, J., Tseng, C. K., et al. (2019). Accurate Esophageal Gross Tumor Volume Segmentation in *PET/CT Using Two-Stream Chained 3D Deep Network Fusion*. China: Springer International Publishing Ag. 182–191. doi: 10.1007/978-3-030-32245-8_21

Jin, D. K., Xu, Z. Y., Tang, Y. B., Harrison, A. P., and Mollura, D. J. (2018). "CT-Realistic Lung Nodule Simulation from 3D Conditional Generative Adversarial Networks for Robust Lung Segmentation" in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*. (SPAIN: Springer International Publishing Ag). 732–740. doi: 10.1007/978-3-030-00934-2_81

Kasnazani, A., and Alipanah, A. (2021). Solving brachistochrone problem via scaling functions of Daubechies wavelets. *Comput. Methods Diff. Equat.* 9, 511–522.

Kiran, R., and Naik, D. L. (2021). Novel sensitivity method for evaluating the first derivative of the feed-forward neural network outputs. *J. Big Data* 8:88. doi: 10.1186/s40537-021-00480-4

Koo, E., and Kim, H. (2021). Empirical strategy for stretching probability distribution in neural-network-based regression. *Neural Netw.* 140, 113–120. doi: 10.1016/j.neunet.2021.02.030

Kshatriya, S., and Prasanna, K. (2021). Jump Interdependencies: stochastic linkages among international stock markets. *North Am. J. Econom. Finance* 57:101418. doi: 10.1016/j.najef.2021.101418

Le Guisquet, S., and Amabili, M. (2021). Identification by means of a genetic algorithm of nonlinear damping and stiffness of continuous structures subjected to large-amplitude vibrations. Part II: one-to-one internal resonances. *Mech. Syst. Signal Process.* 161:107972. doi: 10.1016/j.ymssp.2021.107972

Leow, E. K. W., Nguyen, B. P., and Chua, M. C. H. (2021). Robo-advisor using genetic algorithm and BERT sentiments from tweets for hybrid portfolio optimisation. *Expert Syst. Appl.* 179:115060. doi: 10.1016/j.eswa.2021.115060

Li, J. (2017). Texture Analysis Method Based on Fractional Fourier Entropy and Fitness-scaling Adaptive Genetic Algorithm for Detecting Left-sided and Right-sided Sensorineural Hearing Loss. *Fundam. Inform.* 151, 505–521. doi: 10.3233/FI-2017-1507

Livne-Margolin, M., Latzer, I. T., Pinhas-Hamiel, O., Harari, G., and Achiron, A. (2021). Bodyweight Measures and Lifestyle Habits in Individuals with Multiple Sclerosis and Moderate to Severe Disability. *J. Clin. Med.* 10:2083. doi: 10.3390/jcm10102083

Loizou, C. P., Murray, V., Pattichis, M. S., Seimenis, I., Pantziaris, M., and Pattichis, C. S. (2011). Multiscale Amplitude-Modulation Frequency-Modulation (AM-FM) Texture Analysis of Multiple Sclerosis in Brain MRI Images. *IEEE Transac. Inform. Technol. Biomed.* 15, 119–129. doi: 10.1109/TITB.2010.2091279

Lopez, M. (2017). Multiple Sclerosis Slice Identification by Haar Wavelet Transform and Logistic Regression. *Adv. Eng. Res.* 114, 50–55.

Lu, M. Z., He, J., Chen, C., Okinda, C., Shen, M. X., Liu, L. S., et al. (2018). An automatic ear base temperature extraction method for top view piglet thermal image. *Comput. Electr. Agric.* 155, 339–347. doi: 10.1016/j.compag.2018.10.030

Lu, M. Z., Xiong, Y. J., Li, K. Q., Liu, L. S., Yan, L., Ding, Y. Q., et al. (2016). An automatic splitting method for the adhesive piglets' gray scale image based on the ellipse shape feature. *Comput. Electr. Agric.* 120, 53–62. doi: 10.1016/j.compag.2015.11.008

Machorro-Lopez, J. M., Amezquita-Sanchez, J. P., Valtierra-Rodriguez, M., Carrion-Viramontes, F. J., Quintana-Rodriguez, J. A., and Valenzuela-Delgado, J. I. (2021). Wavelet Energy Accumulation Method Applied on the Rio

Papaloapan Bridge for Damage Identification. *Mathematics* 9:422. doi: 10.3390/math9040422

Misra, I., Rout, L., Arya, S., Bhateja, Y., Moorthi, S. M., and Dhar, D. (2021). Phobos image enhancement using unpaired multi-frame acquisitions from Indian Mars Color Camera. *Planet. Space Sci.* 201:105215. doi: 10.1016/j.pss.2021.105215

Moreno-Andrade, T., Garza-Villarreal, E., Gonzalez-Aquines, A., Villarreal-Garza, E., Lopez-Garza, N., Campos-Coy, M., et al. (2021). Diffusion tensor imaging of the corticospinal pathway and its association with the prognosis of acute cerebral infarction: experience with a cohort in Mexico. *Rev. Neurol.* 72, 16–22. doi: 10.33588/rn.7201.2020232

Motlagh, M. M., Bahar, A., and Bahar, O. (2021). Damage detection in a 3D wind turbine tower by using extensive multilevel 2D wavelet decomposition and heat map, including soil-structure interaction. *Structures* 31, 842–861. doi: 10.1016/j.istruc.2021.01.018

Nayak, D. R., Dash, R., and Majhi, B. (2016). Brain MR image classification using two-dimensional discrete wavelet transform and AdaBoost with random forests. *Neurocomputing* 177, 188–197. doi: 10.1016/j.neucom.2015.11.034

Ng, R. W., Begam, K. M., Rajkumar, R. K., Wong, Y. W., and Chong, L. W. (2021). An improved self-organizing incremental neural network model for short-term time-series load prediction. *Appl. Energy* 292:116912. doi: 10.1016/j.apenergy.2021.116912

Pan, C. (2018). Multiple sclerosis identification by convolutional neural network with dropout and parametric ReLU. *J. Comput. Sci.* 28, 1–10. doi: 10.1016/j.jocs.2018.07.003

Radi, A., Bakalli, G., Guerrier, S., El-Sheimy, N., Sesay, A. B., and Molinari, R. (2019). A Multisignal Wavelet Variance-Based Framework for Inertial Sensor Stochastic Error Modeling. *IEEE Transac. Instrument. Measure.* 68, 4924–4936. doi: 10.1109/TIM.2019.2899535

Rahim, A. A. A., Abdullah, S., Singh, S. S. K., and Nuawi, M. Z. (2021). Fatigue strain signal reconstruction technique based on selected wavelet decomposition levels of an automobile coil spring. *Eng. Fail. Anal.* 125:105434. doi: 10.1016/j.engfailanal.2021.105434

Ren, J. H., Chen, Q., Ma, D. L., Xie, R. F., Zhu, H. L., and Zang, S. Y. (2021). Study on a fast EC measurement method of soda saline-alkali soil based on wavelet decomposition texture feature. *Catena* 203:105272. doi: 10.1016/j.catena.2021.105272

Salem, M., Valverde, S., Cabezas, M., Pareto, D., Oliver, A., Salvi, J., et al. (2020). A fully convolutional neural network for new T2-w lesion detection in multiple sclerosis. *Neuroimage Clin.* 25:102149. doi: 10.1016/j.nicl.2019.102149

Singh, H., Dwivedi, A. K., and Nagaria, D. (2021). Quadrature mirror filter bank design based on hybrid bee colony technique. *Automatika* 62, 264–274. doi: 10.1080/00051144.2021.1933362

So, C., Ho, I. M., Chae, J. S., and Hong, K. H. (2021). PWR core loading pattern optimization with adaptive genetic algorithm. *Ann. Nucl. Energy* 159:108331. doi: 10.1016/j.anucene.2021.108331

Sorkhabi, O. M., Asgari, J., and Simkooei, A. A. (2021). Analysis of Greenland mass changes based on GRACE four-dimensional wavelet decomposition. *Remote Sens. Lett.* 12, 499–509. doi: 10.1080/2150704X.2021.1903608

Sousa, D., Rodrigues, A. I., Carvalho, M., de Almeida, R. F., Castro, S., and Nunez, D. (2021). Neuromyelitis optica as a cause of serious ventilatory compromise. *Postgraduate Med. J.* 97, 403–403. doi: 10.1136/postgradmedj-2020-138472

Srinivas, M., and Patnaik, L. M. (1994). Adaptive probabilities of crossover and mutation in genetic algorithms. *IEEE Transac. Syst. Man Cybernet.* 24, 656–667. doi: 10.1109/21.286385

Tang, C. (2021). Multiple sclerosis detection via 5-layer Convolutional Neural Network. *J. Med. Imaging Health Inform.* 11, 1–8.

To, C., and Liew, A. W. C. (2014). "GENETIC ALGORITHM BASED DETECTION OF GENERAL LINEAR BICLUSTERS" in *International Conference on Machine Learning and Cybernetics*. (CHINA: IEEE). 550–555. doi: 10.1109/ICMLC.2014.7009667

Valizadeh, M., Sohrabi, M. R., and Motiee, F. (2021). Simple spectrophotometric method for simultaneous determination of salmeterol and fluticasone as anti-asthma drugs in inhalation spray based on artificial neural network and support vector regression. *Optik* 240:166879. doi: 10.1016/j.ijleo.2021.166879

van der Goot, R., and van Noord, G. (2017). "Parser Adaptation for Social Media by Integrating Normalization," in *in 55th Annual Meeting of the Association for Computational Linguistics*, (CANADA: Association for Computational Linguistics). 491–497. doi: 10.18653/v1/P17-2078

Vanchurin, V. (2021). Toward a theory of machine learning. *Mach. Lear. Sci. Technol.* 2:035012. doi: 10.1088/2632-2153/abe6d7

Viatkin, D., Garcia-Zapirain, B., and Zorrilla, A. M. (2021). Deep Learning Techniques Applied to Predict and Measure Finger Movement in Patients with Multiple Sclerosis. *Appl. Sci. Basel* 11:3137. doi: 10.3390/app11073137

Westenbrink, S., Presser, K., Roe, M., Ireland, J., and Finglas, P. (2021). Documentation of aggregated/compiled values in food composition databases; EuroFIR default to improve harmonization. *J. Food Compos. Anal.* 101:103968. doi: 10.1016/j.jfca.2021.103968

Yahia, S., Salem, Y. Ben, and Abdelkrim, M. N. (2018). Texture analysis of magnetic resonance brain images to assess multiple sclerosis lesions. *Multimed. Tools Appl.* 77, 30769–30789. doi: 10.1007/s11042-018-6160-9

Zhan, T. M., and Chen, Y. (2016). Multiple Sclerosis Detection Based on Biorthogonal Wavelet Transform, RBF Kernel Principal Component Analysis, and Logistic Regression. *IEEE Access* 4, 7567–7576. doi: 10.1109/ACCESS.2016.2620996

Zhou, Q. (2021). ADVIAN: alzheimer's Disease VGG-Inspired Attention Network Based on Convolutional Block Attention Module and Multiple Way Data Augmentation. *Front. Aging Neurosci.* 13:687456. doi: 10.3389/fnagi.2021.687456

Zhou, Q., and Shen, X. (2018). "Multiple sclerosis identification by grey-level cooccurrence matrix and biogeography-based optimization" in *23rd International Conference on Digital Signal Processing (DSP)*. (China: IEEE). 1–5. doi: 10.1109/ICDSP.2018.8631873

Zucatelli, P. J., Nascimento, E. G. S., Santos, A. A. B., Arce, A. M. G., and Moreira, D. M. (2021). An investigation on deep learning and wavelet transform to nowcast wind power and wind power ramp: a case study in Brazil and Uruguay. *Energy* 230:120842. doi: 10.1016/j.energy.2021.120842

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.