Multimedia Tools and Applications (2023) 82:24101–24125
https://doi.org/10.1007/s11042-022-14180-5

Adaptive Weiner filtering with AR-GWO based optimized fuzzy wavelet neural network for enhanced speech enhancement

Amarendra Jadda[1] **· Inty Santi Prabha**[1]

Received: 6 September 2021 / Revised: 17 June 2022 / Accepted: 27 October 2022 /
Published online: 9 December 2022

Abstract

Speech signal enhancement is a subject of study in which a large number of researchers are working to improve the quality and perceptibility of speech signals. In the existing Kalman Filter method, the short-time magnitude or power spectrum due to random variations of noise was a serious problem and the signal-to-noise ratio was very low. This issue severely reduced the perceived qualityand intelligibility of enhanced speech. Thus, this paper intent to develop an improved speech enhancement model and it includes“training phase and testing phase”. In the training phase, the input noise corrupted signal is initially fed as input to both STFT-based noise estimation and NMF-based spectrum estimation forestimating the noise spectrum and signal spectrum, respectively. The obtained noise spectrum and the signal spectrum are fed as input to the Wiener filter and these filtered signals are subjected to Empirical Mean Decomposition (EMD).Since, tuning factor $\boldsymbol{\eta}$ plays a key role in Wiener filter, it has to be determined for each signal and from the denoised signal the bark frequency is evaluated. The computed bark frequency is fed as input to the learning algorithm referred as Fuzzy Wavelet Neural Network (FW-NN)for detecting the suited tuning factor η for the entire input signal in Wiener filter.An Adaptive Randomized Grey Wolf Optimization (AR-GWO) is proposed for proper tuning of the tuning factor η referred as tuned tuning factor (η^{tuned}). The proposed AR-GWO is the improved version of standard Grey wolf optimization (GWO). In the testing phase, the training is accomplished initially and from which the tuning factor is gathered for each of the relevant input signal. Then, the properly tuned tuning factor (η^{tuned}) from FW-NN is fed as input to EMD via adaptive wiener filter for decomposing the spectral signal and

✉ Amarendra Jadda
amarendrajadda0@gmail.com

[1] Department of Electronics and Communication Engineering, Jawaharlal Nehru Technological University, Kakinada, Kakinada 533 003 Andhra Pradesh, India

the output of EMD is denoised enhanced speech signal. At last, the performance of the adopted approach is evaluated to the existing approaches in terms of various metrics. In particular, the computation time of the adopted AR-GWO model is 34.07%, 43.57%, 28.86%, 38.88%, and 16.03% better than the existing GA, ABC, PSO, FF, and GWO approaches respectively.

Keywords STFT-based noise estimation · NMF-based spectrum estimation · EMD · Weiner filter · FW-NN

Abbreviations

ABC	Artificial Bee Colony optimization
CSED	Cumulative Squared Euclidean Distance
DNN	Deep Neural Network
ESTOI	Extended STOI
FF	Firefly optimization
FRBM	Fuzzy Restricted Boltzmann Machines
GA	Genetic Algorithm
GWO	Grey Wolf optimization
HMM	Hidden Markov Model
IFD	Instantaneous Frequency Deviation
IMF	Intrinsic Mode Functions
IRM	Ideal Ratio Mask
JT-FS	Joint Time-Frequency Segmentation Algorithm
KCF	Kalman Filter-Based
LT-FD	linear time-frequency distribution
MMSE	Minimum Mean Square Error
MSE	Mean Square Error
NMF	Nonnegative Matrix Factorization
P-ASE	Phase-Aware Speech Enhancement
PDF	Power Spectral Density
PESQ	Perceptual Evaluation Of Speech Quality
P-SJL	Phase-Sensitive Joint Learning Algorithm
PSD	Power Spectral Density
PSM	Phase-Sensitive Mask
PSO	Particle Swarm Optimization
PWFT	perceptual wavelet filter bank
RMSE	Root-Mean-Square Error
SDR	Source To- Distortion Ratio
S-MSE	Single-Microphone Speech Enhancement
SNR	Signal-To-Noise Ratio
SSD	Supervised Speech Denoising
STFT	Short-Time Fourier Transform
STOI	Extended STOI
T-F	Time-Frequency
VoIP	Voice over Internet Protocol

1 Introduction

In the present era, speech enhancement plays a major role in the field of speech processing as it is related to the speaking as well as listening skills. In general, speech enhancement is employed with a desire of processing the noisy speech signals, thereby enhancing human perception [3, 38, 41]. Generally, the quality of the speech is related to the attributes of the speaker, like the naturalness and speaker recognizability, whereas the intelligibility of the speech is related to the meaning or information content that is hidden behind the words [45] [27]. Speech signals are utilized in many purposes and in recent times, COVID-19 [28, 36] has been detected from speech signals. Hence, it is vivid that, the ability to communicate (speak and listen) diminishes in the noisy environment.

The speech enhancement is performed with the intention of reducing the impact of the communication problem [22]. Most of the research proved that, it is a complex task to reduce the noise of the signal without distorting speech and this is the major reason behind the non-availability of an ideal enhancement systems [4, 30]. Beyond this, efforts to enhance the "higher quality and/or intelligibility of noisy speech" will definitely end up with a mass increment in the performance of the speech signal and hence it can be employed in the fields of "speech coding/compression and speech recognition, hearing aids, voice communication systems and so on" [21, 35]. Further, the goal behind each of the speech recognition might be different and they are application based, such as diminishing the listener fatigue, boosting the overall speech quality, enhancing the intelligibility and improving the performance of the voice communication device, etc. But, the major benchmark behind all the research is to diminish the noise level and to enhance the quality as well asthe intelligibility of the signal. Hence,"speech enhancement is necessary to avoid the degradation of speech quality and to overcome the limitations of human auditory systems" [2, 43].

A vast amount of automatic speech processing systems are playing a major role in human life, like the "mobile communication, speech and speaker recognition, hearing impaired and numerous other applications". Moreover, the quality and intelligibility of speech areof utmost importance with the intention of enhancing the accuracy of information exchange [31]. Beyond this, in the controlled environment, human as well as automatic speech communications are found to be much more effective [5, 47]. The Spectral Subtraction algorithm suffers from the problem of restoration in the basis parameters of the speech like the power spectrum or the magnitude spectrum and here only the additive noise available in the signal can be removed [1]. Then, in the Sub-space analysis algorithm, there was a difficulty in enhancing the noise spectrum and updating the noise spectrum from period to period was a complex task. [46]. Thus, in order to override these entire problems, there is a necessity to have an optimal speech enhancement method.

Nowadays, literature works have come up with several techniques for speech enhancement as relates to speaking as well as listening skills. Tantibundhit et al. [44] proposed JT-FS with the desire of decomposing the speech signal into "transient" as well as "non-transient components" only the basis of the wavelet packets. Lee et al. [30] proffered P-SJL with the intention of enhancing single-channel speech. The phase-related information of the speech signal was represented using PSM which was similar to the T-F mask. Furthermore, the P-ASEalgorithm [48] was formulated on the basis of DNN. Shao and Chang [40] developed a framework of wavelet-based techniques with the intention of enhancing the performance of automatic speech recognition by eliminating the background noise. AKCF algorithm was introduced in [13]with the aim of enhancing speech. The noise as well as the speech parameters was estimated using the Estimate-Maximize (EM) method. Mohammadiha et al. [34] proposed SSD algorithms on the basis of NMF. In addition to this, the Bayesian Formulation of NMF (BNMF) was used for generating the novel speech enhancement method. Additionally, Chazan et al. [6] proposed the S-MSE algorithm in order to enhance the speech signal. Samuiet al. [39] proposed time-frequency masking in the basis of DNN with the intention of enhancing the speech signal and here the pre-training of the signal was accomplished using FRBM. Moreover, the advantages and challenges of the few works are listed in Table 1. These challenges have kept the main stand for motivating and accomplishing the new speech enhancement model.

In addition, many optimization algorithms have been introduced recently [8–10, 25] and utilized in many fields for better outcomes [7]. In this research work, a modified version of a popular meta-heuristic algorithm is employed. The major contributions of this research are listed below:

- In this research work, STFT-based noise estimation and NMF-based spectrum estimation are utilized for the estimation of the noise spectrum and signal spectrum of the noisy signal.

Table 1 Features and Challenges of of the state-of-art Speech Enhancement models

Author [citation]	Adopted methodology	features	Challenges
Tantibundhit et al. [44]	JT-FS Algorithm	Enhanced the word recognition rate. Increased speech intelligibility in background Noise.	Low SNR No consideration on the constant Or slowly varying frequency information
Lee et al. [30]	P-SJL algorithm	High SNR	Not Applicable to magnitude spectra domain.
		Had consideration on the dynamic range of speech and noise spectral Features.	Do not solve the phase mismatch problem.
Zheng andZhang [48]	P-ASE Algorithm	Solved phase wrapping problem Enhanced speech quality and intelligibility.	Low SNR BW is high
Shao and Chang [40]	HMM	Degraded the residual noise energy. High speech recognition accuracy.	Sensitivity of signal is high SNR is low
Gannot et al. [13]	KF algorithm	Achieves higher SNR and quality of speech	Difficult to construct reliable VAD (Voice Activity Detector)
		Applicable to noise sources with fast changing Spectrum	
Mohammadiha et al. [34]	SSD algorithm	Noise classification and speech enhancement were carried out at the same time Solved the mismatch problem	Highly affected by nonnegative decomposition problem
Chazan et al. [6]	DNN	Preserves the continuity of speech.	Do not address the problem of noisy speech received by a single microphone
		Reduces the artifacts	High distortions in the enhanced speech
Samui *et al.* [39]	Fuzzy based DNN	Solves the problem of local optima and total reconstruction error.	Not suitable for real-world noise.
		Provides good performance under noisy reverberant conditions.	Training of the signal is complex

- To minimize the error, a Wiener filter is employed and the tuning factor $\eta$of Wiener filter is obtained for different signals.
- Introducing a Fuzzy Wavelet Neural Network (FW-NN) for detecting the suited tuning factor η for the entire input signal in Wiener filter.
- Proposed an Adaptive Randomized Grey Wolf Optimization (AR-GWO) for proper tuning of the tuning factorηreferred as tuned tuning factor(η^{tuned}). The proposed AR-GWO algorithm is an improved version of the traditional GWO algorithm.

The rest of the paper is organized as below: Section 2 portrays the proposed architecture of the speech enhancement model. Section 3 depicts the processed steps for enhanced speech enhancement. The results and discussions are exhibited in Section 4, and Section 5 concludes the paper.

2 Proposed architecture of speech enhancement model

2.1 Architectural representation

Figure 1 demonstrates the architecture of the proposed speech enhancement model in which the overall process takes place in "two major phases (i) training phase (ii) testing phase". In the training phase, initially, the noise corrupted signal is fed as input to STFT-based noise estimation as well as NMF-based spectrum estimation, for estimating the noise spectrum and signal spectrum, respectively. The obtained spectrum (noise and signal) are given as input to the Wiener filter. These, filtered signals are subjected to EMD, from which the denoised signal can be obtained. Since, tuning factor η plays a key role in Wiener filter, it has to be determined for each signals, and is trained in FW-NN. Then, from the denoised signal the bark frequency is evaluated. The computed bark frequency is fed as input to the learning algorithm referred as FW-NNfor detecting the suited tuning factorη for the entire input signal in Weiner filter. The AR-GWO is employed for proper tuning of the tuning factorη. Moreover, in the testing phase of a signal, the training is accomplished initially, from which the tuning factor$\eta$is gathered for the corresponding input signal. Then, the properly tuned $\eta$from FW-NNis fed as input to EMD via adaptive Wiener filter for decomposing the spectral signal and the output of EMD is denoised signal.

Consider the clear signal as $T(n)$, when the noise Wgets corrupted into it, and the signal becomes noisy signal $\overline{T}(n)$. This noisy signal is fed as input to the STFT-based noise estimation and NMF-based spectrum estimation, from which the noise spectrum W^T and signal spectrum $\overline{W}^T$are obtained. The obtained noise and signal spectrum are subjected to filtration using Wiener filtering process; at the end of filtration the filtered signal $\overline{T}_u(n)$ is generated. Then, $\overline{T}_u(n)$ is decomposed using EMD as a result of this, the bark frequency $c'(u')$ is obtained. This bark frequency is utilized to train FW-NN classifier. From the spectrum W^T and $\overline{W}^T$ as well as from FW-NN, 'tuned η' referred as η^{tuned} is acquired for all the inputs signals with AR-GWO. In the testing process, the tuned η^{tuned} is acquired for the corresponding signal with the aid of the AR-GWO; this η^{opt} is fed as input to the adaptive Wiener filtering process with the intention of tuning the input signal $\overline{T}(n)$. The outcomes of the adaptive Wiener filter are the filtered signal $\overline{\overline{T_u(n)}}$. Again, $\overline{\overline{T_u(n)}}$ is decomposed using EMD and the result is the enhanced denoised signal $\overline{\overline{T_o(n)}}$.

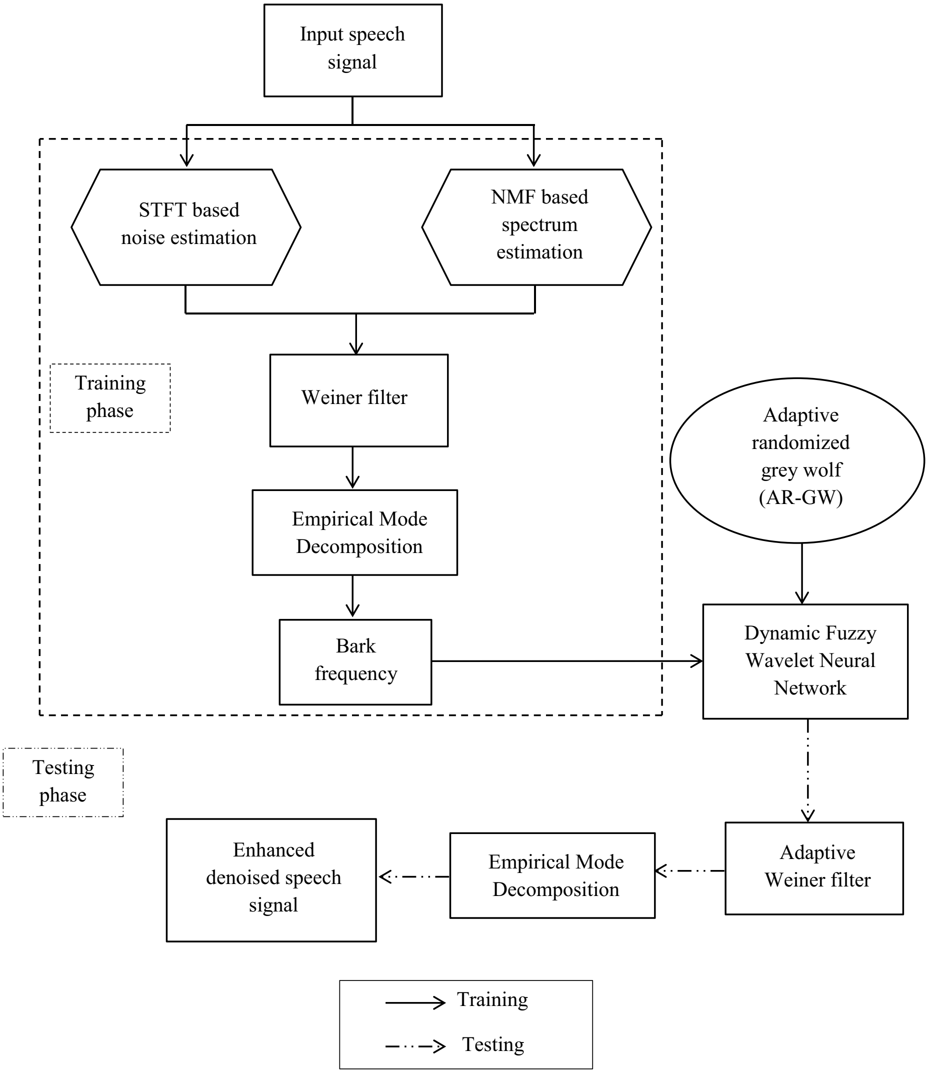

Fig. 1 Proposed intelligence architecture for speech enhancement model

3 Processed steps for enhanced speech enhancement

3.1 STFT-based noise estimation

The noise power spectral density estimatorisbased onminimum statistics to track the minima from the noisy signal [26]. The STFT coefficient of the frame γ is depicted as $T(\gamma, p)$ and their mathematical formula is exhibited in Eq. (1) [14].

$$T(\gamma,p) = \tau(\gamma,p)T(\gamma-1,p) + (1-\tau(\gamma,p))\left|T\left(\gamma,p\right)\right|^2 \tag{1}$$

Here, the frequency bin is manifested as p. The frequency and time-dependent smoothing parameters is portrayed as $\tau(\gamma, p)$. With the intention of observing the mean power, the bias compensation factor is employed. The variance estimator of the smoothened PSD is represented as $var\{T(\gamma, p)\}$ and the function corresponding to the length of minimum search interval is defined by the bias compensation factor $K_{\min}$. The variance estimator relating the smoothened PSD is indicated as$var\{T(\gamma, p)\}$ and this assist in evaluating the variance of $T(\gamma, p)$ by fixing the length of the search interval in the algorithm. Eq. (2) depicts the mathematical formula for evaluating the variance estimator at the frame γ relating the frequency bin p. In Eq. (2), the mean smoothened periodograms is represented as $\overline{T}(\gamma,p)$, and $\overline{T^2}(\alpha,b)$ indicates the first-order recursive average of smoothened periodograms [14].

$$var\{T(\gamma,p)\} = \overline{T^2}(\gamma,p) - \overline{T}^2(\gamma,p) \tag{2}$$

This paper deal with STFT-based noise estimation and the graphical representation of the power spectrum corresponding to the noise estimated by FFT as well as STFT is exhibited in Fig. 2. The power spectrum varies by the magnitude of the frequency component. Moreover, in determining the phase content of the signal and varying sine wave frequency that alter over time are predicted using STFT. In general, the time signals which are larger in size are sub-divided into smaller equal size signals and to each of the segments the Fourier transform is employed. In addition, in the filtering process, STFT can also be interpreted. The estimation strategy is satisfied by two major properties viz. magnitude based shift invariance property and LT-FD properties. The noise spectrum W^T is obtained as the resultant.

3.2 NMF-based Spectrum estimation

In the time-frequency (γ, p) domain, the voicing of the noisy signal $\overline{T}(n)$ takes place via STFT as per Eq. (3), to enhance the speech signal [42]. In Eq. (3), the STFT of the clear speech $T(p, \gamma)$, the STFT of the noisy speech $\overline{T}(p,\gamma)$ and the STFT of the noise signal $W(p, \gamma)$ are used in p^{th} frequency bin of γ frame. The mathematical formula for "noisy speech's magnitude spectrum" approximation, which is most commonly, utilized assumption for NMF-based processing of speech and audio signal, is show in Eq. (4) [14].

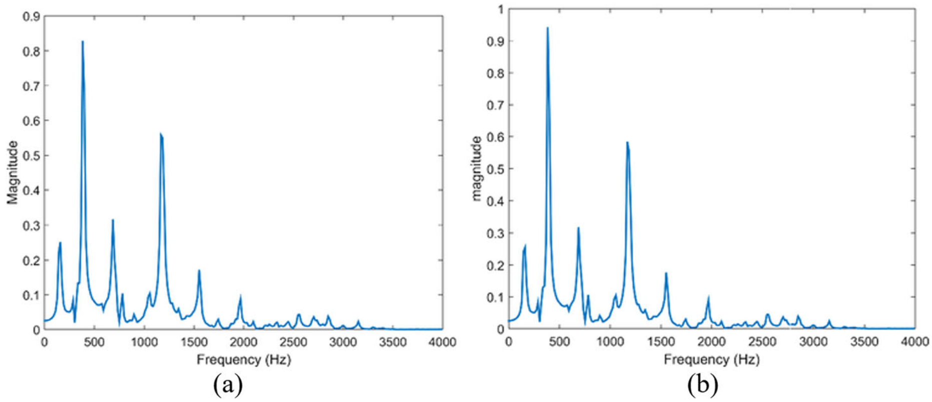

Fig. 2 Noise power spectrum (**a**) estimated by FFT (**b**) estimated by STFT - minimum statistics

$$\overline{T}(p,\gamma) = T(p,\gamma) + W(p,\gamma) \tag{3}$$

$$|\overline{T}(p,\gamma)| = |T(p,\gamma)| + |W(p,\gamma)| \tag{4}$$

The magnitude spectrum matrices of the varied signal are indicated as per Eq. (5) and magnitude spectral value corresponding to γ frame for the p^{th} bin is depicted as $j_{p,\gamma}$. The count of the frequency bins is represented as H and the time frames are indicated as I.

$$J = \left[j_{p,\gamma}\right] \in N_{+}^{H \times I} \tag{5}$$

For the training data $J_T \in N_{+}^{H \times I_T}$ as well as $J_W \in N_{+}^{H \times I_W}$, the Eq. (5) is employed separately in the training stage and the outcome of these data is the basis matrices in terms of clear speech $F_T = \left[r_{Hl}^{T}\right] \in N_{+}^{H \times L_T}$ and noise $F_W = \left[r_{Hl}^{W}\right] \in N_{+}^{H \times L_W}$, respectively. The count of base vectors is indicated as L. In Eq. (6) ζ represents a $H \times I$ matrix, whose entities is equal to one and the transpose of the matrix, is represented as T'. In addition, the basis matrices are fixed in the enhancement stage as $F_T = [F_T \ F_W] \in N_{+}^{H \times (L_T + L_W)}$.The activation matrix $E_{\widehat{T}} = \left[E_T^{T'} E_W^{T'}\right]^{T'} \in N_{+}^{(L_T + L_W) \times I_{\widehat{T}}}$ corresponding to the noisy speech is estimated from $J_{\widehat{T}} \in N_{+}^{H \times I_{\widehat{T}}}$ by means of employing the NMF activation update. Further, with the assistance got from the Wiener Filter (WF), the clear speech spectrum is evaluated from the speech signal only after obtaining the activation matrix as per Eq. (7). The estimated PSD matrices corresponding of the clear speech is manifested as $V'_T = [V'_T(p, \gamma)]$ and the evaluated PSD matrices corresponding to the noisy signal is represented as $V'_W = [V'_W(p,\gamma)] \in N_{+}^{H \times I_{\widehat{T}}}$ in Eq. (7). Further, as per Eqs. (8) and (9) the next solution is obtained via the temporal smoothing of the period grams. The temporal smoothing factor of speech ω_T and noise ω_W is shown in Eqs. (8) and (9), respectively.

$$\begin{gathered} F \leftarrow F \otimes \frac{(J/F.E)E}{\zeta E}, \\ E \leftarrow E \otimes \frac{F(J/FE)}{F^{T'} \zeta} \end{gathered} \tag{6}$$

$$Q = \frac{V'_T}{V'_T + V'_W} \otimes \widehat{T} \tag{7}$$

$$V'_T(p,\gamma) = \rho_T V'_T(p,\gamma - 1) + (1 - \rho_T)\left(\left[F_T E_T\right]_{p\gamma}\right)^2 \tag{8}$$

$$V'_W(p,\gamma) = \rho_W V'_W(p,\gamma - 1) + (1 - \rho_W)\left(\left[F_W E_W\right]_{p\gamma}\right)^2 \tag{9}$$

The signal spectrum $\overline{W}^{T}$ is obtained as the outcomes.

The obtained noise spectrum W^T and signal spectrum $\overline{W}^{T}$ are subjected to filtration using Wiener filtering process.

3.3 WienerFilter

In the signal enhancement technique, the Wiener filter has been employed in large scale [15]. The Wiener filter works on the principle of producing an estimate of the clean signal from the corrupted noise signal. The estimation is accomplished by minimizing MSE in between the desired signal and additive noise corrupted signal. The filter transfer functionis shown in Eq. (10) and it gives the solution to this optimization problem in the frequency domain. This equation is generated by considering the signal spectrum $\overline{W}^{T}$ and the noise spectrum W^T as uncorrelated and stationary signals. The power spectral density of $\overline{W}^{T}$ is represented as $G_T(\omega)$ and the power spectral density of W^T is depicted as $G_W(\omega)$. The mathematical formula for SNR is exhibited in Eq. (11) and the SNR formula can be incorporated in the filter transfer function as per Eq. (12). The estimated signal magnitude spectrum is indicated as $\widehat{G}_W(\omega)$.

$$F(\omega) = \frac{G_T(\omega)}{G_T(\omega) + G_W(\omega)} \tag{10}$$

$$SNR = \frac{G_T(\omega)}{\widehat{G}_W(\omega)} \tag{11}$$

$$F(\omega) = \left[1 + \frac{1}{SNR}\right]^{-1} \tag{12}$$

At the end of filtration the filtered signal $\overline{T}_u(n)$ is generated.

The Wiener filteroften fails at all the frequencies due to the drawback of fixed frequency response and requirement of estimating the clean signal and noise signal's power spectral density prior to filtering.

3.4 Empirical model decomposition

EMD [16] was introduced by Huang as an adaptive technique in which small number of orthogonal empirical modes referred as IMF were added to represent the complex data. The symmetric envelope is present in each of the mode in terms of local maximums and minimums. Thus at all locations of the envelope, mean is zero and in the underlying signal, there is no requirement of linearity or time invariance. Further, by the process of shifting, the riding waves are eliminated. The shifting process of EMD algorithm can be depicted as shown below. Two main properties are obeyed by EMD during the splitting of $\overline{T}_u(n)$ into its IMF components. They are (a) In between two subsequent zero crossing, the IMF has only one extremum and (b) Mean value of IMF is zero.

The data set $\overline{T}_u(n)$ is decomposed into IMFs $y_e(n)$ and residue $q(n)$. The mathematical formula corresponding to this decomposition is described in Eq. (13).

$$y(n) = \sum_e y_e(n) + y(n) \tag{13}$$

Furthermore, the detailed steps of EMD are given below.

- At first, initialization is processed i.e., $d := 1, q_0(n) = y(n)$
- As per the following steps, d^{th} IMF is extracted

Set $k_0(n) := q_{d-1}(n), m := 1$ and local maxima and minima of whole $k_{m-1}(n)$ are identified. Then the envelope $UB_{m-1}(n)$ for $k_{m-1}(n)$ defined by the maxima and $LB_{m-1}(n)$ by the minima using the cubic splines interpolation.For both the envelopes belonging to $k_{m-1}(n)$, the mean $z_{m-1}(n)$ is determined as $z_{m-1}(n) = \frac{1}{2}(UB_{m-1}(n) - LB_{m-1}(n))$. This running mean is referred as low frequency local trend. Further, via the process of shifting, the evaluation of high-frequency local detail takes place.

Further, the m^{th} component is formed as $k_m(n) := k_{m-1}(n) - z_{m-1}(n)$. In case if $k_m(n)$ is not found to be accordance with whole IMF criteria, then the process of shifting is continued by increasing $mm + 1$. In case, if all IMF criteria is satisfied by $k_m(n)$, then set $y_d(n) := k_m(n)$ and $q_d(n) := q_{d-1}(n) - y_d(n)$.

- The shifting process can be stopped, if $q_d(n)$ represents a residuum and if not, then continue the shifting process by increasing d, $d + 1$ and again begin the process.

Further, EMD algorithm achieves the completeness of the decomposition process automatically as $y(n) = \sum_{d=1}^{v} y_d + q$ and this represents an identity. The locally orthogonal IMFs are generated by EMD algorithm and lacks to guarantee the global orthogonality, since identical frequencies might be utilized by neighboring IMFs at different time points. As a result of this, the bark frequency $c'(u')$ is obtained. This bark frequency is utilized to train FW-NN classifier.

3.5 Fuzzy wavelet neural network (FW-NN) classifier

Classification is the most frequently used prediction type [37]. Generally, the wavelet functions are combined with neural nets to provide better results [17–19]. In this work, a FW-NN model is employed and it is combination of fuzzy logic concepts and wavelet neural network. In FW-NN, each fuzzy rule corresponds to a WNN comprised of numerous wavelets with changeable translation and dilation parameters. The fuzzy rules are being the consequent part of theFW-NN architecture and it is described only by wavelet functions. The output of WNN is expressed as per Eq. (14).

$$Y = \sum_{j=1}^{k} \delta_j \kappa_j(X) \tag{14}$$

In which κ_j is j^{th} layers wavelet activation function corresponding to the hidden layer. In addition, δ_j is the weight between the hidden (hid) and output layer.

The FWNN combines the wavelet functions and the TSK fuzzy system. A MF is shown by each of the region in the TSK fuzzy model. The FWNN has the properties of high precision and fast convergence. The FW-NN has six layers and they are discussed in the below section.

Layer 1 (input layer) The input signal vector $In = (In_1, In_2, \ldots, In_n)$ is fed as input to the next layer and the whole FW-NN model is trained with the bark frequency $c'(u')$.

Layer 2 (fuzzification layer) The fuzzy MFs are shown by each of the neuron in IF part of the rules. The MFs values are the outcomes' from this layer. In the first layer there is l^1 count of MFs and in the second layer there is l^2 count of MFs. For the i^{th} input variable, the Gaussian membership function is shown as per Eq. (15).

$$A^i_{j_i} = \exp\left(-\left(\frac{X_i - \varpi_{ji}}{\varsigma_{ji}}\right)^2\right); \quad i = 1, 2, ., n \ and \ j_i = 1, 2, \ldots, l_i \tag{15}$$

Layer 3 Grey wolf (fuzzy rule layer) In this layer, each neurons show fuzzy rule. The l^{th} nodes outcome is denoted as per Eq. (2). Here, each of the input MFs based possible combinations describes a fuzzy rule.

$$\eta^l = \prod_{i=1}^{n} A^i_{j_i}(X_i) \tag{16}$$

Layer 4 Grey wolf (normalization layer) Normalization factor is computed for each of the neurons in this layer. The l^{th} nodesnormalization factor is expressed as per Eq. (17).

$$\overline{\eta^l} = \frac{\eta^l}{\sum_{j=1}^{m} \eta^l} \tag{17}$$

Layer 5 The weighted output value is computed in this layer as per Eq. (18).

$$F^l = \overline{\eta^l} \chi^l \tag{18}$$

Layer 6 The overall output is calculated in this layer by summing the previous layers outputs. This is mathematically shown in Eq. (19).

$$Out = \sum_{l=1}^{m} F^l \tag{19}$$

During the training phase, the MSE is selected as the performance index and this MSE minimization is being the major objective of the current research work. The mathematical formula for MSE based training is shown in Eq. (20). Here, the actual FWNN outcome is *Act* and the desired outcome is *Pre*.

$$Er = \frac{1}{N} \sum_{k=1}^{N} (Act - Pre) \tag{20}$$

4 Adaptive randomizatized grey wolf algorithm: solution encoding and objective function

4.1 Objective function and solution encoding

The major objective of the current research work is to minimize the error Er of the FW-NN. This is expressed mathematically in Eq. (21).

$$Obj = Min(Er) \tag{21}$$

The AR-GWO is employed for properly tuning the tuning factor η, which is accomplished by means of optimizing the hidden neurons (hid) of FW-NN. The solution fed as input to AR-GWO is exhibited in Fig. 3.

4.2 Standard GWO

GWO [11, 32] was introduced by Mirjalili on the basis of the natural behavior of the grey wolves and it belongs to the category of swarm intelligence algorithm. Three are four types of grey wolves and these wolves stay in groups. The highest authority among them is the α (alpha) and it has the responsibility of taking decision. The supporter of α in taking decisions is β (beta), the lowest among these wolves is ω (omega) and it has to bow other wolves. The leftovers are referred as δ (delta). The main phases of GWO are "hunting, chasing and approaching the prey, encircling the prey and attacking the prey". The upcoming section portrays the mathematical model of GWO.

Mathematical model of GWO

(i) **Search for prey (exploitation)**: In the search process, the 1st, 2nd and 3rd best solutions are obtained during the search process of unique α, β and δ

(ii) **Encircling prey:** The mathematical formula for prey encircling during the hunting process is represented in Eqs. (22) and (23). In Eq. (24) the current iteration and the localization of the prey is represented as $x\&C_g(x)$. The coefficient vectors are indicated as Y and D. In addition, $C(x)$ represents the position of the grey wolf and the random values are manifested as b_1 & b_2. In addition Eqs. (24) and (25) are the mathematical formula for calculating the coefficient vectors Y and D, here there is a gradual decrease in the value of c from 2 to 0 over the course of iterations.

$$A = \left| D.C_g(x) - C(x) \right| \tag{22}$$

$$C(x+1) = C_g(x) - Y.A \tag{23}$$

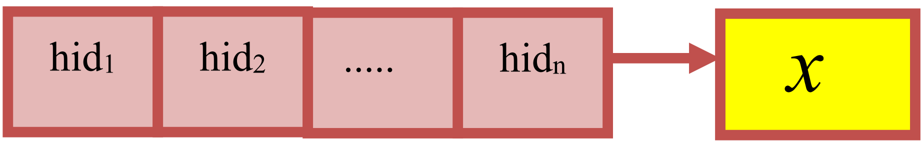

Fig. 3 Solution encoding

$$Y = 2.cb_1 - c \tag{24}$$

$$D = 2.b_2 \tag{25}$$

(iii) **Hunting the prey:** There lacks no information on the location of the prey in the search space. An assumption is made here that a better knowledge on the potential location of prey can be acquired from α, β and δ. This is the reason behind the storage first three results by discarding the others. The mathematical formula for hunting of prey is depicted in Eqs. (26) to (32) [32].

$$A_\alpha = |D.C_\alpha - C| \tag{26}$$

$$C_1 = C_\alpha - Y_1.(A_\alpha) \tag{27}$$

$$A_\beta = \left|D.C_\beta - C\right| \tag{28}$$

$$C_2 = C_\beta - Y_2.\left(A_\beta\right) \tag{29}$$

$$A_\delta = \left|D.C_\beta - C\right| \tag{30}$$

$$C_3 = C_\delta - Y_3.(A_\delta) \tag{31}$$

$$C(x+1) = \frac{C_1 + C_2 + C_3}{3} \tag{32}$$

(iv) **(iv) Attacking the prey (exploitation):** This is the end process of hunting behaviour of grey wolf and this process take place, when the prey is stationary.

4.3 AR-GWO

The conventional GWO suffers from the drawbacks of "bad local searching ability, low solving precision and slow convergence". So, the AR-GWO is formulated. In the conventional GWO, the random values $b_1$and b_2 are within the range [0, 1] and they are utilized to find the coefficient vectors Y and D in Eqs. (24) and (25). But, in the proposed model, instead of random numbers the proposed algorithm determines the random values b_{i1} and b_{i2} on the basis the fitness functions. The coefficient vectors are presented as Yi and Di are computed by utilizing Eqs. (33) and (34). Here i denote α, β and δ wolves. Further, the random values b_{i1} and b_{i2} are determined by using Eqs. (35) and (36), in which fitness of the best wolves either α, β $_{\text{or}}$ δ is represented as Fi,

$$Yi = 2.cb_{i1} - c \tag{33}$$

$$Di = 2.b_{i2} \tag{34}$$

$$b_{i1} = Fi \tag{35}$$

$$b_{21} = \frac{Fi}{\frac{1}{3}\sum_{i=\alpha,\beta,\delta} Fi} \tag{36}$$

The resultant from AR-GWO is the properly tuned tuning factor η^{tuned}, which is fed as input to adaptive Wiener filtering.

4.4 Adaptive WienerFiltering

The role of tuning ratio η^{tuned} is highly substantiated. The estimated tuning ratio by the FW-NN, on the basis of the $c'(u')$ (bark frequency) of the NMF-based filtered EMD signal $\overline{T}_o(n)$ is fine-tuned by AR-GWO. Mathematically, $c'(u')$ can be expressed as per Eq. (37)

$$c'\left(u'\right) = 13\arctan\left(0.76u'\right) + 3.5\arctan\left[(0.33u')^2\right] \tag{37}$$

The properly tuned tuning ratio η^{tuned} acquired from AR- GWO is fed as input to the wiener filter, instead of the constant η. The outcomes of the Adaptive Wiener filter are the filtered signal $\overline{\overline{T_u(n)}}$. Again, $\overline{\overline{T_u(n)}}$ is decomposed using EMD and the result is the enhanced denoised signal $\overline{\overline{T_o(n)}}$.

In the training process, the training library is constructed by giving the known $c'(u')$ (bark frequency) and tuning ratio η^{tuned} as inputs. The testing process is said to be the online process, while the training process is an offline process. The appropriate tuning factor for diverse noises are identified in the offline process and with this, the FW-NN is trained. The actual enhancement process takes place in the online mechanism, where the tuning factor is identified with the trained network.

5 Results and discussion

5.1 Experimental setup

The proposed speech enhancement model using GWO with FW-NN was implemented in MATLAB and the resultant of each of the analysis is observed. The data set for the research work is gathered from [23]. In this database, the five noise types, namely, "airport noise, exhibition noise, restaurant noise, station noise and street noise" are added to the speech signals. The performance of the proposed model (AR-GWO) is compared with the extant modelslike GA [29], PSO [20], ABC [24], FF [12] and GWO [14] in terms of "SDR, PESQ, SNR, RMSE, Correlation, ESTOI and CSED". Also, statistical analysis and computational time analysis are performed. Figure 4 exhibits the noisy and denoised signal for different approaches like GA, PSO, ABC, FF and GWO.

5.2 Performce analysis of airport noise

The performance evaluation of the proposed model over the existing model for airport noise at varying SNR levels is shown in Table 2. whenSNR = 0 dB, the SDR of the proposed model is 2.13%, 1.04%, 0.67%, 0.56% and 2.4% superior to the extant models like GA based η tuning,

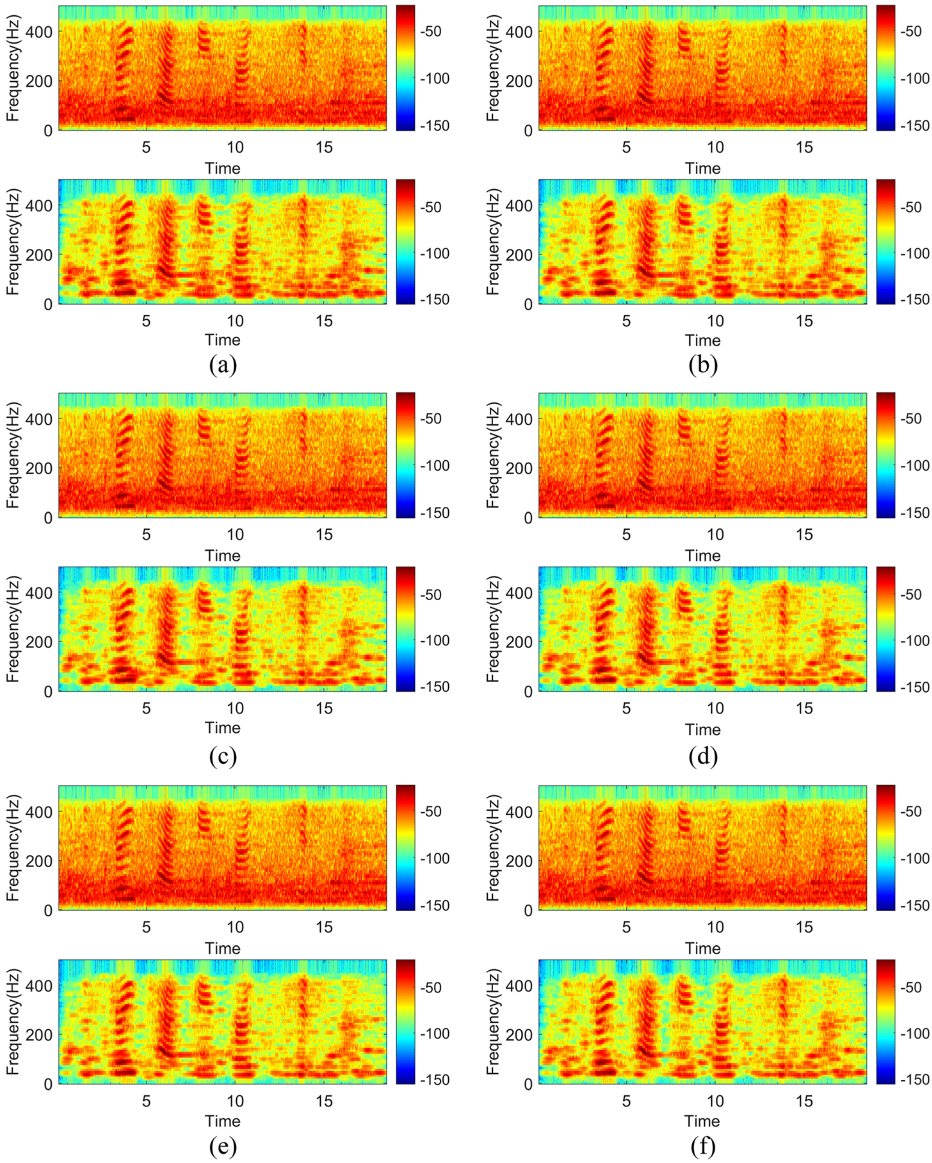

Fig. 4 The noisy and denoised signal of various approaches (**a**) GA (**b**) ABC (**c**) PSO (**d**) FF(**e**) GWO and (**f**) AR-GWO

Table 2 Performance evaluation of proposed model over existing for airport noise at varying SNR

Airport Noise at SNR = 0 dB							
Approaches	SDR	PESQ	SNR	RMSE	CORR	ESTOI	STOI
GA [29] based η tuning	4.7437	1.9181	5.0273	0.024047	0.83179	0.50627	0.7133
ABC [24] based η tuning	4.6986	1.9643	5.165	0.023663	0.82836	0.50418	0.71323
PSO [20] based η tuning	4.6933	1.9661	5.2589	0.023316	0.83346	0.49648	0.70853
FF [12] based η tuning	4.7909	1.9454	5.1998	0.023503	0.829	0.50408	0.71318
GWO [33] based η tuning	4.8551	1.9531	5.1905	0.023567	0.82864	0.50189	0.71254
AR-GWO based η tuning	4.9841	1.9703	5.2668	0.02327	0.82995	0.50845	0.71535
Airport Noise at SNR = 5 dB							
Approaches	SDR	PESQ	SNR	RMSE	CORR	ESTOI	STOI
GA [29] based η tuning	9.5432	2.3844	8.4923	0.016008	0.92871	0.6604	0.8186
ABC [24] based η tuning	9.71	2.3865	8.5657	0.015902	0.92942	0.66107	0.81988
PSO [20] based η tuning	9.4551	2.4003	8.6167	0.01572	0.93146	0.66148	0.82061
FF [12] based η tuning	9.5676	2.3794	8.5016	0.016011	0.9287	0.66177	0.81667
GWO [33] based η tuning	9.6879	2.3984	8.6911	0.015614	0.93199	0.66267	0.82075
AR-GWO based η tuning	9.746	2.4058	8.7794	0.01546	0.9324	0.66586	0.82486
Airport Noise at SNR = 10 dB							
Approaches	SDR	PESQ	SNR	RMSE	CORR	ESTOI	STOI
GA [29] based η tuning	12.373	2.7008	10.51	0.012779	0.95782	0.77512	0.88209
ABC [24] based η tuning	12.533	2.7093	10.843	0.01222	0.96042	0.77079	0.88111
PSO [20] based η tuning	12.493	2.7153	10.54	0.01263	0.95905	0.76772	0.87947
FF [12] based η tuning	12.517	2.7075	10.754	0.012384	0.95943	0.76992	0.88066
GWO [33] based η tuning	12.57	2.7293	10.765	0.012342	0.95975	0.77607	0.88368
AR-GWO based η tuning	12.671	2.732	11.128	0.011823	0.96271	0.78377	0.89057
Airport Noise at SNR = 15 dB							
Approaches	SDR	PESQ	SNR	RMSE	CORR	ESTOI	STOI
GA [29] based η tuning	14.046	2.9512	11.823	0.011128	0.96883	0.83878	0.91063
ABC [24] based η tuning	14.387	2.9856	12.257	0.010448	0.97199	0.84432	0.91623
PSO [20] based η tuning	14.023	2.9525	11.73	0.011077	0.96909	0.83847	0.91185
FF [12] based η tuning	14.37	2.9732	12.34	0.010484	0.97152	0.84457	0.91637
GWO [33] based η tuning	14.303	2.9651	12.18	0.010559	0.9715	0.84048	0.91378
AR-GWO based η tuning	14.616	2.988	12.576	0.010081	0.97384	0.84906	0.9207

ABC based η tuning, PSO based η tuning, FF based η tuning and GWO based η tuning, respectively.PESQ of the proposed model at SNR = 0 dB exhibits an improvement of 3.7%, 2.6%, 4.5%, 1.3% and 1% over the extant models like GA based η tuning, ABC based η tuning, PSO based η tuning, FF based η tuning and GWO based η tuning, respectively.In airport noise at SNR = 5 dB, RMSE of the proposed model is 3.4%, 2.7%, 1.6%, 3.44% and 0.9% superior to the traditional models like GA based η tuning, ABC based η tuning, PSO based η tuning, FF based η tuning and GWO based η tuning, respectively. At SNR = 10 dB, ESTOI of the proposed model is 1.11% better than GA based η tuning, 1.6% better than ABC based η tuning, 2% better than PSO based η tuning, 1.7% better than FF based η tuning and 0.9% better than GWO based η tuning. Further, at SNR = 10 dB, STOI of the proposed model shows an improvement of 0.9%, 1%, 1.2%, 1.1% and 0.7% better than classical model like GA based η tuning, ABC based η tuning, PSO based η tuning, FF based η tuning and GWO based η tuning, respectively.

5.3 Performce analysis of exhibition noise

Table 3 exhibits the performance analysis of the proposed model over exiting for exhibition noise at different SNR levels. At SNR = 0 dB, the proposed model shows an improvement of 10.3%, 3.6% 10.2%, 3.1% and 1.6% over the classical models like GA based η tuning, ABC

Table 3 Performance evaluation of proposed model over existing for exhibition noise at varying SNR

Exhibition Noise at SNR=0 dB							
Approaches	SDR	PESQ	SNR	RMSE	CORR	ESTOI	STOI
GA [29] based η tuning	4.8816	1.8847	4.9769	0.024181	0.83386	0.5536	0.73186
ABC [24] based η tuning	5.3046	1.9195	5.2921	0.023453	0.83881	0.55774	0.7336
PSO [20] based η tuning	4.934	1.9288	4.9329	0.024466	0.83118	0.5594	0.73382
FF [12] based η tuning	5.3942	1.9284	5.3336	0.023355	0.83972	0.55461	0.73462
GWO [33] based η tuning	5.5763	1.9176	5.404	0.023082	0.84347	0.56018	0.73376
AR-GWO based η tuning	5.6496	1.9372	5.4944	0.022868	0.84495	0.56032	0.73619
Exhibition Noise at SNR=5 dB							
Approaches	SDR	PESQ	SNR	RMSE	CORR	ESTOI	STOI
GA [29] based η tuning	8.8427	2.268	8.0345	0.016949	0.9184	0.66867	0.82182
ABC [24] based η tuning	9.2962	2.2925	8.338	0.016348	0.92256	0.67082	0.82384
PSO [20] based η tuning	9.383	2.3008	8.3461	0.016334	0.92438	0.67058	0.81908
FF [12] based η tuning	9.4158	2.3035	8.2985	0.016411	0.92435	0.67239	0.822
GWO [33] based η tuning	8.8893	2.2617	8.3604	0.016278	0.92534	0.67411	0.82333
AR-GWO based η tuning	9.497	2.3164	8.4358	0.016145	0.92541	0.67664	0.82735
Exhibition Noise at SNR=10 dB							
Approaches	SDR	PESQ	SNR	RMSE	CORR	ESTOI	STOI
GA [29] based η tuning	12.766	2.6089	10.685	0.012575	0.95814	0.77244	0.88091
ABC [24] based η tuning	12.819	2.626	10.922	0.01215	0.96033	0.77738	0.88182
PSO [20] based η tuning	12.838	2.6333	10.654	0.012527	0.95848	0.77733	0.88133
FF [12] based η tuning	12.791	2.63	10.615	0.012602	0.95782	0.77547	0.88071
GWO [33] based η tuning	12.894	2.6031	10.727	0.012474	0.95847	0.77665	0.88108
AR-GWO based η tuning	12.951	2.6351	11.112	0.011962	0.96048	0.77782	0.88468
Exhibition Noise at SNR=15 dB							
Approaches	SDR	PESQ	SNR	RMSE	CORR	ESTOI	STOI
GA [29] based η tuning	15.159	2.8586	12.591	0.010224	0.97272	0.84184	0.91566
ABC [24] based η tuning	15.389	2.9084	12.878	0.00981	0.97467	0.85189	0.92211
PSO [20] based η tuning	15.296	2.9074	12.375	0.01035	0.97274	0.84318	0.91555
FF [12] based η tuning	15.201	2.8987	12.537	0.01017	0.97332	0.84546	0.91675
GWO [33] based η tuning	15.247	2.9056	12.532	0.010191	0.97319	0.84327	0.91605
AR-GWO based η tuning	15.602	2.9094	13.251	0.009483	0.97579	0.85294	0.92215

based η tuning, PSO based η tuning, FF based η tuning and GWO based η tuning, respectivelyin terms of SNR. Further, RMSE of the proposed model is 5.4%, 2.4%, 6.5%, 2% and 0.9% better than the extant models like GA based η tuning, ABC based η tuning, PSO based η tuning, FF based η tuning and GWO based η tuning, respectively at SNR = 0 dB. For the exhibition noise at SNR = 5 dB, the SDR of the proposed model is improvedover the existing model as 15.7%, 2.16%, 1.2%, 0.86%, and 6.83% by GA based η tuning, ABC based η tuning, PSO based η tuning, FF based η tuning and GWO based η tuning, respectively. Then, in terms of Correlation at SNR = 10 dB, the proposed model is found to be better than the existing approaches GA based η tuning, ABC based η tuning, PSO based η tuning, FF based η tuning and GWO based η tuning, respectively. ESTOI of the proposed model is 0.6% better than GA based η tuning, 0.05% better than ABC based η tuning, 0.06% better than PSO based η tuning, 0.3% better than FF based η tuningand 0.15% better than GWO based η tuningat SNR = 10 dB.

5.4 Performce analysis of restaurant noise

Table 4 portrays the performance evaluation of the proposed model over the existing for restaurant noise at different SNR levels. From, which SDR of the signal at SNR = 0 dB is 8.5%, 5.7%, 9.7%, 5.7% and 4.3% superior to the classical models like GA based η tuning,

Table 4 Performance evaluation proposed model over existing for restaurant noise at varying SNR

Restaurant Noise at SNR=0 dB							
Approaches	SDR	PESQ	SNR	RMSE	CORR	ESTOI	STOI
GA [29] based η tuning	3.7234	1.9355	4.459	0.025493	0.80808	0.51043	0.70067
ABC [24] based η tuning	3.8386	1.9501	4.5555	0.025202	0.8115	0.51518	0.70489
PSO [20] based η tuning	3.6758	1.9541	4.3721	0.025728	0.8088	0.51486	0.70261
FF [12] based η tuning	3.838	1.9409	4.5605	0.025168	0.81113	0.51307	0.70103
GWO [33] based η tuning	3.8974	1.9553	4.5775	0.025115	0.81268	0.51203	0.70278
AR-GWO based η tuning	4.073	1.9597	4.6674	0.024878	0.81594	0.517	0.7076
Restaurant Noise at SNR=5 dB							
Approaches	SDR	PESQ	SNR	RMSE	CORR	ESTOI	STOI
GA [29] based η tuning	8.8132	2.3121	8.1199	0.016642	0.92163	0.65082	0.81008
ABC [24] based η tuning	8.8225	2.316	8.2539	0.016383	0.92313	0.64843	0.80827
PSO [20] based η tuning	8.8684	2.3083	8.137	0.016599	0.92168	0.64538	0.8055
FF [12] based η tuning	8.6289	2.3145	8.2625	0.016515	0.92195	0.64878	0.80944
GWO [33] based η tuning	8.8698	2.3071	8.2049	0.016374	0.92304	0.64866	0.80906
AR-GWO based η tuning	8.9533	2.3219	8.2702	0.01639	0.92368	0.65163	0.81352
Restaurant Noise at SNR=10 dB							
Approaches	SDR	PESQ	SNR	RMSE	CORR	ESTOI	STOI
GA [29] based η tuning	11.913	2.6448	10.313	0.013046	0.95504	0.7633	0.87382
ABC [24] based η tuning	12.11	2.6647	10.638	0.012508	0.95762	0.76751	0.87699
PSO [20] based η tuning	11.981	2.6569	10.2	0.013148	0.95455	0.76607	0.87308
FF [12] based η tuning	11.915	2.6478	10.452	0.012784	0.95646	0.76327	0.87696
GWO [33] based η tuning	12.089	2.6708	10.565	0.012615	0.95741	0.76792	0.87634
AR-GWO based η tuning	12.139	2.6774	10.696	0.012403	0.95866	0.7743	0.88253
Restaurant Noise at SNR=15 dB							
Approaches	SDR	PESQ	SNR	RMSE	CORR	ESTOI	STOI
GA [29] based η tuning	14.189	2.8967	12.05	0.010765	0.97066	0.84242	0.91343
ABC [24] based η tuning	14.469	2.9243	12.45	0.010169	0.97351	0.84493	0.91598
PSO [20] based η tuning	14.298	2.9281	12.014	0.010704	0.97139	0.84393	0.91434
FF [12] based η tuning	14.268	2.9093	12.153	0.010616	0.97125	0.84181	0.91481
GWO [33] based η tuning	14.281	2.9194	12.14	0.010567	0.97167	0.84285	0.91416
AR-GWO based η tuning	14.617	2.9452	12.676	0.009908	0.9749	0.85112	0.92113

ABC based η tuning, PSO based η tuning, FF based η tuning and GWO based η tuning. PESQ of the proposed modelis found to be 1.2%, 0.4% better than GA and ABC, 0.2%, 0.9% and0.7% better than PSO, FF and GWO, respectively at SNR = 0 dB. Then, for SNR = 5 dB, RMSE of the proposed model exhibits superiority to the traditional models like GA based η tuning, ABC based η tuning, PSO based η tuning, FF based η tuning and GWO based η tuning by 1.54%, 0.04%, 1.2%, 0.7% and 0.09%, respectively. Further, in terms of SNR, there is an improvement of 1.8%, 0.19%, 1.6%, 0.9% and 0.7% in the proposed model over the existing model like GA, ABC, PSO, FF and GWO, respectively at SNR = 5 dB. Moreover, from SNR = 10 dB, STOI of the proposed model is 0.8%, 0.5%, 0.7%, 0.6% and 0.76% better than state-of-art models like GA based η tuning, ABC based η tuning, PSO based η tuning, FF based η tuning and GWO based η tuning. CSED at SNR = 10 dB is 6.9%, 2%, 6.7%, 5.9% and 5.13% superior to the extant modelsGA based η tuning, ABC based η tuning, PSO based η tuning, FF based η tuning and GWO based η tuning, respectively.

5.5 Performce analysis of station noise

From 5 represents the performance analysis of the proposed model over exiting for station noise at different SNR values as (Table 5 shows the performance evaluation of the proposed model over existing model for station noise at varying SNR).. From the table, at SNR = 0 dB, the proposed model overtakes the extant modelsGA based η tuning, ABC based η tuning, PSO based η tuning, FF based η tuning and GWO based η tuning by 2.6%, 4.5%, 1.2%, 2.7% and 1%, respectively in terms of SDR. Moreover, SEQ of the proposed model at SNR = 0 dB is 3.7%, 1.14%, 2.2%, 1.7% and 1% better than the extant models like GA based η tuning, ABC based η tuning, PSO based η tuning, FF based η tuning and GWO based η tuning, respectively. Then, for SNR = 5 dB, the proposed model is better than extant models, 5.4% by GA, 2.4% by ABC, 1.2% by PSO, 3.1% by FF and 1.6% by GWO. SEI of the proposed model at SNR = 10 dB, an improvement of 0.3%, 0.2%, 0.1%, 0.5% and 0.02% over the state-of-art models. Then, atSNR = 15 dB, the proposed model is 0,6%, 0.02%, 0.9%, 0.3% and 0.4% better than extant modelsGA based η tuning, ABC based η tuning, PSO based η tuning, FF based η tuning and GWO based η tuning in terms of STOI. In terms of CSED at SNR = 15 dB, the proposed model is 9.5%, 4.5%, 10.9%, 9.8% and 7.7% better than the traditional models GA based η tuning, ABC based η tuning, PSO based η tuning, FF based η tuning and GWO based η tuning, respectively.

Table 5 Performance evaluation proposed model over existing for station noise at varying SNR

Station Noise at SNR=0 dB							
Approaches	SDR	PESQ	SNR	RMSE	CORR	ESTOI	STOI
GA [29] based η tuning	6.4259	2.037	5.7614	0.022214	0.86557	0.47946	0.70202
ABC [24] based η tuning	6.3054	2.0924	5.6648	0.022488	0.86031	0.4714	0.70522
PSO [20] based η tuning	6.5155	2.0598	5.7821	0.022054	0.86512	0.47993	0.70432
FF [12] based η tuning	6.4156	2.0788	5.8548	0.021868	0.86597	0.47663	0.70174
GWO [33] based η tuning	6.5284	2.0458	5.7819	0.022073	0.86721	0.47881	0.70517
AR-GWO based η tuning	6.5994	2.1167	5.9058	0.021723	0.86727	0.48062	0.70639
Station Noise at SNR=5 dB							
Approaches	SDR	PESQ	SNR	RMSE	CORR	ESTOI	STOI
GA [29] based η tuning	9.6888	2.4452	8.168	0.01657	0.92905	0.65479	0.81531
ABC [24] based η tuning	9.8802	2.4501	8.5373	0.015867	0.93236	0.64988	0.81377
PSO [20] based η tuning	9.8902	2.4446	8.356	0.016186	0.93232	0.65575	0.81579
FF [12] based η tuning	9.894	2.4484	8.3898	0.016162	0.93177	0.65616	0.81535
GWO [33] based η tuning	9.924	2.4529	8.4232	0.016112	0.9314	0.65599	0.816
AR-GWO based η tuning	10.04	2.4669	8.6676	0.015649	0.93302	0.6575	0.81666
Station Noise at SNR=10 dB							
Approaches	SDR	PESQ	SNR	RMSE	CORR	ESTOI	STOI
GA [29] based η tuning	12.654	2.7153	10.69	0.012513	0.9598	0.76866	0.87386
ABC [24] based η tuning	12.746	2.7219	10.816	0.012267	0.96001	0.76982	0.87447
PSO [20] based η tuning	12.482	2.7009	10.411	0.012815	0.95793	0.76372	0.86911
FF [12] based η tuning	12.647	2.7094	10.755	0.012396	0.95977	0.76721	0.87422
GWO [33] based η tuning	12.654	2.7334	10.744	0.012352	0.96011	0.77175	0.87556
AR-GWO based η tuning	12.758	2.7343	10.988	0.012003	0.9621	0.77158	0.87685
Station Noise at SNR=15 dB							
Approaches	SDR	PESQ	SNR	RMSE	CORR	ESTOI	STOI
GA [29] based η tuning	14.281	2.9701	11.881	0.011026	0.96941	0.83574	0.91008
ABC [24] based η tuning	14.569	2.9968	12.436	0.010197	0.97367	0.84294	0.91569
PSO [20] based η tuning	14.096	2.967	11.657	0.011177	0.96897	0.83418	0.90697
FF [12] based η tuning	14.394	2.9911	12.166	0.010578	0.97211	0.83742	0.91285
GWO [33] based η tuning	14.338	2.9964	12.044	0.010711	0.97133	0.83899	0.91214
AR-GWO based η tuning	14.732	2.9993	12.671	0.009969	0.97436	0.84346	0.91593

5.6 Performce analysis of street noise

The performance evaluation of the proposed model over the existing model for the street noise is shown in Table 6. For SNR = 0db, the proposed model exhibits an improvement of 1.7%, 0.5%, 1.3%, 1.72% and 1.7% over the classical models like GA based η tuning, ABC based η tuning, PSO based η tuning, FF based η tuning and GWO based η tuning, respectively in terms of CSED. Then, for the same SNR, the STOI of the projected model is 0.4%, 0.3%, 0.9%, 0.7% and 0.3% superior to the state-of-art models like GA based η tuning, ABC based η tuning, PSO based η tuning, FF based η tuning and GWO based η tuning, respectively. For SNR = 5 dB, ESTOI of the proposed model is 1.3%, 0.9%, 1.4%, 0.5% and 0.4% superior to the conventional models GA based η tuning, ABC based η tuning, PSO based η tuning, FF based η tuning and GWO based η tuning, respectively. The correlation of the signal for the same SNR is 0.35%, 0.017%, 0.13%, 0.2% and 0.29% superior to the existing approaches GA based η tuning, ABC based η tuning, PSO based η tuning, FF based η tuning and GWO based η tuning, respectively. Then, in case of SNR = 15 dB, the PESQ of the proposed model is 1.8%, 0.27%, 0.8%, 0.9% and 0.89% better than GA based η tuning, ABC based η tuning,

Table 6 Performance evaluation proposed model over existing for street noise at varying SNR

Street Noise at SNR=0 dB							
Approaches	SDR	PESQ	SNR	RMSE	CORR	ESTOI	STOI
GA [29] based η tuning	6.4479	2.0587	5.8042	0.02216	0.86391	0.51601	0.72324
ABC [24] based η tuning	6.7136	2.1067	5.8664	0.021787	0.86832	0.51561	0.72419
PSO [20] based η tuning	6.3911	2.0922	6.0489	0.021285	0.86341	0.51708	0.71952
FF [12] based η tuning	6.8776	2.0922	5.9977	0.021426	0.86787	0.51849	0.7213
GWO [33] based η tuning	6.7047	2.0819	5.9561	0.021506	0.86648	0.51715	0.72374
AR-GWO based η tuning	6.9113	2.1121	6.0753	0.021276	0.86971	0.51968	0.72652
Street Noise at SNR=5 dB							
Approaches	SDR	PESQ	SNR	RMSE	CORR	ESTOI	STOI
GA [29] based η tuning	10.343	2.4608	8.459	0.016038	0.93221	0.65462	0.80808
ABC [24] based η tuning	10.477	2.4569	8.7538	0.015468	0.9354	0.65674	0.81378
PSO [20] based η tuning	10.408	2.4771	8.5652	0.015798	0.9343	0.65378	0.80799
FF [12] based η tuning	10.409	2.4678	8.6381	0.015704	0.9335	0.65946	0.81023
GWO [33] based η tuning	10.378	2.4718	8.5641	0.015827	0.93281	0.6603	0.81101
AR-GWO based η tuning	10.508	2.4865	8.8397	0.015302	0.93556	0.6632	0.81415
Street Noise at SNR=10 dB							
Approaches	SDR	PESQ	SNR	RMSE	CORR	ESTOI	STOI
GA [29] based η tuning	13.28	2.759	11.182	0.011819	0.96423	0.75127	0.87123
ABC [24] based η tuning	13.483	2.7811	11.623	0.01118	0.96669	0.75718	0.87601
PSO [20] based η tuning	13.477	2.7518	11.243	0.01165	0.96555	0.75744	0.87482
FF [12] based η tuning	13.473	2.7955	11.401	0.011478	0.96623	0.75838	0.87631
GWO [33] based η tuning	13.451	2.7753	11.307	0.0116	0.96554	0.75485	0.87494
AR-GWO based η tuning	13.619	2.7998	11.643	0.011159	0.96756	0.75995	0.87721
Street Noise at SNR=15 dB							
Approaches	SDR	PESQ	SNR	RMSE	CORR	ESTOI	STOI
GA [29] based η tuning	14.94	2.9263	12.472	0.010375	0.97234	0.82301	0.90012
ABC [24] based η tuning	15.24	2.9737	12.86	0.009771	0.97544	0.82681	0.90526
PSO [20] based η tuning	14.869	2.9574	12.26	0.010444	0.97267	0.81959	0.89918
FF [12] based η tuning	15.025	2.9532	12.597	0.010133	0.97394	0.82197	0.9005
GWO [33] based η tuning	14.958	2.9554	12.435	0.010269	0.97339	0.8188	0.89966
AR-GWO based η tuning	15.286	2.982	13.042	0.009615	0.976	0.82982	0.90647

PSO based η tuning, FF based η tuning and GWO based η tuning, respectively. Then, for the same SNR, the proposed model is 2.2%, 0.3%, 2.7%, 1.7% and 2.12% better than the traditional models like GA based η tuning, ABC based η tuning, PSO based η tuning, FF based η tuning and GWO based η tuning, respectively in terms of SDR at 15 dB.

5.7 Statistical analysis

The evaluation of statistical analysis of the adopted and existing approaches is depicted in Fig. 5. The outcomes are provided based on the error and thus the proposed model value is lower than the existing works. On considering the results, the best value of the adopted AR-GWO scheme is 9.41%, 3.51%, 8.99%, 3.84%, and 4.53% superior to the existing GA, ABC, PSO, FF, and GWO approaches. Furthermore, in mean case scenario, the suggested approach value is 5.21%, 2.57%, 3.37%, 2.81%, and 2.33% superior to the existing GA, ABC, PSO, FF, and GWO approaches. Moreover, the median value of the AE-GWO approach is 0.013641 and it is 5.22%, 2.99%, 3.77%, 3.92%, and 2.41% better than the existing GA, ABC, PSO, FF, and GWO methods. Therefore, the effectiveness of the proposed speech enhancement model is proved.

5.8 Computational time analysis

In this section, the computational time of the proposed and existing methods is evaluated and it is depicted in Fig. 6. From the graph, the computation time of the proposed AR-GWO method is 227.86 and it is 34.07%, 43.57%, 28.86%, 38.88%, and 16.03% better than the existing GA, ABC, PSO, FF, and GWO approaches respectively. Thus, the effectiveness of the adopted AR-GWO based speech enhancement method is validated.

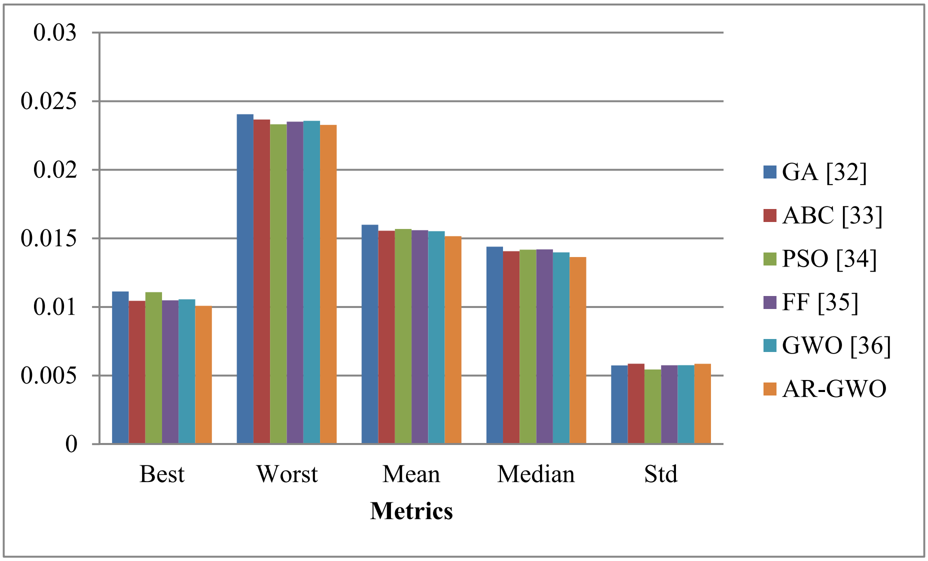

Fig. 5 Statistical analysis of the proposed and existing approaches

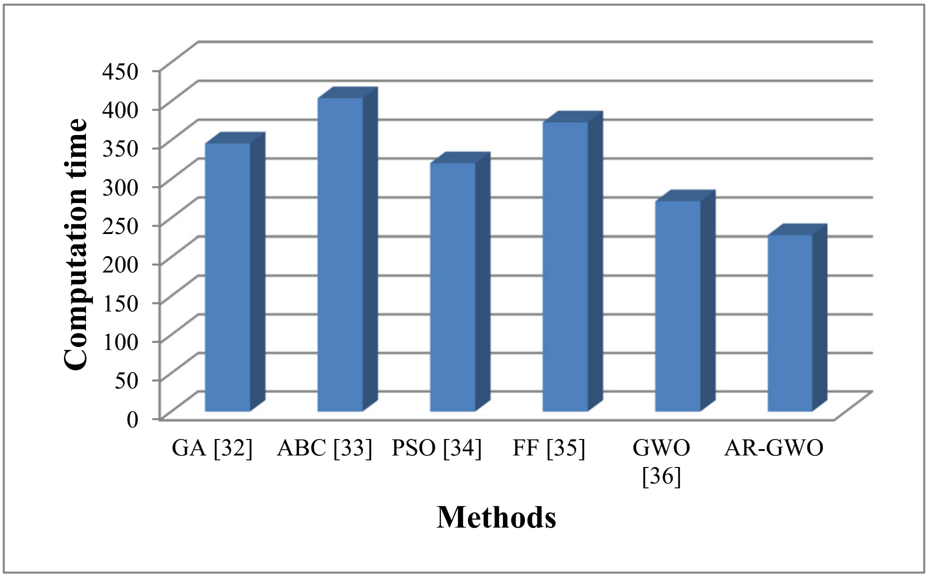

Fig. 6 Computational time analysis of the proposed and existing approaches

5.9 Practical implications

The major aim of the proposed speech enhancement is to suppress the noise in a noisy speech signal and improve the quality and intelligibility of speech. The proposed speech enhancement approach utilizes in real-time applications such as speech recognition, mobile phones, VoIP, teleconferencing systems and hearing aids.

6 Conclusion

In this paper, an optimized fuzzy wavelet neural network based speech enhancement model is proposed. In the training phase, the input noise corrupted signal was initially provided as input to both STFT-based noise estimation and NMF-based spectrum estimation for estimating the noise spectrum and signal spectrum, respectively. The obtained noise spectrum and the signal spectrum are fed as input to the wiener filter and these filtered signals are subjected to EMD.Since, tuning factorη plays a key role in wiener filter, it has to be determined for each signals, and is trained in FW-NN. Then, from the denoised signal the bark frequency is evaluated. The computed bark frequency is fed as input to the learning algorithm referred as FW-NN for detecting the suited tuning factorη for the entire input signal in Weiner filter. The AR-GWO is employed for proper tuning of the tuning factor η referred as tuned tuning factor (η^{tuned}). In the testing phase, the training is accomplished initially and from which the tuning factor is gathered for each of the relevant input signal. Then, the properly tuned tuning factor (η^{tuned}) from FW-NN is fed as input to EMD via adaptive wiener filter for decomposing the spectral signal and the output of EMD is denoised enhanced speech signal.Theresultant acquired is compared over the existing models in terms of various measures. In case of street noise, at SNR = 0db, the proposed model exhibits an improvement of 1.7%, 0.5%, 1.3%, 1.72% and 1.7% over the classical models like GA based η tuning, ABC based η tuning, PSO based η tuning, FF based η tuning and GWO based η tuning, respectively in terms of CSED. Thus, the effectiveness of the work is validated via the result analysis. However, in statistical analysis, the standard deviation metric value is not better than the existing ones. Hence, in the future work, we enhanced our proposed work by utilizing the recent optimization algorithms and validate the work in real-time applications.

Acknowledgements I would like to express my very great appreciation to the co-authors of this manuscript for their valuable and constructive suggestions during the planning and development of this research work.

Author contribution All authors have made substantial contributions to conception and design, revising the manuscript, and the final approval of the version to be published. Also, all authors agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Data availability The data that support the findings of this study is "NOIZEUS" openly available in https://ecs.utdallas.edu/loizou/speech/noizeus/.

Declarations

Ethical approval This paper does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Not Applicable.

Conflict of interest The authors declare no conflict of interest.

References

1. Abel J, Fingscheidt T (2018) Artificial speech bandwidth extension using deep neural networks for wideband spectral envelope estimation. IEEE/ACM Trans Audio, Speech, Lang Process 26(1):71–83
2. Arcos CD, Vellasco M, Alcaim A (2018) Ideal neighbourhood mask for speech enhancement. Electron Lett 54(5):317–318
3. Bai H, Ge F, Yan Y (2018) DNN-based speech enhancement using soft audible noise masking for wind noise reduction. China Commun 15(9):235–243
4. Bando Y, Itoyama K, Konyo M, Tadokoro S, Nakadai K, Yoshii K, Kawahara T, Okuno HG (2018) Speech enhancement based on Bayesian low-rank and sparse decomposition of multichannel magnitude spectrograms. IEEE/ACM Trans Audio, Speech, Lang Process 26(2):215–230
5. Bao F, Abdulla WH (2019) A new ratio mask representation for CASA-based speech enhancement. IEEE/ACM Trans Audio, Speech Lang Process 27(1):7–19
6. Chazan SE, Goldberger J, Gannot S (2016) A hybrid approach for speech enhancement using MoG model and neural network phoneme classifier. IEEE/ACM Trans Audio, Speech, Lang Process 24(12):2516–2530
7. Dehghani M, Montazeri Z, Dhiman G, Malik OP, Morales-Menendez R, Ramirez-Mendoza RA, Dehghani A, Guerrero JM, Parra-Arroyo L (2020) A spring search algorithm applied to engineering optimization problems. Appl Sci 10(18):6173
8. Dhiman G, Kaur A (2019) STOA: a bio-inspired based optimization algorithm for industrial engineering problems. Eng Appl Artif Intell 82:148–174
9. Dhiman G, Kumar V (2017) Spotted hyena optimizer: a novel bio-inspired based metaheuristic technique for engineering applications. Adv Eng Softw 114:48–70
10. Dhiman G, Kumar V (2018) Emperor penguin optimizer: a bio-inspired algorithm for engineering problems. Knowl-Based Syst 159:20–50

11. Fahad M, Aadil F, Rehman Z, Khana S, Shah PA, Muhammad K, Lloret J, Wang H, Lee JW, Mehmoode I (2018) Grey wolf optimization based clustering algorithm for vehicular ad-hoc networks. Comput Electric Eng 70:853–870
12. Fister I, Iztok Fister X-SY Jr, Brest J (2013) A comprehensive review of firefly algorithms. Swarm Evol Comput 13:34–46
13. Gannot S, Burshtein D, Weinstein E (2008) Iterative and sequential Kalman filter-based speech enhancement algorithms. IEEE Trans Speech Audio Process 6(4):373–385
14. Garg A, Sahu OP (2020) Enhancement of speech signal using diminished empirical mean curve decomposition-based adaptive Wiener filtering. Pattern Anal Appl 23(1):179–198
15. Grimble M (1984) Weiner and Kalman filters for systems with random parameters. IEEE Trans Autom Control 29(6):552–554
16. Grispino AS, Petracca GO, Dominguez AE (2013) Comparative analysis of wavelet and EMD in the filtering of radar signal affected by Brown noise. IEEE Latin Am Trans 11(1):81–85
17. Guido RC (2011) A note on a practical relationship between filter coefficients and scaling and wavelet functions of discrete wavelet transforms. Appl Math Lett 24(7):1257–1259
18. Guido RC (2017) Effectively interpreting discrete wavelet transformed signals [lecture notes]. IEEE Signal Process Mag 34(3):89–100
19. Guido RC, Vieira LS, Junior SB, Sanchez FL, Maciel CD, Fonseca ES, Pereira JC (2007) A neural-wavelet architecture for voice conversion. Neurocomputing 71(1–3):174–180
20. Hamza D, Tashan T (2021) Dual channel speech enhancement using particle swarm optimization. Indonesian J Electric Eng Comput Sci 23(2):821–828
21. He Q, Bao F, Bao C (2017) Multiplicative update of auto-regressive gains for codebook-based speech enhancement. IEEE/ACM Trans Audio, Speech, Lang Process 25(3):457–468
22. Hou J, Wang S, Lai Y, Tsao Y, Chang H, Wang H (2018) Audio-visual speech enhancement using multimodal deep convolutional neural networks. IEEE Trans Emerg Topics Comput Intell 2(2):117–128
23. https://ecs.utdallas.edu/loizou/speech/noizeus/ (n.d.) (Access Date: 01-03-2019)
24. Karaboga D, Basturk B (2008) On the performance of artificial bee colony (ABC) algorithm. Appl Soft Comput 8(1):687–697
25. Kaur S, Awasthi LK, Sangal AL, Dhiman G (2020) Tunicate Swarm Algorithm: A new bio-inspired based metaheuristic paradigm for global optimization. Eng Appl Artif Intell 90:103541
26. Krawczyk M, Gerkmann T (2014) STFT phase reconstruction in voiced speech for an improved Single-Channel speech enhancement. IEEE/ACM Trans Audio, Speech Lang Process 22(12):1931–1940
27. Krawczyk-Becker M, Gerkmann T (2018) On speech enhancement under PSD uncertainty. IEEE/ACM Trans Audio, Speech, Lang Process 26(6):1144–1153
28. Kuqi, B, Elezaj E, Millaku B, Dreshaj A, Hung NT (2021) "The impact of COVID-19 (SARS-CoV-2) in tourism industry: evidence of Kosovo during Q1, Q2 and Q3 period of 2020." J Sustain Finance Invest 1–12
29. LeBlanc, R, Selouani SA (2019) "Self-adaptive tuning for speech enhancement algorithm based on evolutionary approach." In 2019 IEEE First International Conference on Cognitive Machine Intelligence (CogMI), pp. 16–22. IEEE
30. Lee J, Skoglund J, Shabestary T, Kang H (2018) Phase-sensitive joint learning algorithms for deep learning-based speech enhancement. IEEE Signal Process Lett 25(8):1276–1280
31. Martín-Doñas JM, Gomez AM, Gonzalez JA, Peinado AM (2018) A deep learning loss function based on the perceptual evaluation of the speech quality. IEEE Signal Process Lett 25(11):1680–1684
32. Ming J, Crookes D (2017) Speech enhancement based on full-sentence correlation and clean speech recognition. IEEE/ACM Trans Audio, Speech, Lang Process 25(3):531–543
33. Mirjalili S (2014) Seyed Mohammad Mirjalili, Andrew Lewis, "Grey wolf optimizer". Adv Eng Softw 69:46–61
34. Mohammadiha N, Smaragdis P, Leijon A (2013) Supervised and unsupervised speech enhancement using nonnegative matrix factorization. IEEE Trans Audio Speech Lang Process 21(10):2140–2151
35. Ou S, Song P, Gao Y (2018) Soft Decision Based Gaussian-Laplacian Combination Model for Noisy Speech Enhancement. Chin J Electron 27(4):827–834
36. Parente G, Gargano T, Di Mitri M, Cravano S, Thomas E, Vastano M, Maffi M, Libri M, Lima M (2021) Consequences of COVID-19 lockdown on children and their pets: dangerous increase of dog bites among the paediatric population. Children 8(8):620
37. Prasanalakshmi B, Farouk A (2019) Classification and prediction of student academic performance in king khalid university-a machine learning approach. Indian J Sci Technol 12:14
38. Rehr R, Gerkmann T (2018) On the importance of super-Gaussian speech priors for machine-learning based speech enhancement. IEEE/ACM Trans Audio, Speech, Lang Process 26(2):357–366
39. Samui S, Chakrabarti I, Ghosh SK (2019) Time–frequency masking based supervised speech enhancement framework using fuzzy deep belief network. Appl Soft Comput 74:583–602

40. Shao Y, Chang C (2011) Bayesian separation with sparsity promotion in perceptual wavelet domain for speech enhancement and hybrid speech recognition. IEEE Trans Syst Man Cybern Syst Hum 41(2):284–293
41. Stahl J, Mowlaee P (2018) A pitch-synchronous simultaneous detection-estimation framework for speech enhancement. IEEE/ACM Trans Audio, Speech, Lang Process 26(2):436–450
42. Sun M, Li Y, Gemmeke JF, Zhang X (2015) Speech enhancement under low SNR conditions via noise estimation using sparse and low-rank NMF with Kullback–Leibler divergence. IEEE/ACM Trans Audio, Speech Lang Process 23(7):1233–1242
43. Tan K, Chen J, Wang D (2019) Gated residual networks with dilated convolutions for monaural speech enhancement. IEEE/ACM Trans Audio, Speech Lang Process 27(1):189–198
44. Tantibundhit C, Pernkopf F, Kubin G (2010) Joint time–frequency segmentation algorithm for transient speech decomposition and speech enhancement. IEEE Trans Audio Speech Lang Process 18(6):1417–1428
45. Wang Y, Brookes M (2018) Model-based speech enhancement in the modulation domain. IEEE/ACM Trans Audio, Speech, Lang Process 26(3):580–594
46. Wang J, Xie X, Kuang J (2018) Microphone array speech enhancement based on tensor filtering methods. China Commun 15(4):141–152
47. Yilmaz S, Oysal Y (2010) Fuzzy wavelet neural network models for prediction and identification of dynamical systems. IEEE Trans Neural Netw 21(10):1599–1609
48. Zheng N, Zhang X (2019) Phase-aware speech enhancement based on deep neural networks. IEEE/ACM Trans Audio, Speech, Lang Process 27(1):63–76