

Article

Not peer-reviewed version

---

# Multiple sclerosis recognition via wavelet entropy and PSO-based neural network

---

[Dimas Lima](#) \*

Posted Date: 10 October 2023

doi: 10.20944/preprints202310.0642.v1

Keywords: multiple sclerosis; wavelet entropy; particle swarm optimization; neural network; machine learning



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Article*

# Multiple Sclerosis Recognition via Wavelet Entropy and PSO-Based Neural Network

Dimas Lima

Department of Electrical Engineering, Federal University of Santa Catarina, 88040-900, Florianópolis, Brazil;  
dimaslima@ieee.org

**Abstract:** This paper presents a novel approach for the recognition of multiple sclerosis (MS) using wavelet entropy and a Particle Swarm Optimization (PSO)-based neural network. MS is a complex neurological disorder with diverse clinical manifestations, often challenging to diagnose accurately. In this study, we leverage wavelet entropy (WE) as a feature extraction method to capture intricate patterns within brain imaging data. These extracted features are then employed to train a neural network model optimized through PSO to enhance classification accuracy. Experimental results on a dataset comprising MS and healthy control subjects demonstrate the effectiveness of our proposed approach, viz., WE-PSONN, which achieves a sensitivity of  $91.95 \pm 1.15$ , a specificity of  $92.36 \pm 0.88$ , a precision of  $92.28 \pm 0.88$ , an accuracy of  $92.16 \pm 0.90$ . The combination of advanced signal processing techniques and machine learning optimization holds promise for improving the early diagnosis and management of MS, offering potential benefits to both patients and healthcare providers. **Keywords:** multiple sclerosis; wavelet entropy; particle swarm optimization; neural network; machine learning

## 1. Introduction

Multiple sclerosis (MS) is a chronic and often disabling neurological condition that primarily affects the central nervous system [1], comprising the brain and spinal cord. It is considered an autoimmune disease, which means that the immune system mistakenly targets and damages the protective covering of nerve fibers called myelin [2]. This damage disrupts the normal flow of electrical impulses along the nerves, leading to a wide range of neurological symptoms [3].

The symptoms of MS can vary significantly from person to person, making it a highly unpredictable condition. Common symptoms include fatigue, muscle weakness, difficulty walking, numbness or tingling sensations, problems with coordination and balance, and visual disturbances [4]. These symptoms can come and go, or they may persist and worsen over time. In some cases, individuals with MS may experience more severe symptoms, such as difficulty with speech, cognitive impairments [5], and issues with bladder and bowel control [6].

While the exact cause of MS remains unknown, it is believed to involve a complex interplay of genetic and environmental factors. There is currently no cure for multiple sclerosis [7], but various treatments are available to manage symptoms, reduce the frequency of relapses, and slow the progression of the disease. These treatments often include medications, physical therapy, and lifestyle modifications to improve overall well-being and quality of life for individuals living with MS [8].

Traditional MS recognition methods have several shortcomings, which have led to the development and adoption of more advanced diagnostic techniques. Some of the main shortcomings [9] of traditional MS recognition methods include (i) Reliance on Clinical Symptoms: Traditional MS diagnosis often relies on clinical symptoms alone. Since MS symptoms can be nonspecific and mimic other conditions [10], this approach may lead to delayed or misdiagnosis, as many other diseases can produce similar symptoms [11]. This delay can hinder early intervention and treatment, potentially allowing the disease to progress. (ii) Lack of Objectivity: Clinical assessment of MS is subjective and can vary from one healthcare provider to another. The lack of objective, quantifiable measures [12] make it challenging to establish a definitive diagnosis and monitor disease progression accurately. (iii) Invasive Procedures: Historically, diagnosing MS required invasive procedures such as a lumbar

puncture (spinal tap) [13] to analyze cerebrospinal fluid for abnormalities. These procedures can be uncomfortable and carry some risks.

Recently, scholars have tended to use machine learning methods to recognize MS. Machine learning methods [14] are a subset of artificial intelligence techniques that enable computers to learn from data and make predictions or decisions without being explicitly programmed [15]. These methods involve the development of algorithms that can analyze and interpret patterns in large datasets, automatically adjusting their parameters to improve performance over time. Machine learning encompasses various approaches, including supervised learning [16] (where models are trained on labeled data to make predictions), unsupervised learning [17] (for discovering patterns and structures in data), and reinforcement learning [18] (for decision-making in dynamic environments). These methods have wide-ranging applications, from natural language processing and image recognition to autonomous robotics and recommendation systems, and are fundamental in enabling computers to perform tasks that require learning from experience. Lopez [19] employed the Haar wavelet transform and logistic regression (LR) method. Han, et al. [20] used an adaptive genetic algorithm (AGA) to detect MS. Zhao, et al. [21] proposed the Dirichlet mixture of Gaussian processes with split-kernel (DMGPS) to identify MS. Han, et al. [22] proposed using Hu moment invariant (HMI) to classify MS.

This paper proposes a novel method that combines wavelet entropy (WE) and a particle swarm optimization-based neural network (PSOINN). The contributions of our study are below:

- We are the first to propose the WE-PSOINN method to recognize MS.
- Our WE-PSOINN gets better results than four state-of-the-art approaches.

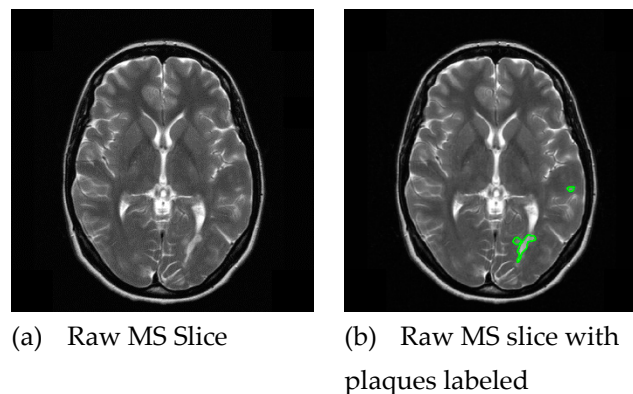
## 2. Dataset

The dataset is from Ref [20]. Its demographic description is listed in **Error! Reference source not found.**, where two categories exist: (i) MS; and (ii) healthy control (HC). **Error! Reference source not found.** shows one sample of MS.

**Table 1.** Demographic characteristics.

Category	NS	NS	Gender (m/f)	Age
MS [23]	38	676	17/21	34.1 ± 10.5
HC [24]	26	681	12/14	33.5 ± 8.3

(NS: Number of Subjects, NS: Number of slices).



**Figure 1.** Sample of MS.

Data harmonization [25] is the process of bringing together and standardizing data from different sources or formats to make it consistent and compatible for analysis, integration, or other purposes. The goal of data harmonization is to create a unified and cohesive dataset that can be used for various analytical, reporting, or decision-making tasks. Histogram stretching (HS) [26], also known as contrast stretching or intensity stretching, is a technique in image processing that aims to improve the contrast and visibility of an image by expanding the range of pixel intensities. It involves

linearly scaling the intensity values of an image so that the minimum and maximum values span the entire available range, typically from 0 to 255 in an 8-bit image [27]. HS can enhance the visual appearance of an image, making details more distinguishable and improving its overall quality. HS is often used in image enhancement and preprocessing to make images more suitable for subsequent analysis or visualization. HS is used in our method due to its ease of implementation. Suppose  $\alpha$  denotes the original brain slice, and  $\beta$  stands for the processed brain slice. Let  $(w, v)$  denote the coordinates, HS operation  $HS: \alpha \mapsto \beta$  is defined via:

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

where  $(\alpha_l, \alpha_h)$  stand for the lowest and highest grayscale intensity values of  $\alpha$ , respectively.

$$\begin{cases} \alpha_l = \min_w \min_v \alpha(w, v) \\ \alpha_h = \max_w \max_v \alpha(w, v) \end{cases} \quad (2)$$

### 3. Methodology

#### 3.1. Wavelet Entropy

Wavelet entropy [28] is a mathematical and statistical measure used in signal processing and data analysis to characterize the complexity or irregularity of a signal or time series data. It is derived from the concept of wavelet transform, which is a mathematical technique for decomposing a signal into different frequency components and analyzing their variations over time [29]. Wavelet entropy quantifies the amount of information or disorder within these frequency components and is particularly useful in fields like neuroscience [30], finance, and environmental science for analyzing complex and non-stationary signals. To compute wavelet entropy, one typically follows these steps:

##### Step 1: Wavelet Transform

The first step involves applying the wavelet transform to the time series data [31]. This transform breaks down the signal into different scales, allowing for the analysis of both low and high-frequency components.

##### Step 2: Probability Distribution

After obtaining the wavelet coefficients, a probability distribution  $P(R)$  is created to represent the magnitude of each scale. This distribution characterizes how the signal's energy is distributed across different scales or frequencies.

##### Step 3: Entropy Calculation

Wavelet entropy [32] is then calculated based on the probability distribution. It measures the unpredictability or randomness within the signal. Higher wavelet entropy values indicate greater complexity, while lower values suggest more regular or predictable patterns.

Shannon entropy is used since it is the originator of modern information theory. Suppose a variable  $R$  is discrete and random, and it falls within the value set  $(r_1, r_2, \dots, r_n)$  with probability mass function of  $P(R)$ , the Shannon entropy is defined as:

$$H(R) = E[-\ln P(R)], \quad (3)$$

where  $H$  is the entropy, and  $E$  represents the expected value operation. If the set of possible values of  $R$  is finite, the formula above is transformed as:

$$H(R) = - \sum_{i=1}^n P(r_i) \log_b P(r_i), \quad (4)$$

where  $b$  is the base of the log function, determining the unit of Shannon entropy.

There are seven coefficient matrices (**Error! Reference source not found.**) after a 2-level 2D-DWT [33], with four in size of 64×64 and three 128-by-128. By using this method, the dimension of the feature vector is reduced from the origin 256×256=65,536 to only seven effectively.

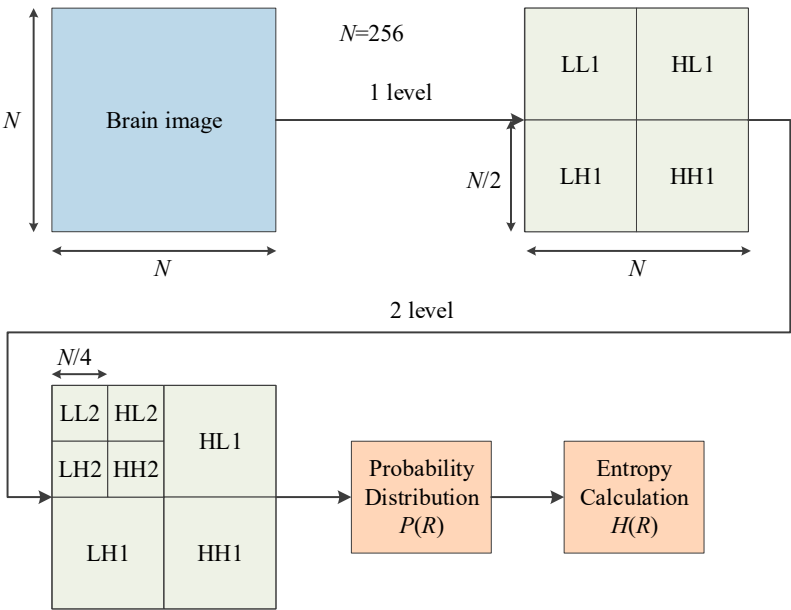


Figure 2. Diagram of 2-level 2D-DWT.

In all, wavelet entropy offers advantages over traditional entropy measures, especially for analyzing signals with varying frequency content and non-stationary behavior. It can capture intricate patterns and fluctuations in time series data, making it a valuable tool for understanding complex systems and processes in various scientific and engineering disciplines.

3.2. Feedforward Neural Network

A feedforward neural network (FNN) [34], often simply called a feedforward network, is a fundamental type of neural network architecture used in machine learning and deep learning. It represents the simplest form of neural network, where information flows in one direction [35], from input to output, without any feedback loops or recurrent connections [36]. Here's a more detailed explanation in four paragraphs:

A FNN consists of multiple layers of interconnected nodes or neurons (**Error! Reference source not found.**). These layers typically include an input layer, one or more hidden layers, and an output layer [37]. Each neuron in a layer is connected to every neuron in the subsequent layer, but there are no connections that loop back within the same layer or to previous layers. Each connection between neurons has an associated weight, which is adjusted during training to enable the network to learn and make predictions [38].

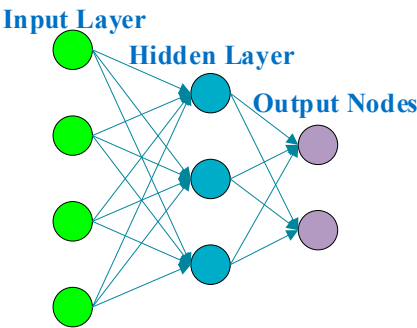


Figure 3. Diagram of an FNN.

The operation of a FNN involves a simple forward propagation of information. Input data is fed into the input layer, and it propagates through the network layer by layer [39]. Neurons in each layer perform a weighted sum of their inputs and pass the result through an activation function. This output then becomes the input for the next layer. The final output is obtained at the output layer [40]. This process is deterministic and does not involve any recurrent or feedback connections.



Activation functions play a crucial role in feedforward networks. They introduce nonlinearity into the model, allowing the network to learn complex relationships in data [41]. Common activation functions include the sigmoid, hyperbolic tangent (tanh), and rectified linear unit (ReLU) functions [42]. These functions determine the output of a neuron based on its weighted input sum and introduce the concept of thresholds and saturation points [43].

FNNs are trained using various supervised learning algorithms, with the most popular being backpropagation. During training, the network learns to adjust its weights to minimize the difference between its predictions (output) and the target values in the training data. This is typically done by computing gradients of an error or loss function with respect to the network's weights and then adjusting the weights [44] in the direction that reduces the error. This process is repeated iteratively until the network's performance on the training data reaches a satisfactory level [45].

### 3.3. PSO-Based Neural Network

A particle swarm optimization-based neural network (PSONN) is a combination of two distinct computational techniques: particle swarm optimization (PSO) [46] and feedforward neural networks (FNNs).

PSO [47] is a heuristic optimization algorithm inspired by the social behavior of birds or fish swarming, where individual particles adjust their positions in search of an optimal solution. FNNs, on the other hand, are a class of machine learning models inspired by the human brain that can learn and make predictions from data. Combining PSO with FNNs aims to improve the training and optimization process of neural networks [48].

PSO is an optimization algorithm that simulates the behavior of a swarm of particles in a multidimensional search space. Each particle represents a potential solution to the optimization problem. These particles iteratively update their positions and velocities based on their own experiences and the experiences of their peers in the swarm. PSO is particularly useful for finding global optima in complex, high-dimensional search spaces. In the context of neural networks, PSO can be applied to optimize the weights and biases of the network to minimize a specific cost or error function [49].

In a PSONN, the PSO algorithm [50] is used to optimize the weights and biases of the neural network. Each particle in the PSO swarm corresponds to a set of neural network weights and biases. The objective is to find the combination of weights and biases that minimizes the error or cost function associated with the neural network's performance on a specific task. PSO guides the search in weight space [51], helping the neural network converge to a more optimal solution. The fitness function used in PSO is often related to the neural network's error on a training dataset.

PSONNs offer advantages in optimizing neural network architectures, making them more efficient and capable of solving complex problems. They are used in various applications, including pattern recognition, time series prediction, feature selection, and optimization of deep neural networks. By combining the strengths of PSO's global search capabilities with the learning and generalization abilities of FNNs, this approach can enhance the performance and efficiency of neural network models, making them more suitable for real-world applications. In the future, we shall try more advanced and recent optimization methods [52].

### 3.4. 10-Fold Cross Validation

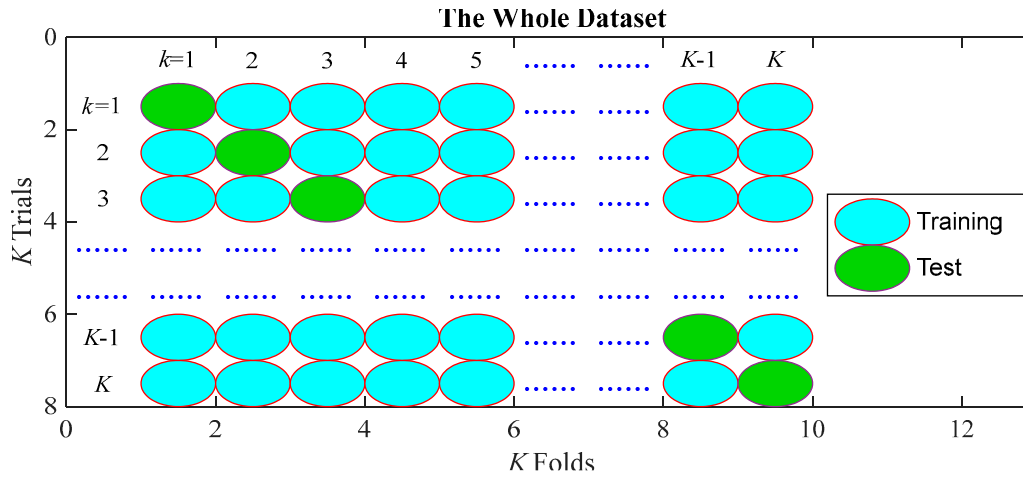
$K$ -fold cross-validation is a widely used technique in machine learning and model evaluation that helps assess the performance of a predictive model while maximizing the use of available data [53]. It involves splitting the dataset into  $K$  subsets, or "folds," where  $K$  is typically a positive integer like 10. The process can be summarized in four key aspects:

**Step 1: Dataset Splitting.** The first step in  $K$ -fold cross-validation is to divide the dataset into  $K$  roughly equal-sized parts or folds. Each fold represents a subset of the data, and these subsets are often randomly selected to ensure that the cross-validation process is not biased by the order of the data.

**Step 2: Training and Testing.**  $K$ -fold cross-validation then proceeds through  $K$  iterations. During each iteration, one of the  $K$  folds is held out as the test set, while the remaining  $K-1$  folds are used as the training set. This means that every data point in the dataset is used for testing exactly once.

Step 3: Model Training and Evaluation. In each iteration, a predictive model (e.g., a machine learning algorithm) is trained on the training set, and its performance is evaluated on the corresponding test set. Common performance metrics, such as accuracy, precision, recall, or mean squared error, are computed to assess how well the model generalizes to unseen data [54].

Step 4: Cross-Validation Results. After completing all  $K$  iterations,  $K$  separate performance metrics are obtained, one for each fold. These metrics are typically averaged to produce a single, overall performance measure that provides a more robust estimate of how well the model is likely to perform on new, unseen data. This average performance metric is often used to compare different models or tuning parameters [55]. **Error! Reference source not found.** shows the diagram of  $K$ -fold cross validation.



**Figure 4.** Diagram of  $K$ -fold cross validation.

$K$ -fold cross-validation is a valuable technique for several reasons. It helps in assessing a model's performance more reliably than a single train-test split because it uses all available data for both training and testing [56]. It also provides insights into the model's stability and generalization performance by examining its performance on different subsets of the data. Additionally,  $K$ -fold cross-validation is particularly useful when the dataset is limited in size, as it allows for more efficient and effective use of the available data for model evaluation. Overall, it is a key tool for model selection, hyperparameter tuning, and assessing a model's expected performance in real-world applications.

### 3.5. Measure on $Z$ runs

Suppose we carry out  $K$ -fold cross-validation  $Z$  runs. Suppose the confusion matrix  $\mathbf{M}(z)$  over  $z$ -th run ( $z = 1, 2, \dots, Z$ ) is

$$\mathbf{M}(z) = \begin{bmatrix} m_{11}(z) & m_{12}(z) \\ m_{21}(z) & m_{22}(z) \end{bmatrix} \quad (1)$$

where the four elements  $(m_{11}, m_{12}, m_{21}, m_{22})$  represent TP, FN, FP, and TN, respectively. Here P means the positive category, MS, and N is the negative category, HC.

The sensitivity [57] (symbolized as  $f_1$ ), specificity [58] (symbolized as  $f_2$ ), precision (symbolized as  $f_3$ ), and accuracy (symbolized as  $f_4$ ) of  $z$ -th run are defined as:

$$\begin{cases} f_1(z) = \frac{m_{11}(z)}{m_{11}(z) + m_{12}(z)} \\ f_2(z) = \frac{m_{22}(z)}{m_{21}(z) + m_{22}(z)} \\ f_3(z) = \frac{m_{11}(z)}{m_{11}(z) + m_{21}(z)} \\ f_4(z) = \frac{m_{11}(z) + m_{22}(z)}{m_{11}(z) + m_{12}(z) + m_{21}(z) + m_{22}(z)} \end{cases}, z = 1, 2, \dots, Z. \quad (2)$$

The F1 score [59] of  $z$ -th run  $f_5(z)$  is defined as

$$f_5(z) = 2 \times \frac{f_3(z) \times f_1(z)}{f_3(z) + f_1(z)} = \frac{2 \times m_{11}(z)}{2 \times m_{11}(z) + m_{12}(z) + m_{21}(z)}. \quad (3)$$

Matthews correlation coefficient (MCC) [60] of  $z$ -th run  $f_6(z)$  is:

$$f_6(z) = \frac{m_{22}(z) \times m_{11}(z) - m_{21}(z) \times m_{12}(z)}{\sqrt{[m_{21}(z) + m_{11}(z)] \times [m_{11}(z) + m_{12}(z)] \times [m_{22}(z) + m_{21}(z)] \times [m_{22}(z) + m_{12}(z)]}} \quad (4)$$

Fowlkes–Mallows index (FMI) [61] of  $z$ -th run  $f_7(z)$  is defined as:

$$f_7(z) = \sqrt{\frac{m_{11}(z)}{m_{11}(z) + m_{21}(z)} \times \frac{m_{11}(z)}{m_{11}(z) + m_{12}(z)}} \quad (5)$$

After running up all the  $Z$  runs, we deduce the mean and standard deviation (MSD, symbolized as  $a_k \pm b_k$ ,  $k = 1, \dots, 7$ ) are defined below:

$$\begin{cases} a_k = \frac{1}{Z} \times \sum_{z=1}^Z f_k(z) \\ b_k = \sqrt{\frac{1}{Z-1} \times \sum_{z=1}^Z [f_k(z) - a_k]^2}, k = 1, 2, \dots, 7 \end{cases} \quad (6)$$

The receiver operating characteristic (ROC) curve and the area under the curve (AUC) are reported based on  $Z$  runs.

## 4. Results and Discussions

### 4.1. Statistical Analysis

**Error! Reference source not found.** shows the statistical analysis of our WE-PSONN method, which achieves a sensitivity of  $91.95 \pm 1.15$ , a specificity of  $92.36 \pm 0.88$ , a precision of  $92.28 \pm 0.88$ , an accuracy of  $92.16 \pm 0.90$ , an F1 score of  $92.11 \pm 0.92$ , an MCC value of  $84.32 \pm 1.79$ , and an FMI of  $92.12 \pm 0.91$ .

**Table 2.** Statistical Analysis (Unit: %).

Run	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$
1	92.60	91.63	91.65	92.11	92.13	84.23	92.13
2	90.38	90.90	90.79	90.64	90.59	81.28	90.59
3	93.05	93.83	93.74	93.44	93.39	86.88	93.39
4	90.68	93.10	92.88	91.89	91.77	83.81	91.77
5	90.09	91.34	91.17	90.71	90.62	81.43	90.63
6	92.46	92.66	92.59	92.56	92.52	85.11	92.52
7	92.75	92.51	92.48	92.63	92.61	85.26	92.61
8	93.20	92.80	92.78	93.00	92.99	86.00	92.99
9	91.86	92.80	92.69	92.34	92.27	84.67	92.27
10	92.46	92.07	92.05	92.26	92.25	84.53	92.25
MSD	91.95	92.36	92.28	92.16	92.11	84.32	92.12
	$\pm 1.15$	$\pm 0.88$	$\pm 0.88$	$\pm 0.90$	$\pm 0.92$	$\pm 1.79$	$\pm 0.91$



4.2. ROC Curve

The ROC curve is a graphical representation used in machine learning and statistics to evaluate the performance of binary classification models. It is a plot that illustrates the trade-off between a model's true positive rate (TPR, sensitivity) and its false positive rate (FPR, i.e., 1 - specificity) across different threshold values. The ROC curve is created by plotting TPR (sensitivity) on the y-axis against FPR (1 - specificity) on the x-axis at various threshold values. Each point on the curve corresponds to a different threshold setting for the classification model. A diagonal line in the ROC space (from [0,0] to [1,1]) represents random guessing, where the model's performance is no better than chance.

A perfect classifier would have a ROC curve that hugs the upper-left corner of the plot, with TPR equal to 1 and FPR equal to 0. The closer the ROC curve is to the upper-left corner, the better the model's performance in distinguishing between the two classes. The area under the ROC curve (AUC) is also a commonly used metric to quantify the overall performance of a classification model. AUC ranges from 0 to 1, with higher values indicating better model performance. **Error! Reference source not found.** shows the AUC of our WE-PSO method is 0.9489.

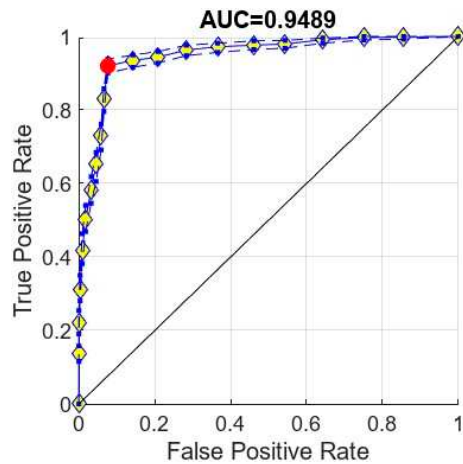


Figure 5. RoC curve of our WE-PSO.

4.3. PSO versus AGA

Han, et al. [20] use WE as a feature and AGA as the optimization algorithm. The AGA is a variant of the traditional Genetic Algorithm (GA) that incorporates adaptive mechanisms to improve the algorithm's performance during the optimization process [62]. Genetic Algorithms are a type of evolutionary algorithm inspired by the process of natural selection and genetics. They are used for solving optimization and search problems by iteratively evolving a population of candidate solutions to find the best solution or approximate the optimal solution [63].

We compare our method with AGA [20], and the comparison results are displayed in **Error! Reference source not found.**. It shows the PSO is better than AGA in all seven measures. The reason may be the parameter sensitivity. PSO typically has fewer algorithm parameters to tune compared to AGAs, making it easier to set up and less sensitive to parameter choices. AGAs often require careful tuning of parameters such as mutation rates, crossover rates, and selection mechanisms, which can be challenging but also provide more flexibility for customization.

Table 3. PSO versus AGA.

Method	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$
AGA [20]	91.91±1.24	91.98±1.36	91.97±1.32	91.95±1.19	91.92±1.20	83.89±2.41	91.92±1.19
PSO (Ours)	91.95±1.15	92.36±0.88	92.28±0.88	92.16±0.90	92.11±0.92	84.32±1.79	92.12±0.91

4.4. Comparison with State-of-the-Art Algorithms

We compare our method with three state-of-the-art approaches: LR [19], DMGPS [21], and HMI [22]. **Error! Reference source not found.** shows the proposed WE-PSO gives better results than other state-of-the-art approaches. The reason is twofold, shown below.

First, WE is a good feature extractor [64] because it quantifies the complexity or irregularity of signals in a multi-scale fashion, allowing it to capture important information at various levels of detail. By decomposing a signal into different frequency components using wavelet transform and then measuring the entropy of these components, it provides a robust and informative representation of the signal's characteristics. High wavelet entropy values indicate increased complexity and fine-grained details in the signal, while low values suggest more regular and predictable patterns. This adaptability to different signal complexities makes wavelet entropy a valuable feature for a wide range of applications, including signal processing, image analysis, and biomedical signal classification, as it can effectively discriminate between various signal types and highlight relevant features.

Second, PSO can optimize the hyperparameters of the neural network, such as the learning rate, the number of hidden layers, and the number of neurons in each layer. This automatic tuning process can lead to better convergence and faster training. PSO is a global optimization technique, which means it explores a wide range of parameter settings for the neural network. This can help ensure that the model does not get stuck in local optima and finds a better overall solution.

Table 4. Method Comparison.

Method	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$
LR [19]	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
DMGPS [21]	88.99±1.20	88.56±1.13	88.54±1.05	88.78±0.95	88.76±0.96	77.56±1.91	88.77±0.96
HMI [22]	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
WE-PSO	91.95±1.15	92.36±0.88	92.28±0.88	92.16±0.90	92.11±0.92	84.32±1.79	92.12±0.91
(Ours)							

5. Conclusions

In conclusion, this study demonstrates a promising approach to improving the accuracy of multiple sclerosis diagnosis. By leveraging wavelet entropy for feature extraction and optimizing neural network parameters with a Particle Swarm Optimization (PSO) algorithm, the research shows the potential for more precise and reliable identification of this complex neurological condition. The combination of advanced signal processing [65,66] techniques and machine learning methods offers a valuable contribution to the field of medical image analysis [67,68], paving the way for enhanced early detection and management of multiple sclerosis, ultimately benefiting patients and healthcare providers alike.

Further research and validation of this approach hold the potential to advance our understanding and treatment of multiple sclerosis. We shall explore advanced feature selection techniques [69] to identify the most relevant wavelet entropy features for multiple sclerosis recognition. This could involve employing machine learning algorithms or domain-specific knowledge to optimize feature extraction [70]. Also, conducting larger-scale clinical studies to validate the WE-PSO model's performance on diverse patient populations. Gathering data from multiple healthcare institutions can help assess the model's generalizability and robustness.

**Funding:** This research did not receive any grants.

**Acknowledgment:** We thank all the anonymous reviewers for their hard reviewing work.

**Conflict of Interest:** The author declares there is no conflict of interest regarding the publication of this paper.

## References

1. A. B. Kornbluh and I. Kahn, "Pediatric Multiple Sclerosis," *Seminars in Pediatric Neurology*, vol. 46, Article ID: 101054, 2023.
2. S. Quigley, M. Asad, C. Doherty, D. Byrne, S. Cronin, and H. Kearney, "Concurrent diagnoses of Tuberous sclerosis and multiple sclerosis," *Multiple Sclerosis and Related Disorders*, vol. 71, Article ID: 104586, 2023.
3. J. Maroto-García, A. Martínez-Escribano, V. Delgado-Gil, M. Mañez, C. Mugueta, N. Varo, *et al.*, "Biochemical biomarkers for multiple sclerosis," *Clinica Chimica Acta*, vol. 548, Article ID: 117471, 2023.
4. C. Chagot, M. B. Vlaicu, S. Frismand, S. Colnat-Coulbois, J. P. Nguyen, and S. Palfi, "Deep brain stimulation in multiple sclerosis-associated tremor. A large, retrospective, longitudinal open label study, with long-term follow-up," *Multiple Sclerosis and Related Disorders*, vol. 79, Article ID: 104928, 2023.
5. S. M. Lunin, E. G. Novoselova, O. V. Glushkova, S. B. Parfenyuk, A. A. Kuzekova, T. V. Novoselova, *et al.*, "Protective effect of exogenous peroxiredoxin 6 and thymic peptide thymulin on BBB conditions in an experimental model of multiple sclerosis," *Archives of Biochemistry and Biophysics*, vol. 746, Article ID: 109729, 2023.
6. D. E. Freedman, K. M. Krysko, and A. Feinstein, "Intimate partner violence and multiple sclerosis," *Multiple Sclerosis Journal*, 2023.
7. S. Cerri, A. Hoopes, D. N. Greve, M. Muhlau, and K. Van Leemput, "A Longitudinal Method for Simultaneous Whole-Brain and Lesion Segmentation in Multiple Sclerosis," *Arxiv*, 2020.
8. M. Atapour, F. Maghaminejad, S. Ghiyasvandian, M. Rahimzadeh, and S. Hosseini, "The psychometric properties of the Persian version of the Multiple Sclerosis Self-Management Scale-Revised: A cross-sectional methodological study," *Health Science Reports*, vol. 6, Article ID: e1515, 2023.
9. A. Shafiee, H. Soltani, M. M. T. Athar, K. Jafarabady, and P. Mardi, "The prevalence of depression and anxiety among Iranian people with multiple sclerosis: A systematic review and meta-analysis," *Multiple Sclerosis and Related Disorders*, vol. 78, Article ID: 104922, 2023.
10. J. J. Eisler, G. Disanto, R. Sacco, C. Zecca, and C. Gobbi, "Influence of Disease Modifying Treatment, Severe Acute Respiratory Syndrome Coronavirus 2 Variants and Vaccination on Coronavirus Disease 2019 Risk and Outcome in Multiple Sclerosis and Neuromyelitis Optica," *Journal of Clinical Medicine*, vol. 12, Article ID: 5551, 2023.
11. A. Turcic, B. Radovani, Z. Vogrinc, M. Habek, D. Rogic, T. Gabelic, *et al.*, "Higher MRI lesion load in multiple sclerosis is related to the N-glycosylation changes of cerebrospinal fluid immunoglobulin G," *Multiple Sclerosis and Related Disorders*, vol. 79, Article ID: 104921, 2023.
12. Z. Karaaslan, B. Sengul-Yediel, H. Yuceer-Korkmaz, E. Sanlia, D. Gezen-Ak, E. Dursun, *et al.*, "Chloride intracellular channel protein-1 (CLIC1) antibody in multiple sclerosis patients with predominant optic nerve and spinal cord involvement," *Multiple Sclerosis and Related Disorders*, vol. 78, Article ID: 104940, 2023.
13. C. G. Chisari, J. Guadagno, P. Adjamian, C. V. Silvan, T. Greco, M. Bagul, *et al.*, "A post hoc evaluation of the shift in spasticity category in individuals with multiple sclerosis-related spasticity treated with nabiximols," *Therapeutic Advances in Neurological Disorders*, vol. 16, Article ID: 17562864231195513, 2023.
14. Y. Zhang, "Pathological brain detection in MRI scanning via Hu moment invariants and machine learning," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 29, pp. 299-312, 2017.
15. M. Lasalvia, C. Gallo, V. Capozzi, and G. Perna, "Discrimination of Healthy and Cancerous Colon Cells Based on FTIR Spectroscopy and Machine Learning Algorithms," *Applied Sciences-Basel*, vol. 13, Article ID: 10325, 2023.
16. E. O. Kiyak, B. Ghasemkhani, and D. Birant, "High-Level K-Nearest Neighbors (HLKNN): A Supervised Machine Learning Model for Classification Analysis," *Electronics*, vol. 12, Article ID: 3828, 2023.
17. G. Han, Y. F. Wang, J. X. Liu, and F. Y. Zeng, "Low-light images enhancement and denoising network based on unsupervised learning multi-stream feature modeling," *Journal of Visual Communication and Image Representation*, vol. 96, Article ID: 103932, 2023.
18. X. Liu and Z. Q. Li, "Dynamic multiple access based on deep reinforcement learning for Internet of Things," *Computer Communications*, vol. 210, pp. 331-341, 2023.
19. M. Lopez, "Multiple Sclerosis Slice Identification by Haar Wavelet Transform and Logistic Regression," *Advances in Engineering Research*, vol. 114, pp. 50-55, 2017.
20. J. Han and S.-M. Hou, "Multiple Sclerosis Detection via Wavelet Entropy and Feedforward Neural Network Trained by Adaptive Genetic Algorithm," *Lecture Notes in Computer Science*, vol. 11507, pp. 87-97, 2019.

21. Y. J. Zhao and T. Chitnis, "Dirichlet Mixture of Gaussian Processes with Split-kernel: An Application to Predicting Disease Course in Multiple Sclerosis Patients," in *International Joint Conference on Neural Networks (IJCNN)*, Padua, ITALY, 2022, pp. 1-8.
22. J. Han and S.-M. Hou, "A Multiple Sclerosis Recognition via Hu Moment Invariant and Artificial Neural Network Trained by Particle Swarm Optimization," *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, vol. 327, pp. 254-264, 2020.
23. "MRI Lesion Segmentation in Multiple Sclerosis Database," ed. eHealth laboratory, University of Cyprus, 2021.
24. Y.-D. Zhang, "Multiple sclerosis identification by convolutional neural network with dropout and parametric ReLU," *Journal of Computational Science*, vol. 28, pp. 1-10, 2018.
25. J. L. Gleason, R. Tamburro, and C. Signore, "Promoting Data Harmonization of COVID-19 Research in Pregnant and Pediatric Populations," *Jama-Journal of the American Medical Association*, vol. 330, pp. 497-498, 2023.
26. C. J. Norris and S. Kim, "A Use Case of Iterative Logarithmic Floating-Point Multipliers: Accelerating Histogram Stretching on Programmable SoC," in *56th IEEE International Symposium on Circuits and Systems (ISCAS)*, Monterey, CA, 2023, pp. 1-11.
27. S.-H. Wang, "Polarimetric synthetic aperture radar image segmentation by convolutional neural network using graphical processing units," *Journal of Real-Time Image Processing*, vol. 15, pp. 631-642, 2018.
28. D. Dwivedi, A. Chamoli, and S. K. Rana, "Wavelet Entropy: A New Tool for Edge Detection of Potential Field Data," *Entropy*, vol. 25, Article ID: 240, 2023.
29. F. Hua, T. H. Ling, W. C. He, and X. J. Liu, "Wavelet Entropy-Based Method for Migration Imaging of Hidden Microcracks by Using the Optimal Wave Velocity," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 36, Article ID: 2254021, 2022.
30. S. Wang, "Detection of Dendritic Spines using Wavelet Packet Entropy and Fuzzy Support Vector Machine," *CNS & Neurological Disorders - Drug Targets*, vol. 16, pp. 116-121, 2017.
31. J. J. Wang, "COVID-19 Diagnosis by Wavelet Entropy and Particle Swarm Optimization," in *18th International Conference on Intelligent Computing (ICIC)*, Xian, PEOPLES R CHINA, 2022, pp. 600-611.
32. A. S. Janani, S. A. Rezaeieh, A. Darvazehban, S. E. Keating, and A. M. Abbosh, "Portable Electromagnetic Device for Steatotic Liver Detection Using Blind Source Separation and Shannon Wavelet Entropy," *Ieee Journal of Electromagnetics Rf and Microwaves in Medicine and Biology*, vol. 6, pp. 546-554, 2022.
33. Y. T. Tao, T. Scully, A. G. Perera, A. Lambert, and J. Chahl, "A Low Redundancy Wavelet Entropy Edge Detection Algorithm," *Journal of Imaging*, vol. 7, Article ID: 188, 2021.
34. D. Valluri, R. Campbell, and Ieee, "On the Space of Coefficients of a Feedforward Neural Network," in *International Joint Conference on Neural Networks (IJCNN)*, Broadbeach, AUSTRALIA, 2023.
35. K. Yotov, E. Hadzhikolev, S. Hadzhikoleva, and S. Cheresharov, "A Method for Extrapolating Continuous Functions by Generating New Training Samples for Feedforward Artificial Neural Networks," *Axioms*, vol. 12, Article ID: 759, 2023.
36. L. N. Wu, "Improved image filter based on SPCNN," *Science In China Series F-Information Sciences*, vol. 51, pp. 2115-2125, 2008.
37. M. R. Admon, N. Senu, A. Ahmadian, Z. A. Majid, and S. Salahshour, "A new efficient algorithm based on feedforward neural network for solving differential equations of fractional order," *Communications in Nonlinear Science and Numerical Simulation*, vol. 117, Article ID: 106968, 2023.
38. M. Abernot, C. Delacour, A. Suna, J. M. Gregg, S. Karg, A. Todri-Sanial, *et al.*, "Two-Layered Oscillatory Neural Networks with Analog Feedforward Majority Gate for Image Edge Detection Application," in *56th IEEE International Symposium on Circuits and Systems (ISCAS)*, Monterey, CA, 2023.
39. S. Sharma, P. D. Achlerkar, P. Shrivastava, A. Garg, B. K. Panigrahi, and Ieee, "Combined SoC and SoE Estimation of Lithium-ion Battery using Multi-layer Feedforward Neural Network," in *10th IEEE International Conference on Power Electronics, Drives and Energy Systems (IEEE PEDES)*, Malaviya Natl Inst Technol Jaipur, Dept Elect Engn, Jaipur, INDIA, 2022.
40. X. K. Le, N. Wang, and X. Jiang, "Nuclear mass predictions with multi-hidden-layer feedforward neural network," *Nuclear Physics A*, vol. 1038, Article ID: 122707, 2023.
41. K. Yurtkan, A. Adalier, and U. Tekgüç, "Student Success Prediction Using Feedforward Neural Networks," *Romanian Journal of Information Science and Technology*, vol. 26, pp. 121-136, 2023.

42. S.-H. Wang, "Advances and Challenges of Deep Learning," *Recent Patents on Engineering*, vol. 17, pp. 1-2, 2023.
43. J. Zhou, "Logistics inventory optimization method for agricultural e-commerce platforms based on a multilayer feedforward neural network," *Pakistan Journal of Agricultural Sciences*, vol. 60, pp. 487-496, 2023.
44. H. Ben Braiek and F. Khomh, "Testing Feedforward Neural Networks Training Programs," *Acm Transactions on Software Engineering and Methodology*, vol. 32, Article ID: 105, 2023.
45. R. Sapna, S. N. Sheshappa, P. Vijayakarthish, and S. P. Raj, "Global Pattern Feedforward Neural Network Structure with Bacterial Foraging Optimization towards Medicinal Plant Leaf Identification and Classification," *International Journal of Advanced Computer Science and Applications*, vol. 13, pp. 63-70, 2022.
46. K. Bidyanath, S. D. Singh, and S. Adhikari, "Implementation of genetic and particle swarm optimization algorithm for voltage profile improvement and loss reduction using capacitors in 132 kV Manipur transmission system," *Energy Reports*, vol. 9, pp. 738-746, 2023.
47. M. R. Karami, B. Jaleh, M. Eslamipana, A. Nasri, and K. Y. Rhee, "Design and optimization of a TiO<sub>2</sub>/RGO-supported epoxy multilayer microwave absorber by the modified local best particle swarm optimization algorithm," *Nanotechnology Reviews*, vol. 12, Article ID: 20230121, 2023.
48. S. Wang, "Magnetic resonance brain classification by a novel binary particle swarm optimization with mutation and time-varying acceleration coefficients," *Biomedical Engineering-Biomedizinische Technik*, vol. 61, pp. 431-441, 2016.
49. J. Ghaeb, I. Al-Naimi, and M. Alkayali, "Intelligent Integrated Approach for Voltage Balancing Using Particle Swarm Optimization and Predictive Models," *Journal of Electrical and Computer Engineering*, vol. 2023, Article ID: 8864216, 2023.
50. E. Karpat and F. Imamoglu, "Optimization and Comparative Analysis of Quarter-Circular Slotted Microstrip Patch Antenna Using Particle Swarm and Fruit Fly Algorithms," *International Arab Journal of Information Technology*, vol. 20, pp. 624-631, 2023.
51. O. M. Katipoglu, S. N. Yesilyurt, H. Y. Dalkilic, and F. Akar, "Application of empirical mode decomposition, particle swarm optimization, and support vector machine methods to predict stream flows," *Environmental Monitoring and Assessment*, vol. 195, Article ID: 1108, 2023.
52. S.-H. Wang, "Diagnosis of COVID-19 by Wavelet Renyi Entropy and Three-Segment Biogeography-Based Optimization," *International Journal of Computational Intelligence Systems*, vol. 13, pp. 1332-1344, 2020.
53. I. A. D. Sousa, I. A. Reis, A. S. Pagano, J. Telfair, and H. D. Torres, "Translation, cross-cultural adaptation and validation of the sickle cell self-efficacy scale (SCSES)," *Hematology Transfusion and Cell Therapy*, vol. 45, pp. 290-296, 2023.
54. Q. Yu, J. Q. Guo, and F. Y. Gong, "Construction and Validation of a Diagnostic Scoring System for Predicting Active Pulmonary Tuberculosis in Patients with Positive T-SPOT Based on Indicators Associated with Coagulation and Inflammation: A Retrospective Cross-Sectional Study," *Infection and Drug Resistance*, vol. 16, pp. 5755-5764, 2023.
55. V. Ton, "Validation of a mathematical model of ultrasonic cross-correlation flow meters based on industrial experience," *Flow Measurement and Instrumentation*, vol. 93, Article ID: 102438, 2023.
56. Q. K. Mahmood, A. Jalil, M. Farooq, M. S. Akbar, and F. Fischer, "Development and validation of the Post-Pandemic Fear of Viral Disease scale and its relationship with general anxiety disorder: a cross-sectional survey from Pakistan," *Bmc Public Health*, vol. 23, Article ID: 1739, 2023.
57. V. Liyanage, M. Q. Tao, J. S. Park, K. N. Wang, and S. Azimi, "Malignant and non-malignant oral lesions classification and diagnosis with deep neural networks," *Journal of Dentistry*, vol. 137, Article ID: 104657, 2023.
58. I. G. Ivanov, Y. Kumchev, and V. J. Hooper, "An Optimization Precise Model of Stroke Data to Improve Stroke Prediction," *Algorithms*, vol. 16, Article ID: 417, 2023.
59. S. Bansal, K. Gowda, and N. Kumar, "Multilingual personalized hashtag recommendation for low resource Indic languages using graph-based deep neural network," *Expert Systems with Applications*, vol. 236, Article ID: 121188, 2024.
60. L. Abirami and J. Karthikeyan, "Digital Twin-Based Healthcare System (DTHS) for Earlier Parkinson Disease Identification and Diagnosis Using Optimized Fuzzy Based k-Nearest Neighbor Classifier Model," *IEEE Access*, vol. 11, pp. 96661-96672, 2023.



61. G. Alkawsi, R. Al-amri, Y. Baashar, S. Ghorashi, E. Alabdulkreem, and S. K. Tiong, "Towards lowering computational power in IoT systems: Clustering algorithm for high-dimensional data stream using entropy window reduction," *Alexandria Engineering Journal*, vol. 70, pp. 503-513, 2023.
62. M. Dejene, H. Palnivel, H. Senthamarai, V. Varadharajan, S. V. Prabhu, A. Yeshtila, *et al.*, "Optimisation of culture conditions for gesho (*Rhamnus prinoides*.L) callus differentiation using Artificial Neural Network-Genetic Algorithm (ANN-GA) Techniques," *Applied Biological Chemistry*, vol. 66, Article ID: 64, 2023.
63. H. Nakamura, J. W. Zhang, K. Hirose, K. Shimoyama, T. Ito, and T. Kanaumi, "Generating simplified ammonia reaction model using genetic algorithm and its integration into numerical combustion simulation of 1 MW test facility," *Applications in Energy and Combustion Science*, vol. 15, Article ID: 100187, 2023.
64. Y. Zhang, "Feature Extraction of Brain MRI by Stationary Wavelet Transform and its Applications," *Journal of Biological Systems*, vol. 18, pp. 115-132, 2010.
65. Y. F. Chen, Q. J. Yu, and Y. J. Lu, "Pulsar Timing Array Detections of Supermassive Binary Black Holes: Implications from the Detected Common Process Signal and Beyond," *Astrophysical Journal*, vol. 955, Article ID: 132, 2023.
66. C. R. Zhou, Z. T. Wang, S. W. Yang, Y. H. Xie, P. Liang, D. Cao, *et al.*, "Application Progress of Chemometrics and Deep Learning Methods in Raman Spectroscopy Signal Processing," *Chinese Journal of Analytical Chemistry*, vol. 51, pp. 1232-1242, 2023.
67. M. Kolhar and S. M. Aldossary, "Privacy-Preserving Convolutional Bi-LSTM Network for Robust Analysis of Encrypted Time-Series Medical Images," *Ai*, vol. 4, pp. 706-720, 2023.
68. D. Y. Wu, L. Li, J. Y. Wang, P. H. Ma, Z. C. Wang, and H. Z. Wu, "Robust zero-watermarking scheme using DT CWT and improved differential entropy for color medical images," *Journal of King Saud University-Computer and Information Sciences*, vol. 35, Article ID: 101708, 2023.
69. Y. Zhang, "Pathological brain detection in MRI scanning by wavelet packet Tsallis entropy and fuzzy support vector machine," *SpringerPlus*, vol. 4, Article ID: 716, 2015.
70. S. K. Shandilya, A. Srivastav, K. Yemets, A. Datta, and A. K. Nagar, "YOLO-base d segmente d dataset for drone vs. bird detection for deep and machine learning algorithms," *Data in Brief*, vol. 50, Article ID: 109355, 2023.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.