Multiscale Adaptive Multifractal Detrended Fluctuation Analysis-Based Source Identification of Synchrophasor Data

Yi Cui, *Senior Member, IEEE*, Feifei Bai, *Senior Member, IEEE,* Hongzhi Yin, *Senior Member IEEE,* Tong Chen, *Member IEEE,* David Dart, Matthew Zillmann, Ryan K. L. Ko, *Member, IEEE*

***Abstract*- As a typical cyber-physical system, dispersed Phasor Measurement Units (PMUs) are networked together with advanced communication infrastructures to record the Distribution Synchrophasor (DS) which represents the states and dynamics of distribution power networks. Source information of DS is critical for many DS-based applications which is potentially vulnerable to data integrity attacks. To ensure the reliability of DS-based applications, it is imperative to efficiently authenticate the DS source locations before any DS data analytics is initiated. This letter presents a cost-effective method for accurate source identification by realising the multifractality of DS data. First, Multiscale Adaptive Multifractal Detrended Fluctuation Analysis (MSA-MFDFA) is executed to reveal the scale which possesses the most significant multifractality of the time-series DS. Subsequently, Adaptive Multifractal Interpolation (AMFI) is proposed to generate quasi high-resolution DS where unique time-frequency signatures are extracted. Such signatures are further fed into a deep learning model - deep forest for source identification. Experimental results using real-life DS of a distribution network illustrate the excellent performance of the proposed approach.**

***Index Terms*- Source identification, synchrophasor, distribution power networks, cybersecurity, PMU.**

I. Introduction

An increasing deployment of PMUs into modern power networks has been witnessed in recent decades. Through the high-resolution DS recorded by PMUs, system-wide states and dynamics can be synchronously captured which enables various DS-based applications (e.g., event identification and localisation, dynamic stability assessment, et al [1]). The DS-based applications are highly dependent on the source information of recorded DS which could be altered maliciously [2]. Therefore, it is imperative to verify the source information of DS data and determine whether the DS is coming from the genuine measurement locations before they become actionable signals.

In general, the mainstream approaches of DS source identification in the literature can be classified into two groups. The first group is model-based methods. By utilising state estimation, these methods take advantage of DS redundancy and network model and parameters to recognise DS anomalies and authenticate source information [3]. However, unavoidable errors may be introduced into network modelling and system identification which could severely affect their performance. The second group generally refers to data-driven alternatives. Without the need of any network models or parameters, these methods identify DS source locations through spatio-temporal DS signature extraction and various artificial intelligence algorithms (e.g., support vector machine [4], neural networks [2], random forest [5] and deep learning methods [6]). However, completely realising the above signatures and achieving accurate identification remains challenging in the research community.

The performance of DS source identification by data-driven methods is highly dependent on the reporting rate of PMUs as DS with a high resolution contains more time-frequency information that could be separated by the artificial intelligence algorithms. However, collecting DS with fine granularity requires a substantial effort to improve the PMU hardware, communication networks and server storage, which may not be feasible for network operators. Therefore, reconstructing quasi high-resolution DS through data interpolation becomes a more promising method in practical power grids. Previous studies realised that multifractality is an intrinsic characteristic that is widely observed in DS [7]. By utilising the multifractality, the quasi high-resolution DS can be generated through traditional multifractal interpolation or Weighted Multifractal Surface Interpolation (WMFSI) [8]. However, the multifractality of DS varies from point to point along with the signal and existing methods use a fixed scale parameter (usually it is the whole data length) to estimate the multifractality which may oversimplify the more complex, multiscale, multifractal structures of DS. In addition, the WMFSI heavily relies on the neighbouring locations of which the DS quality may severely degrade the overall performance. Therefore, the aim of this letter is to propose a cost-effective data-driven method by characterising the multifractality of DS in a wide range of scales simultaneously through MSA-MFDFA and generating quasi high-resolution data through AMFI of DS at individual locations. The proposed method can achieve accurate source identification of DS without the knowledge of power networks and costly upgrading of existing PMUs.

II. Proposed Source Identification Method

The proposed source identification method (illustrated in Fig. 1) consists of three components: (1) MSA-MFDFA is executed on the original DS to determine the scale which presents the strongest multifractality of DS; (2) Based on the identified multifractal scale, AMFI is proposed to dynamically assign the Vertical Scaling Factor (VSF) for generating high-resolution DS data; (3) Informative time-frequency signatures are extracted by continuous wavelet transform from the interpolated high-resolution DS which are used by deep forest algorithm [9] for source identification.

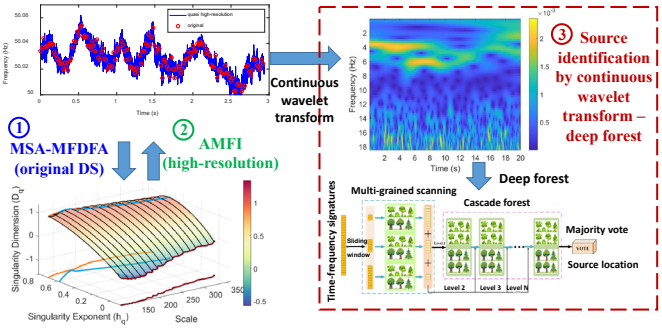

Fig. 1 Proposed data-driven DS source identification method.

Corresponding author: Feifei Bai (e-mail: f.bai@uq.edu.au) is with the School of Information Technology and Electrical Engineering (ITEE), The University of Queensland (UQ), Brisbane, 4072 and is also with School of Engineering and Built Environment at Griffith University, Gold Coast, 4222.

Yi Cui, Hongzhi Yin, Tong Chen and Ryan K. L. Ko are with School of ITEE, UQ , Brisbane, 4072, Australia (e-mail: y.cui3@uq.edu.au, h.yin1@uq.edu.au, tong.chen@uq.edu.au, ryan.ko@uq.edu.au). Yi Cui is also with School of Engineering, University of Southern Queensland, 4300, Australia.

David Dart is with NOJA Power, Brisbane, 4172 (email: DavidD@nojapower.com.au).

Matthew Zillmann is with the Energy Queensland , Brisbane, 4006, Australia (e-mail: matthew.zillmann@energyq.com.au).

A. MSA-MFDFA of DS Measurement

For MSA-MFDFA, the cumulative DS deviation sequence $\theta(i)$ is first derived as (1).

$$\theta(i)=\sum_{l=1}^{i}\left[f(l)-\bar{f}\right],\quad \bar{f}=\frac{1}{N}\sum_{l=1}^{N}f(l),\quad i=1,2,\cdots N \tag{1}$$

where *f(l)* is the DS measurement at time step *l*, $\bar{f}$ denotes the average value of DS, *N* is the total sample number.

Then $\theta(i)$ is scanned using a series of moving windows with certain overlapping (i.e., 5 out of 11 data points) to capture the global smooth trend $\hat{\theta}(i)$. For each window, the original $\theta(i)$ is fitted using a 2-order polynomial function. For the overlapping part in any two adjacent windows, the values are updated as a weighted summation of the original values of two windows using a linear weight as [10]. This makes the trend at the adjacent points of the two windows continuously smooth. Both $\theta(i)$ and $\hat{\theta}(i)$ are further non-overlappingly split into *k* segments with each having a sample number of *T* (i.e., $k = N / T$). For *s-th* segment *(s=1, 2,...,2k),* the root mean square $F^2(T,s)$ of each segment is obtained as (2).

$$F^2(T,s)=\begin{cases}\frac{1}{T}\sum_{i=1}^{T}\left\{\theta\left[(s-1)T+i\right]-\hat{\theta}\left[(s-1)T+i\right]\right\}^2, s=1,2,\cdots k\\ \frac{1}{T}\sum_{i=1}^{T}\left\{\theta\left[N-(s-k)T+i\right]-\hat{\theta}\left[N-(s-k)T+i\right]\right\}^2, s=k+1,\cdots 2k\end{cases} \tag{2}$$

Based on (2), the overall root mean square of all DS segments at the order of q_j is determined as (3).

$$F_{q_j}(T)=\begin{cases}\left\{\frac{1}{2k}\sum_{s=1}^{2k}\ln[F^2(T,s)]^{q_j/2}\right\}^{1/q_j}, q_j\in R, q_j\neq 0\\ \exp\left\{\frac{1}{4k}\sum_{s=1}^{2k}\ln[F^2(T,s)]\right\}, q_j=0\end{cases} \tag{3}$$

The overall root mean square is further non-overlappingly scanned using a series of moving windows with a length of *c* (i.e., scale). For $F_{q_j}(T)$ in each window, if the multifractality exists in the DS, it can be expressed as a power-law function of the segment length within the window as (4) [11].

$$F(q,T)_{\omega_i}\propto T_{\omega_i}^{H(q,c)_{\omega_i}} \tag{4}$$

where ω_i is the *i*-th moving window, *c* is the length of the window (i.e., sample number of the window), $F(q,T)_{\omega_i}$, $H(q,c)_{\omega_i}$, T_{ω_i} are the overall root mean square, Hurst exponent and the segment length of the window. The Hurst exponent can be determined by calculating the slope of $\log\left[F(q,T)_{\omega_i}\right]$ and $\log\left(T_{\omega_i}\right)$ using the least square method. Specifically, the Hurst exponent is above 0.5 when the DS contains multifractal structures. For random signals (such as white noise) without fractal structures, the Hurst exponent is 0.5.

Once the Hurst exponent of each scale is determined, the multifractal spectrum can be calculated as (5)-(6)[11].

$$t(q_j,c)=H(q_j,c)\cdot(q_j-1), j=1,2,\cdots m \tag{5}$$

$$\begin{cases}h(q_j,c)=t(q_j,c)-t(q_{j-1},c)/(q_j-q_{j-1})\\ D(q_j,c)=q_j\cdot h(q_j,c)-t(q_j,c), j=2,\cdots m\end{cases} \tag{6}$$

where $t(q,c)$, $h(q,c)$ and $D(q,c)$ are mass exponent, singularity exponent and singularity dimension at scale *c*.

The strength of the multifractality of the DS data is quantified by the spectrum width which is a plot of $h(q_j,c)$ versus $D(q_j,c)$. Usually, the time series with stronger multifractal structures would have a bell shape spectrum with wider width (as shown in Fig. 3b). In contrast, the spectrum becomes a constant without width when monofractal structures exhibit in the time series.

B. Adaptive Multifractal Interpolation

Based on MSA-MFDFA of DS data, AMFI is proposed to generate quasi high-resolution DS. Unlike the method in [8] which performs multifractal interpolation on the bulk DS data of each location using a constant VSF, the AMFI is piecewise implemented on the short DS fragments using a suitable scale (i.e., sample number) which preserves the strongest multifractality characteristic quantified by the MSA-MFDFA and dynamically adjust the VSF. Assuming the DS dataset $\left\{(t_i,f_i)\,|\,i=0,1,\cdots N'\right\}$, AMFI requires building an iterated function system $\left\{\varphi_n\,|\,n=1,2,\cdots N'\right\}$ as (7) and (8) [12].

$$\varphi_n\begin{bmatrix}t_i\\ f_i\end{bmatrix}=\begin{bmatrix}a_n & 0\\ c_n & d_n\end{bmatrix}\begin{bmatrix}t_i\\ f_i\end{bmatrix}+\begin{bmatrix}e_n\\ g_n\end{bmatrix} \tag{7}$$

$$\varphi_n\begin{bmatrix}t_0\\ f_0\end{bmatrix}=\begin{bmatrix}t_{n-1}\\ f_{n-1}\end{bmatrix},\varphi_n\begin{bmatrix}t_{N'}\\ f_{N'}\end{bmatrix}=\begin{bmatrix}t_n\\ f_n\end{bmatrix} \tag{8}$$

where f_i and t_i denote the DS data at time instance *i*, *N'+1* denotes the total DS data points number, a_n, c_n, d_n ,e_n, g_n are free parameters of φ_n which are determined by (9).

$$\begin{cases}a_n=L^{-1}\left(t_n-t_{n-1}\right)\\ g_n=L^{-1}\left[t_{N'}f_{n-1}-t_0f_n-d_n\left(t_{N'}f_0-t_0f_{N'}\right)\right]\\ e_n=L^{-1}\left(t_{N'}t_{n-1}-t_0t_n\right)\\ c_n=L^{-1}\left[f_n-f_{n-1}-d_n\left(f_{N'}-f_0\right)\right]\end{cases} \tag{9}$$

where $L=t_{N'}-t_0$, d_n is the VSF and it is subjected to $|d_n|<1$ so that the iterated function system converges after multiple iterations. As the most significant parameter in AMFI, d_n determines how to map the DS variation over the whole measurement time into two interpolation points. Usually, d_n is a constant and it is empirically determined [8]. The advantage of AMFI is that it automatically determines d_n by considering the overall trend of the DS signal which can better map the entire DS variation into each interpolation interval. In AMFI, the original DS data is decomposed into different intrinsic mode functions by empirical mode decomposition and the intrinsic mode function at the lowest frequency is selected to represent the general trend of the original DS data. Finally, d_i of each interpolation interval is calculated as (10) and (11).

$$d_i=r_i/r,\qquad i=1,2,\cdots N' \tag{10}$$

$$r_i=f_i-f_i',\qquad r=\max\left\{|r_i|\,|\,i=1,2,\cdots N'\right\}+\varepsilon \tag{11}$$

where f_i' is the overall trend at time *i*, ε is infinitesimal. In this way, a large d_i is assigned if the difference between the DS and the general trend is large (or vice versa). It also meets the requirement of $|d_i|<1$.

C. Signature Extraction and Source Identification

The DS reconstructed by AMFI is subsequently processed by DS detrending and continuous wavelet transform as described in [6] to extract informative time-frequency signatures which quantify the non-stationary and nonlinearity characteristics of DS in both the time and frequency domains. The extracted signatures are fed into the deep forest algorithm with a multi-grain scanning and cascade forest structure [9] for identifying the source locations. Compared with the convolutional neural network, the deep forest has superior or comparable

performance with an easier configuration than the convolutional neural network.

III. Case Studies and Performance Validation

A. Description of the Experimental Data

The proposed method is validated using a DS dataset which contains 3-months data of seven intra-state locations (detailed in [8]) of a distribution network in Queensland. The DS data is collected at multiple 11 kV feeders which significantly improve the situational awareness capability of distribution network operators for better network management. All measurement locations reside within 90km which are significantly closer than those used by most existing studies. The sampling time of the DS is 20ms and no extra noise is added to the DS data for the experiment. For each location, 1000 segments are used to construct an experimental dataset and each segment has 20-seconds DS measurements. 80% of segments are randomly selected for the deep forest training while the rest 20% of segments are used for testing. The identification accuracy of the testing segments is calculated for evaluating the performance of the algorithm.

B. Results of MSA-MFDFA and AMFI

Fig. 2 depicts the Hurst exponent after MSA-MFDFA of DS at one measurement location. It shows that the Hurst exponent is not a constant but varies at different scales. By examining the variation of the Hurst exponent, two regions are identified which are marked as numbers in circles in Fig. 2a. In region 1, an increasing trend is presented in the Hurst exponent when the scale increases and the maximum Hurst exponent (i.e., 0.75) is located at the scale of 147. The difference between the maximum and minimum Hurst exponent also attains the peak (i.e., 0.55) at this scale (Fig. 2b). Beyond this scale, the general Hurst exponent (i.e., order $q=2$) in Fig. 2b gradually decreases. This indicates the vanishment of the multifractality when a larger scale of the data is used to examine the multifractality.

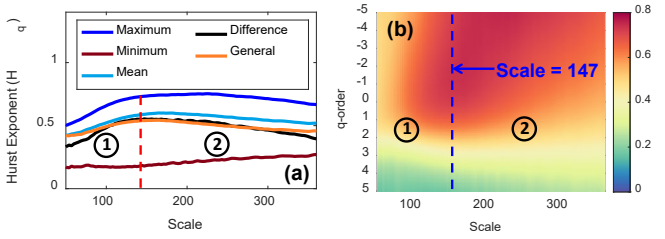

Fig. 2 (a) Maximum, minimum, mean Hurst exponent, the difference between the maximum and minimum Hurst exponent and general Hurst exponent at multiple scales and (b) Hurst exponent obtained by MSA-MFDFA.

After Hurst exponent calculation, the multifractal spectrum of DS at different scales is computed as Fig. 3a. It is clear that MSA-MFDFA is capable of quantifying the multifractality of DS at different granularities as the multifractal spectrum at different scales has different shapes, positions and variation ranges. Fig. 3b depicts the multifractal spectrum at scale 147 where $h_q(c)_{min}$, $h_q(c)_{max}$ and $h_q(c)_0$ represent the minimum, maximum singularity exponents and the singularity exponent with maximum singularity dimension. The width of the half-left spectrum (i.e., W_L) and the half-right spectrum (i.e., W_R) is added to calculate the total width $W(c)$ of the spectrum. By evaluating the spectrum width over all scales, the largest spectrum width of 0.81 appears at the scale of 147 (shown in Fig. 3c), indicating the strongest multifractal strength at this scale. Therefore, the proposed AMFI will be implemented for every 147 samples to construct the quasi high-resolution data.

Fig. 4 compares the performance of two methods (i.e., WMFSI and proposed AMFI) by interpolating the downsampled DS (i.e., 10Hz) with a data length of 2.94s (147×0.02s) back to the original reporting rate (i.e., 50Hz). The interpolation errors and corresponding error distribution are shown in Fig. 4. It is clear that the proposed method outperforms the WMFSI as the error distribution of the proposed method shows a narrow width centred near zero indicating that most interpolation errors have small magnitudes. By calculating the root mean square error between the measured DS and interpolated DS, the proposed AMFI has a much smaller error (i.e., 2.6mHz) compared with WMFSI (i.e., 4.9mHz) due to its dynamic adjustment of the VSF during multifractal interpolation.

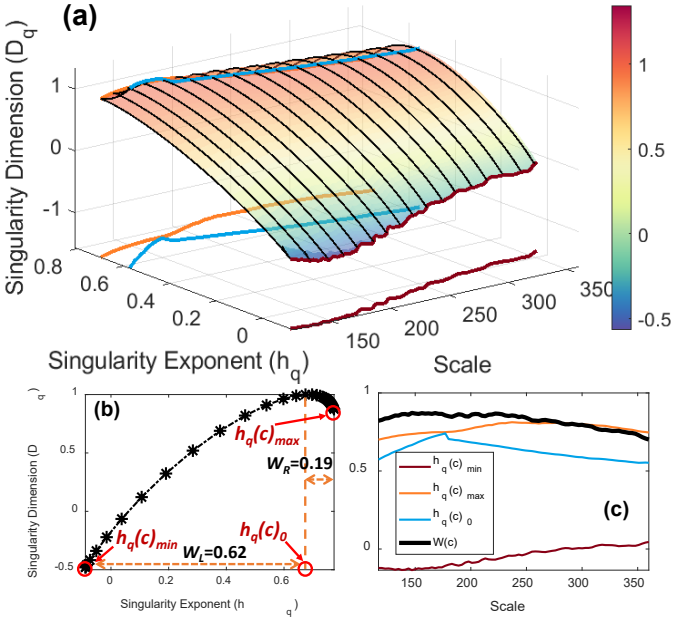

Fig. 3 (a) A multifractal spectrum surface, (b) multifractal spectrum at the scale 147 and (c) minimum, maximum h_q, the h_q with maximum D_q and spectrum width at multiple scales.

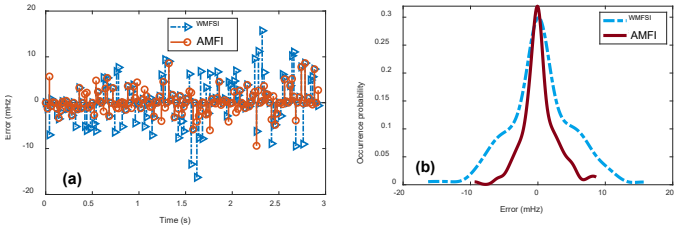

Fig. 4 Comparison of (a) interpolation error and (b) error distribution by WMFSI and AMFI.

C. Performance Comparison of Source Identification

TABLE I compares the source identification performance of the proposed method with the other six methods in the existing literature in terms of the overall identification accuracy and testing time required for each sample. It should be noted that the proposed method and [8] use quasi high-resolution DS with a reporting rate of 1.25kHz for source identification while other methods use the original DS with a 50Hz reporting rate. It is clear that although Ref [4] takes the least time to perform the source identification, it is hard to recognise the source locations of DS (i.e., overall accuracy is only 53%) as it only uses correlation coefficients of DS segments as input features for the artificial intelligence algorithms. By incorporating more informative spatio-temporal signatures (e.g., mean value, variance of DS segments and signatures extracted by wavelet-based signal decomposition), Ref [13] shows a certain improvement in the accuracy which is 68%. The performance of Refs [2], [5] and [6] is comparable as they all extract the spatio-temporal signatures from 50Hz reporting rate DS. Both [8] and

the proposed method achieve a reasonable identification accuracy above 90% where the proposed method outperforms [8] in terms of higher identification accuracy and less computation time. The superior identification accuracy of the proposed method is mainly due to the better reconstruction of the quasi high-resolution DS by taking advantage of the proper multifractal scale indicated by MSA-MFDFA and the dynamic VSF determined by AMFI. The faster computation of the proposed method is achieved since the proposed method does not need to examine the cross-correlation of each pair of PMU locations for DS interpolation. Besides, compared with [8], the proposed method also has better tolerance of unsatisfied DS data quality of neighbouring locations which makes it more applicable in practical power grids.

TABLE I
COMPARISON OF IDENTIFICATION PERFORMANCE OF DIFFERENT ALGORITHMS

Reference	Overall accuracy (%)	Testing time per sample (ms)
[4]	53	0.3
[2]	85	825
[5]	80	1.8
[6]	88	97
[8]	93	90
[13]	68	1.2
Proposed	95.08	45

IV. CONCLUSION

This letter proposes an intelligent data-driven approach for accurate DS source identification. With the help of MSA-MFDFA, it first recognises the scale of DS which possesses the most significant multifractal structures. Then it adaptively adjusts the VSF during the multifractal interpolation for constructing the high-resolution DS data. Essentially, this approach is performed by exploring distinguish time-frequency signatures of the high-resolution DS data and integrating them with the advanced deep forest algorithm. Numerical experiment results with actual DS collected from Queensland distribution networks comprehensively demonstrate its good performance. Compared with other representative methods, it is capable of performing reliable DS source identification in a more cost-effective manner. Given that the frequency measurements of the distribution system and transmission system share the same characteristics and multiple formats of malicious data integrity attacks may happen to synchrophasor data, the proposed source identification method also has the potential to protect part of the synchrophasor data of transmission systems from sophisticated data spoofing attacks.

ACKNOWLEDGMENT

This work was supported by the UQ Cyber Transdisciplinary Research Seed Funding.

REFERENCES

[1] NASPI, "Synchronized Measurements and their Applications in Distribution Systems: An Update Draft,", 2020.

[2] S. Liu, S. You, H. Yin, Z. Lin, Y. Liu, W. Yao and L. Sundaresh, "Model-Free Data Authentication for Cyber Security in Power Systems," *IEEE Trans. Smart Grid*, vol.11, Issue 5, pp. 4565-4568, 2020.

[3] A. S. Musleh, G. Chen and Z. Y. Dong, "A Survey On the Detection Algorithms for False Data Injection Attacks in Smart Grids," *IEEE Trans. Smart Grid*, vol.11, Issue 3, pp. 2218-2234, 2020.

[4] J. Landford, R. Meier, R. Barella, X. Zhao and E. Cotillasanchez, "Fast Sequence Component Analysis for Attack Detection in Synchrophasor Networks," *International Conference on Smart Cities and Green ICT Systems*, 2016, Rome, ITALY, APR 23-25, pp. 1-8.

[5] Y. Cui, F. Bai, Y. Liu and Y. Liu, "A Measurement Source Authentication Methodology for Power System Cyber Security Enhancement," *IEEE Trans. Smart Grid*, vol.9, Issue 4, pp. 3914-3916, 2018.

[6] W. Qiu, K. Sun, W. Yao, S. You, H. Yin, X. Ma and Y. Liu, "Time-Frequency Based Cyber Security Defense of Wide-Area Control System for Fast Frequency Reserve," *Int. J. Electr. Power Energy Syst.*, vol.132, pp. 107151, 2021.

[7] L. Shalalfeh, P. Bogdan and E. Jonckheere, "Modeling of PMU Data Using ARFIMA Models," *Clemson University Power Systems Conference (PSC)*, 2018, pp. 1-6.

[8] Y. Cui, F. Bai, R. Yan, T. Saha, M. Mosadeghy, H. Yin, R. K. L. Ko and Y. Liu, "Multifractal Characterization of Distribution Synchrophasors for Cybersecurity Defense of Smart Grids," *IEEE Trans. Smart Grid*, vol.13, Issue 2, pp. 1658-1661, 2022.

[9] Z. H. Zhou and J. Feng, "Deep Forest: Towards an Alternative to Deep Neural Networks," *International Joint Conference on Artificial Intelligence, 19-25 August*, 2017, Melbourne, Australia, pp. 1-7.

[10] G. Han, F. Zhou and H. Jiang, "Multiscale Adaptive Multifractal Analysis and its Applications," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol.31, Issue 2, pp. 23115, 2021.

[11] E. A. F. Ihlen, "Introduction to Multifractal Detrended Fluctuation Analysis in Matlab," *Front. Physiol.*, vol.3, pp. 1-18, 2012.

[12] M. F. Barnsley, *Fractals Everywhere*, Dover Publications, 2012.

[13] H. Zhou, H. Duanmu, J. Li, Y. Ma and J. Shi, "Geographic Location Estimation From ENF Signals with High Accuracy," *IEEE Signal Processing Cup*, 2016, pp. 1-8.