IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 27 January 2023, accepted 11 February 2023, date of publication 27 February 2023, date of current version 3 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3249967

RESEARCH ARTICLE

U-Shaped Low-Complexity Type-2 Fuzzy LSTM Neural Network for Speech Enhancement

NASIR SALEEM[1], MUHAMMAD IRFAN KHATTAK[2], SALMAN A. ALQAHTANI[3], (Member, IEEE), ATIF JAN[2], IRSHAD HUSSAIN[2,3], MUHAMMAD NAEEM KHAN[2], AND MOSTAFA DAHSHAN[4], (Senior Member, IEEE)

[1]Department of Electrical Engineering, Faculty of Engineering and Technology, Gomal University, Dera Ismail Khan 29050, Pakistan
[2]Faculty of Electrical and Computer Engineering, University of Engineering and Technology Peshawar, Peshawar 25000, Pakistan
[3]Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[4]School of Computing, Mathematics and Engineering, Charles Sturt University, Bathurst, NSW 2795, Australia

Corresponding authors: Irshad Hussain (ee.irshad@gmail.com) and Salman A. Alqahtani (salmanq@ksu.edu.sa)

This Research is funded by Research Supporting Project Number (RSPD2023R585), King Saud University, Riyadh, Saudi Arabia.

ABSTRACT Speech enhancement (SE) aims to improve the intelligibility and perceptual quality of speech contaminated by noise signals through spectral or temporal changes. Deep learning models achieve speech enhancement and estimate the magnitude spectrum. This paper proposes a novel and computationally efficient deep learning model to enhance noisy speech. The model pre-processes the noisy speech magnitude by redistributing energy from high-energy voiced segments to low-energy unvoiced segments using an adaptive power law transformation while maintaining the total energy of the speech signals constant. A U-shaped fuzzy long short-term memory (UFLSTM) estimates the magnitude of a time-frequency (T-F) mask by using the pre-processed data. Residual connections to the similar-shaped layers are added to avoid gradient decay. Attention process is adopted by modifying the forget gate of UFLSTM. To make a causal speech enhancement system, the processing does not include any future audio frames. We compare the proposed speech enhancement to other deep learning models in different noisy environments with signal-to-noise ratios of 0 dB, 5 dB, and 10 dB. The experiments show that the proposed SE system outscores the competing deep learning models and considerably improves speech intelligibility and quality. In terms of STOI and PESQ, the LibriSpeech database improves results by (0.211) 21.1% and (0.95) 36.39%, respectively, over noisy speech in seen noisy conditions, and by (0.199) 19.9% and (0.94) 35.69% over noisy speech in unseen noisy conditions. Further, the cross-corpus analysis shows that proposed SE system performs better when trained with the DNS dataset as compared to the LibriSpeech, VoiceBank, and TIMIT datasets.

INDEX TERMS Energy redistribution, LSTM, residual connections, speech enhancement, and time-frequency masking.

I. INTRODUCTION

The SE recovers the components of clean speech from the noise degraded speech in a single microphone setting. SE systems mostly improve the quality of speech for real-time communication systems, multimedia content that has already been recorded, and better intelligibility to help automatic speech recognition systems and people listen better. SE methods such as spectral subtraction [1], [2], Wiener filtering [3], and statistical [4], [5] have been proposed in the last few decades. These methods were computationally efficient but inadequate in dealing with nonstationary noises. Broadly there are two main classes of speech enhancement, that is, Single-Channel (microphone) SE and Multi-Channel SE [70], [71]. The process that enhances a target's speech signal corrupted by background interferences with multiple microphones is termed as the Multi-Channel Speech Enhancement. One of the main

The associate editor coordinating the review of this manuscript and approving it for publication was Yiming Tang.

advantages of multi-channel speech enhancement is that are able not only to attenuate the background noise but also attain spatial filtering. However, this study focuses on the Single-Channel SE.

In recent years, deep learning has evolved into a mainstream method for SE, addressing the problems of conventional SE methods. The deep learning-based speech enhancement systems use models for speech and noise, where training of speech and/or noise signals estimates the model parameters. With a deep structure of hidden layers between input and output layers, deep learning constructs complex models for nonlinear relations and enables feature representation from the lower layers to model the complex input data. Given a speech dataset of the clean-noisy pairs, a neural model learns to transform the noisy magnitude spectra to their clean counterparts (mapping-based SE) or estimates the time-frequency masks (masking-based SE), such as the ideal binary mask (IBM) [6], [7], ideal ratio mask (IRM) [8], [9], and spectral magnitude mask (SMM) [10]. In spectral mapping, the models are trained using a direct mapping rule, where the noisy spectral features are learned to estimate the clean spectral features. But, over-smoothed spectrum are observed in the output speech. On the other hand, spectral-masking are more successful learning methods where gain parameters of target speech are multiplied to input noisy magnitude spectrum.

Many deep learning models, such as feed-forward DNNs (FDNNs) [11], [12], [13], convolutional neural networks (CNNs) [11], [14], [15], recurrent neural networks (RNNs) [16], [17], [18], [19], gated recurrent units (GRUs) [20], [21], and generative adversarial networks (GANs) [22], [23], [24], are used for SE. To learn the temporal dependencies of speech signals, FDNNs have been extended to RNNs. However, RNN undergoes gradient exploding and vanishing during back propagation through time (BPTT) and impulsively stops to learn the long-term dependencies. A gradient clipping approach can be used to solve the problem. Long-Short-Term Memory (LSTM) [25], [26], [27] enhances gradient vanishing by providing a memory cell framework that facilitates information flow across network layers. In the recent past, LSTM-based SE has gained much attention [25], [28], [29], [30], [31], [32], [33], [34]. The internal structure of LSTM went through various modifications in order to boost its ability to process data in a particular application. A peephole connection is proposed to improve past information learning ability by using the previous cell status as input [35]. To control the data flow between the memory cells, a depth gate is introduced to connect memory cells between network layers [36]. But the aforesaid LSTM variants usually increase the computational complexity.

Fuzzy logic systems, primarily type-2 fuzzy systems, have gained attention in high-order uncertainty applications during the last few years, particularly for prediction challenges including dynamic and non-stationary problems [37].Various sorts of study have been described to represent uncertainty using fuzzification [38], [39]. They have also been used in real-world applications through type-2 fuzzy logic systems, such as in modelling uncertainty, control, and predictions [40]. RNNs have also been used to model, estimate, and model time series data. It has been shown that RNNs, particularly the LSTM, are very capable of handling complex problems. The multi-layered interrelationships between short and long time series may be learned using the LSTM [41], [42], [43]. The advantage of an LSTM cell over a traditional recurrent unit is its cell memory [44].The cell vector of LSTM can capture the concept of forgetting a proportion of its previously recorded memory as well as a proportion of its new information. These characteristics can be applied to non-stationary problems such as SE [29]. Aside from the usual advantages of utilising RNNs for time series prediction, the LSTM network can also learn the temporal dependency from data. As a result, by definition, the LSTM may learn an arbitrary complex mapping from inputs to outputs, much as traditional nonlinear prediction algorithms. It faces difficulties, however, to represent the related uncertainty in non-stationary features [72].

This paper proposes a speech enhancement system with LSTM architectural changes. The UFLSTM with skip connections achieves improved SE performance. The proposed work is inspired by the recent success of the fuzzy-based LSTM network in nonstationary learning tasks. The study uses UFLSTM since we believe that such an arrangement shows better performance in SE. The proposed deep learning framework for SE estimates the time-frequency magnitude spectra. Usually, noise signals strongly mask the weak-energy speech components, which become indistinguishable, resulting in intelligibility and quality deterioration. Therefore, energy in noisy speech is redistributed from the rich energy voiced components to the weak-energy unvoiced components by applying adaptive power law transformation while maintaining the total energy of the speech signals constant. The energy redistribution increases the energy of weak components before actual magnitude estimation. The redistributed speech signals are fed to UFLSTM with residual (skip) connections to avoid gradient decay over the layers. Attention process is adopted by modifying the forget gate of UFLSTM. The speech spectra contain formant-frequencies that are leading in low-frequency segments and demonstrate a sparse distribution in the high-frequency segments. Thus, it is vital to distinguish the different spectral regions with different attention weights. Attention weights are computed to distinguish various spectral segments. No future information is used in the context vector used to train the LSTM model, which leads to a causal SCSE system suitable for real-time signal (speech) processing. The contributions of this study are summarized as. (i) a speech enhancement method is proposed with LSTM architectural changes, applying skip connections to avoid gradient decay and an attention process to differentiate various spectral regions. The ablation study suggested that the LSTM architectural change has improved the intelligibility and quality of noisy speech. (ii) The analysis has suggested that preprocessing with energy redistribution considerably improved the magnitude estimation, thereby increasing the speech intelligibility and quality.

The remaining paper is organized as follows. Section II explains the proposed SE based on the UFLSTM with residual (skip) connections and the attention process. Section III explains the preprocessing of noisy speech using energy redistribution. The experiments are presented in Section IV. Results and discussions are provided in Section V. And finally, the conclusions of this study are drawn in Section VI.

II. THE PROPOSED SPEECH ENHANCEMENT SYSTEM

For a given clean speech x_t and noise signal d_t, the noisy speech signal y_t is formed by the additive mixing as follows:

$$y_t = x_t + d_t \tag{1}$$

where $\{x, y, d\} \in \mathbb{R}^{N\times 1}$ and N shows number of speech frames. The SE systems aims to recover an estimate $\hat{x_t}$ of the clean speech x_t given y_t. The inputs to the UFLSTM are: $Y = [y_1, .., y_t, .., y_N]$, where y_t indicates the spectral magnitudes of the noisy speech at frame t. The high-level features h are extracted by the encoder from the input speech frames as:

$$h^K, h^Q = FLSTMEncoder(y_t) \tag{2}$$

where h^K and h^Q stand for the key and query, respectively. In this study, unidirectional LSTMs are modified, which shows a strong ability to model sequential data, leading to improved performance in speech enhancement. The attention process is applied with key and query as the input to create fixed-length context vectors:

$$C^t = Attention(h^K, h^Q) \tag{3}$$

The output w_t is a recovered enhanced speech signal $\hat{x_t}$ which takes the context vectors C^t, output of the encoder h^Q, and the noisy speech y_t, respectively.

$$w(t) = FLSTMDecoder(C^t, h_t^Q, y_t) \tag{4}$$

The proposed SE system based on UFLSTM is depicted in Fig. 1.

A. REVIEW OF THE LSTM NETWORK

The LSTMs were created to solve the shortcomings of traditional RNNs by improving gradient decay in a deep network. A cell state c_t of LSTM saves long-term memory according to following expressions as input gate, output gate, and forget gate, respectively [45]:

$$i_t = \sigma(\mathbf{W}_i) \times [C_{t-1}, h_{t-1}, x_t] + b_i \tag{5}$$

$$O_t = \sigma \mathbf{W}_o \times [C_{t-1}, h_{t-1}, x_t] + b_o \tag{6}$$

$$f_t = \sigma(\mathbf{W}_f) \times [C_{t-1}, h_{t-1}, x_t] + b_f \tag{7}$$

The gates adjusting the states and hidden cells of the LSTM can be expressed as:

$$k_t = tanh(\mathbf{W}_c \times [C_{t-1}, h_{t-1}, x_t] + b_c) \tag{8}$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes +k_t \tag{9}$$

where $\mathbf{W}_i$, $\mathbf{W}_f$, $\mathbf{W}_o$, are weight matrices of input, forget, and output gate associated with hidden states, x_t is input to the current timestamp, h_{t-1} is hidden state of the previous timestamp, C_{t-1} and C_t are the previous and current timestamps, respectively; whereas b_i, b_f, and b_o are the biased terms of the input, forget, and output gates, respectively.

B. TYPE-2 FUZZY SYSTEM

A type-2 fuzzy system can be represented as B which can be characterized by a type-2 mean function, $\mu_{B_{(k,v)}}$ where $k \in K$ and $v \in J_{k \subseteq [0,1]}$ [46], [47]:

$$B = (k, v), \mu_{B_{(k,v)}} | \forall_{k\in K}, \forall \in J_{k\subseteq[0,1]} \tag{10}$$

where $0 \leq \mu_{B_{(k,v)}} \leq 1$, K and J_k are the domains of the fuzzy set and the secondary mean function at k. The B is given as according to [48]:

$$B = \frac{\int_{k\in K} \int_{v\in J_k} \mu_{B(k,v)}}{k, vJ_k} \subseteq [0, 1] \tag{11}$$

where $\int\int$ shows union of overall admissible k and v [49]. The model considers the Takagi-Sugeno-Kang (TSK) fuzzy rule class, which has higher accuracy than the Mamdani rule [50]. The singleton fuzzifier was used in this study [51]. The Karnik-Mendel (K.M.) approach [52] was used for defuzzification.

C. TYPE-2 FUZZY LSTM CELL STRUCTURE

The sigmoid squashing function is applied to all gates of the FLSTM cell structure in this study, as outlined below:

$$\sigma = \frac{1}{1 + e^{-k}} \tag{12}$$

The inputs to the cell can be described as:

$$Net_{C_{(t)}} = \sum \mathbf{W}_{\lambda j} y(t-1) \tag{13}$$

The cell inputs pass through a non-linear function f_λ. The primary mean functions for all antecedents are defined using a Gaussian distribution with undetermined income as follows:

$$\mu_k^j(M_k) = exp\left[-\frac{1}{2}\left[\frac{N_k - n_k^j}{\sigma_k^j}\right]^2\right] \tag{14}$$

where $n_k^j \in \left[n_{k1}^j \cdot n_{k2}^j\right]$ shows an uncertain mean, k is the number of antecedents, N indicates the number of fuzzy rules, and σ_k^j shows the standard deviation, respectively. Figure 2 demonstrates the cell structure of the FLSTM. The cell state in FLSTM can now be defined as:

$$C_t = i_t C_{t-1} + i_t + i_t g(Net_{C_{(t)}}) \tag{15}$$

The output gate follows the expression, given as:

$$Net_{Out} \sum_n \mathbf{W}_{out} y^n(t-1) + \sum_{j=1}^{C_j} \mathbf{W}_{out} C_t \tag{16}$$

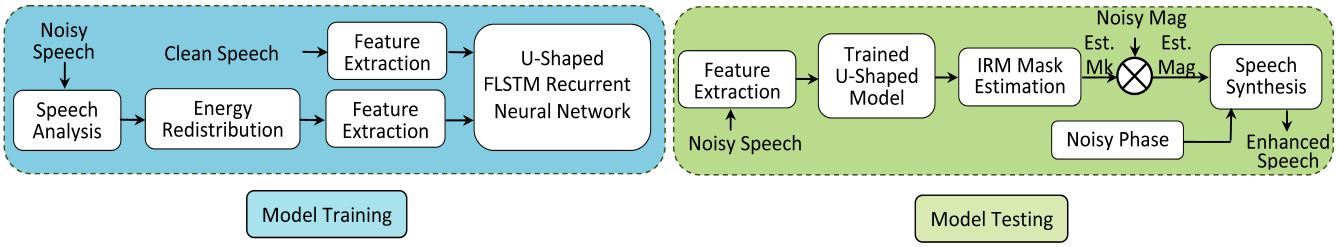

FIGURE 1. **The block diagram of the proposed speech enhancement.**

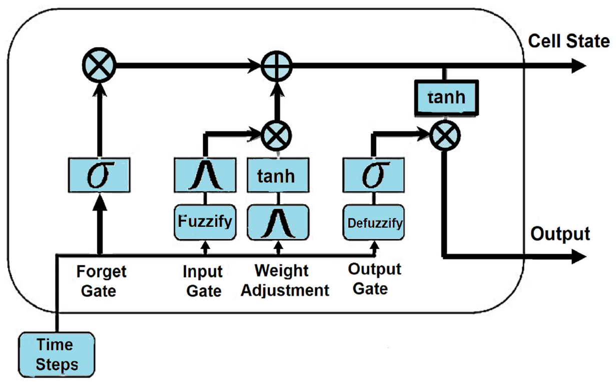

FIGURE 2. **The cell structure of FLSTM.**

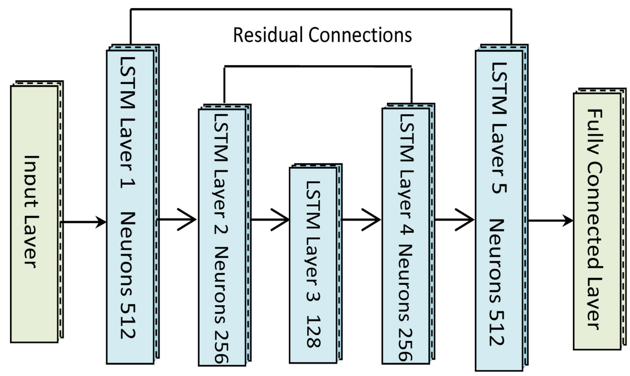

FIGURE 3. **The UFLSTM architecture.**

The activation applied at the output gate is given as:

$$Activation = f_\lambda(Net_{Out}) \tag{17}$$

Finally, the cell output is given as:

$$Output^C(t) = Net_{(Out)_{(t)}} C_{(t)} \tag{18}$$

D. UFLSTM

It is challenging to formulate deep RNNs due to saturation in the activation functions, which results in decaying gradients over the layers. LSTMs with gated mechanisms are able to capture sequential information in speech waveforms. The proposed FLSTM model can effectively deal with the limitations of RNN using new approaches. Firstly, the FLSTM model follows a U-shaped architecture, moving from left to right. For the left side, the time-steps are decreasing whereas the units are increasing. Similarly, on the right side, the time-steps are increasing while the units are decreasing. With this architecture, the FLSTM model can control the high-resolution features without memory overflow and with fewer network parameters. Secondly, residual connections are added to the similar-shape layers from the left to the right-side of the model. Therefore, the decaying gradient over the layers is improved. Thirdly, Attention process is added to the forget gate. The speech spectra contain formant frequencies that are dominant in low-frequency segments and demonstrate a sparse distribution in high-frequency segments. Therefore, it is important to distinguish the different spectral regions with different attention weights. Attention-weights are estimated to differentiate various spectral regions. The FLSTM model consists of five layers and two residual connections. Figure 3 illustrates the architecture of FLSTM. The FLSTM model learns a non-linear relation and transforms the noisy speech into a clean speech signal by estimating a magnitude of a time-frequency mask.

The normalized attention weight can be learnt in the attention parameter vector v using following expression:

$$\alpha_{(t)} = \frac{exp(\mathbf{v}^H y_{(t)})}{\sum_{\tau=1}^{T} exp(\mathbf{v}^H y_{(t)})} \tag{19}$$

where $\alpha_{(t)}$ and h are the weights for output at t^{th} time step y^t and H is a transpose operator, respectively.

III. ENERGY REDISTRIBUTION

The SE system estimates a clean magnitude spectrum from a noisy speech spectrum. The noise signals completely or partially mask the weak segments of the speech signals, and according to differences in the energy levels, the weak speech segments can get attenuated. This leads to deteriorated speech intelligibility and quality. Unvoiced stops and fricatives are mainly composed of low-energy segments. In noisy conditions, these low-energy segments undergo further energy degradation and become indistinguishable from noise segments. Thus, the energy levels of the weak speech segments need to be increased prior to magnitude spectrum estimation such that the low energy segments become distinctive from the noise segments. Speech signals are usually nonstationary;

therefore, a uniform amplification of the frames with different SNR levels may result in an annoying loudness level. As a result, energy is redistributed from the rich energy segments to the weak energy segments in such a way that the total energy of the signal remains intact. A threshold is computed from the average estimated noise energy to obtain noise-only frames such that energy redistribution is avoided in these frames. Here, energy is redistributed in a noisy signal using adaptive power law transformation [53].

The magnitude spectrum of the noisy speech $|Y(\omega)|$ is first segmented into 32 millisecond frames and 8 millisecond overlap is added to the segmented frames. An energy redistribution function is applied to the signal, given as:

$$E_{new}(m) = \beta[1 + d(m))]E_{old}(m)^{[1-(k*d(m))]\eta} \tag{20}$$

where $d(m)$ indicates energy deviation between the mean energy value μ of a noisy speech signal and old energy E_{old} of a speech signal. The parameters β and η control the signal energy in E_{new}. The $d(m)$ and E_{old} are given as:

$$d(m) = E_{old}(m) - \mu(m) \tag{21}$$

$$E_{old}(m) = \sum_{n=0}^{N-1} |y(n)|^2 \tag{22}$$

$$\mu(m) = \sum_{m=1}^{k} E_{old}(m) \tag{23}$$

The adaptive power law transformation function compensates the deviation between μ and E_{old} such that it transforms the old energy values to the new energy values. For a large deviation, the value of new energy is higher than old energy value, and vice versa. To estimate the energy controlling parameters (β and η), particle swarm optimization is applied to maintain the total energy of the noisy speech after distribution. Parameters β and η are initialized in the [0, 1] interval for all audio frames. The redistributed noisy speech signal is given as:

$$\bar{y}(n) = y(n) * \frac{E_{new}(m)}{E_{old}(m)} \tag{24}$$

The energy redistributed noisy speech is fed to the FLSTM for estimating a magnitude of time-frequency mask. Figure 4 demonstrates the speech spectrograms degraded by white noise at 5 dB SNR. The energy redistribution in the noisy speech signal is highlighted (with red boxes). The spectrograms clearly indicate that weak energy parts are boosted after energy transformation.

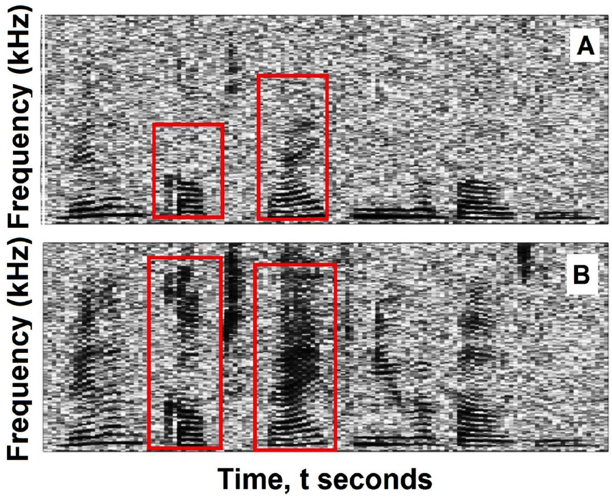

FIGURE 4. Impact of energy redistribution. The spectrogram of a noisy utterance: (A) before energy redistribution and (B) after energy redistribution. The red and blue boxes indicate the impacts of energy redistribution.

IV. EXPERIMENTS AND SETTINGS

A. DATASETS

The data during experiments was prepared from the LibriSpeech dataset [54]. The LibriSpeech dataset is divided into three subsets of 100/360/500 hours, respectively. In the experiments, a 100-hour clean subset is selected, which is uttered by 251 speakers (125 male and 126 female speakers). The clean subset of data is divided into training and testing sets. The clean subset is mixed with noise sources during training. To evaluate the performance of the proposed SE system in various noisy environments, 15 seen noise types are selected from the DEMAND database [55]. The types of noise seen include; N1: Airport, N2: Babble, N3: Buccaneer, N4: Car, N5: Destroyerengine, N6: Destroyerops, N7: Exhibition hall, N8: Factory, N9: Restaurant, N10: Street, N11: Subway, N12: Coffee Shop, N13: White, N14: Volvo, and N15: Pink Noise, respectively. In addition, two unseen noise types (factory2 and cafeteria noise) are also used to examine the SE system. The trained network is tested with factory2 and cafeteria noise, which are unknown to the trained model. During experiments, the noisy utterances are generated by mixing noises with clean utterances at three SNR levels, that is, 0 dB, 5 dB, and 10 dB, with a 5 dB step size. For model training, 2000 clean sentences uttered by 200 speakers (male and female) have been duplicated three times for each SNR level and mixed with 15 noise types. Therefore, a total of 12000 noisy utterances are used to train the SE model. Besides, a set of 2000 utterances from 100 speakers is used to test the SE model.

B. EVALUATION METRICS

To evaluate the performance of SE methods, we have used four objective metrics including Short-time objective intelligibility (STOI) [56], extended STOI (ESTOI) [57], Perceptual evaluation of speech quality (PESQ) [58] and Source-to-distortion ratio (SDR) [59].

C. NETWORK ARCHITECTURE

The proposed UFLSTM model consists of input layer, five unidirectional FLSTM layers with N units followed by a fully connected output layer with 257 units. From the input to the output layer, the quantity of neurons in the UFLSTM layers is given as: [1230/ 512/ 256/ 128/ 256/ 512/ 257] neurons. The number of epochs and learning rate are set to 100 and 0.001, respectively. All weights were randomly initialized and trained with mini-batches of 32 sequences by back-propagation through time with the Adam optimizer.

D. ACOUSTIC FEATURES

The input feature sets are composed of 15-d AMS (Amplitude Modulation Spectrogram, 31-d MFCC (Mel-Frequency Cepstral Coefficients), 13-d RASTA-PLP (Relative Spectral Transformed Perceptual Linear Prediction Coefficients) and 64-d GFE (Gammatone Filter-bank Energies). The GFE features are extracted from the Cochleagram, a T-F representation usually used in CASA (computational auditory scene analysis). A Cochleagram explains the working of the human auditory system. A 64-channel Gammatone Filter-bank is used for GFE features. Furthermore, delta feature coefficients are calculated and appended with the features. RASTAMAT toolbox is used to extract acoustic features.The acoustic features are extracted from the input speech at frame level. The frame length and shift are set to 20 millisecond and 10 millisecond, respectively. Auto-regressive moving average (ARMA) filter is used to further improve the temporal trajectories of features. A context window of eleven (11) frames is used to enclose temporal information. No future frames are used, thereby formulated a causal system which is appropriate for real-time speech processing. A 2706-d (246-d $\times$ 11) feature vector is applied to the FLSTM model. Zero mean and unit variance normalization is applied to all feature vectors before applying to the FLSTM.

E. LOSS FUNCTION AND SPECTRAL MASKS

Speech enhancement systems based on the spectral-masking estimate the masking parameters in order to restore the clean speech components by suppressing the background noise components in all time-frequency units. Spectra- masking is more effective as compared to the spectral-mapping since a time-frequency mask follows a bounded dynamic range; as a result, achieves fast convergence. There are several methods to learn the parameters of time-frequency mask which are based on the training-targets or optimization domain [73]. In mask approximation, a time-frequency mask is estimated such that MSE (mean square error) between reference and estimated mask is minimized, given as:

$$MSE_{(MA)} = \frac{1}{2B}\sum_{t=1}^{B-1}[(M_x(\omega) - \hat{M}_x(\omega))^2] \tag{25}$$

where $M_x(\omega)$ is a reference time-frequency mask whereas $\hat{M}_x(\omega)$ is the estimated time-frequency mask, respectively.

This study estimated the ideal ratio mask (IRM) and ideal amplitude mask (IAM), given as:

$$M_x^{IRM} = \sqrt{\frac{|X(\omega)|^2}{|X(\omega)|^2 + |D(\omega)|^2}} \tag{26}$$

$$M_x^{IAM} = \frac{|X(\omega)|}{|Y(\omega)|} \tag{27}$$

where $|X(\omega)|^2$ and $|D(\omega)|^2$ show the magnitude spectrums of the clean speech and noise signals, respectively. The dynamic range of the IRM is $R = [0, 1]$ whereas $R \geq 0$ for IAM in all time-frequency units.

V. RESULTS AND DISCUSSIONS

A. SPEECH ENHANCEMENT IN SEEN CONDITIONS

We first examine the performance of the proposed SE, denoted by UFLSTM. Table 1-2 shows the average STOI and ESTOI, PESQ, and SDR test scores for various noise types and SNR levels. The results are averaged over all noise types. It can be observed that the proposed SE system obtained better results as compared to other RNN models. The proposed UFLSTM with energy redistribution and architectural changes shows the highest test scores in terms of STOI, ESTOI, PESQ, and SDR in all noisy conditions. While compared to the baseline LSTM [60] and Attention-LSTM [25], the proposed UFLSTM obtained the best SE results. UFLSTM leads to the highest STOI, ESTOI, PESQ, and SDR improvements over the unprocessed noisy speech (UnP).

TABLE 1. Average STOI and ESTOI (in percentage) test scores in five seen noise types and SNR levels. The proposed SE system is denoted as UFLSTM.

Noise	Methods	STOI [57]			ESTOI [58]		
		0dB	5dB	10dB	0dB	5dB	10dB
Airport	Noisy (UnP)	62.7	69.5	78.3	30.8	38.8	50.7
	Att-LSTM [25]	85.8	88.5	91.7	68.9	74.1	80.2
	LSTM [60]	81.9	87.8	91.6	65.3	72.2	78.4
	UFLSTM	88.6	91.3	94.1	71.4	76.8	83.5
Babble	Noisy (UnP)	56.6	63.5	73	23.8	34.5	43.2
	Att-LSTM [25]	78.2	86.8	89.1	52.3	64.4	76.5
	LSTM [60]	80.3	85.1	88.6	54.1	67.2	77.1
	FLSTM	84.6	88.2	91.1	58.3	71.1	80.2
Factory	Noisy (UnP)	46.6	53.1	68.3	21.7	31.4	40.2
	Att-LSTM [25]	72.5	79.3	86.2	46.3	56.5	68.3
	LSTM [60]	71.3	79.1	85.6	47.2	55.8	67.9
	UFLSTM	74.8	83.4	91	50.4	59.3	71.1
Street	Noisy (UnP)	61.4	68.1	77.1	30.3	37.3	49.2
	Att-LSTM [25]	84.3	86.9	90.5	67.4	72.8	79.1
	LSTM [60]	80.5	87.4	91.3	66.1	70.5	76.9
	UFLSTM	87.1	90.7	93.4	70.6	74.4	82.2
Car	Noisy (UnP)	64.8	71.6	80.4	32.9	40	52.6
	Att-LSTM [25]	87.9	89.7	92.6	70.1	75.2	82.1
	LSTM [60]	84	88.9	90.9	67.4	73.4	79.7
	UFLSTM	89.7	92.4	94.6	73.5	77.9	84.4

The best STOI and ESTOI scores at low SNRs are obtained at car noise at SNR$\geq$0dB, i.e. STOI$\geq$ 89%, and ESTOI$\geq$73%, respectively (Table 1). Similarly, the best PESQ and SDR test scores at low SNRs are achieved at car noise at SNR$\geq$0dB, i.e. PESQ$\geq$2.6, and SDR$\geq$5.4, respectively (Table 2). During the analysis of results, it is examined that STOI test score with babble noise is improved from 56.6% with noisy speech (UnP) to 84.6% with UFLSTM and achieved (0.28) 28% improvement in STOI at 0 dB SNR. Similarly, ESTOI test score with babble noise is improved from 52.30% with Attention-LSTM to 58.30% with UFLSTM and achieved (0.06) 6% improvement in ESTOI at 0 dB SNR. At airport noise, the STOI test score is improved from 87.8% with LSTM to 91.3% with UFLSTM

TABLE 2. Average PESQ and SDR test scores in five seen noise types and SNR levels. The proposed SE system is denoted as UFLSTM.

Noise	Methods	PESQ [59]			SDR [60]		
		0dB	5dB	10dB	0dB	5dB	10dB
Airport	Noisy (UnP)	1.55	1.76	2.01	-3.70	-0.80	3.09
	Att-LSTM [25]	2.37	2.58	2.87	5.07	6.86	9.38
	LSTM [60]	2.28	2.58	2.85	4.98	6.71	9.27
	UFLSTM	2.63	2.76	2.99	5.38	7.21	9.62
Babble	Noisy (UnP)	1.49	1.69	1.92	-3.70	-0.80	3.11
	Att-LSTM [25]	2.31	2.52	2.77	2.21	4.64	8.75
	LSTM [60]	2.36	2.54	2.80	2.29	4.70	9.31
	UFLSTM	2.55	2.77	2.94	2.44	4.94	9.49
Factory	Noisy (UnP)	1.32	1.51	1.76	-3.90	-0.90	3.01
	Att-LSTM [25]	1.74	2.06	2.45	3.93	6.09	8.96
	LSTM [60]	1.71	2.00	2.41	3.88	6.01	8.87
	UFLSTM	1.98	2.23	2.65	4.11	6.33	9.20
Street	Noisy (UnP)	1.30	1.50	1.72	-4.10	-1.00	3.00
	Att-LSTM [25]	1.72	2.00	2.43	3.89	6.00	8.90
	LSTM [60]	1.69	1.99	2.39	3.83	5.96	8.81
	UFLSTM	1.96	2.20	2.61	4.02	6.29	9.14
Car	Noisy (UnP)	1.58	1.78	2.08	-3.50	-0.60	3.12
	Att-LSTM [25]	2.41	2.61	2.90	5.10	6.91	9.43
	LSTM [60]	2.32	2.59	2.88	5.02	6.78	9.31
	UFLSTM	2.69	2.79	2.98	5.41	7.25	9.69

and achieved (0.35) 3.5% improvement in STOI at 5 dB SNR. Also, ESTOI test score is improved from 74.1% with LSTM to 76.8% with UFLSTM, thereby achieved (0.27) 2.7% improvement in ESTOI at 5 dB SNR. The proposed SCSE achieved the best test STOI and ESTOI scores in factory noise at all SNRs.

The PESQ and SDR test results show that the PESQ test score at babble noise is improved from 2.36% with LSTM to 2.55% with UFLSTM and achieved 0.19 (8.05%) improvement in PESQ at 0 dB SNR. Similarly, the SDR test score at babble noise is improved from 2.21 dB and 2.29 dB with Attention-LSTM and LSTM to 2.44 dB with UFLSTM and achieved 0.23 dB and 0.25 dB improvements in SDR at 0 dB SNR. In factory noise, the PESQ test score is improved from 1.51 with the noisy speech to 2.23 with UFLSTM and achieved 32.28% improvement in PESQ at 5 dB SNR. Also, the SDR test score is improved from 3.01 dB with noisy speech to 9.20 dB with UFLSTM and achieved 6.19 dB improvements in SDR at 10 dB SNR. The proposed SE system achieved the best PESQ and SDR test scores in airport noise at all SNRs.

B. SPEECH ENHANCEMENT IN UNSEEN CONDITIONS

To evaluate the generalization performance of the proposed SE model, Table 3 provides STOI, ESTOI, PESQ, and SDR results for two unseen noise types (factory2 and cafeteria). The proposed SE model outscored the baseline and competing models with significant margins in unseen noisy conditions. We observed that UFLSTM achieved the best STOI, ESTOI, PESQ, and SDR scores because of its modified network architecture and energy distribution. Since the proposed LSTM model is treated with robust feature sets and residual connections, therefore, the performance is not drastically altered in unseen noisy conditions. For example, the average STOI test scores are improved from 65.4% to 86.5% with UFLSTM, thereby improved the STOI by (0.211) 21.1% over the noisy speech. At SNRs such as 0dB and 5dB, UFLSTM improved the STOI by (0.45) 4.50% and (0.23) 2.30% over the Atten-LSTM. Also, the average PESQ test scores are improved from 1.57 to 2.51 (37.69%) with UFLSTM, thereby improved the PESQ significantly over the noisy speech in unseen noisy conditions. The proposed LSTM model improved the STOI by 2.90% and 3.70% over Att-LSTM and LSTM. In addition, the proposed model improved the PESQ by factors 0.21 (8.33%) and 0.24 (9.52%) over Att-LSTM and LSTM. Finally, the proposed model for the SE improved the SDR by 0.31 dB and 0.30 dB over Att-LSTM and LSTM. The proposed FLSTM achieved the best performance in unseen noise types at various input SNRs.

C. COMPARISON WITH SOTA DL MODELS

To showcase the performance of the proposed SE system, several state-of-the-art (SOTA) deep learning models are selected, which include baseline LSTM (Chen et al.) [60], Attention-LSTM (Liang et al.) [25], CNN (Kounovsky and Malek) [61], GAN (Shah et al.) [62], DNN (Wang and Narayanan) [63], Transformer-DNN (DNN-TGSA by Kim et al.) [64], DeepResGRU (Saleem et al.) [21], LSTM-KF (Yu et al.) [65], and Deepxi (Zhang et al.) [66].

The average STOI and ESTOI test scores over 15 noise types and SNR levels are given in Table 4. In contrast to the competing DL models, the UFLSTM provides the highest STOI and ESTOI scores in all noisy conditions. For example, the STOI score with noisy speech increased from 66.13% to 87.17% with the proposed model and achieved a (0.21) 21.04% gain in STOI. Also, STOI scores with GAN, CNN, and DNN improved from 83.83%, 84.16%, and 79.59% to 87.17% and achieved (0.334) 3.34%, (0.301) 3.01%, and (0.758) 7.58% improvements, respectively. It is clear from the average scores that UFLSTM largely improved the PESQ and SDR test scores. For example, the PESQ score with noisy speech is increased from 1.66 to 2.61 with UFLSTM and improved the PESQ by 0.95 (35.71%). Moreover, PESQ scores with GAN, CNN, and DNN are improved from 2.40, 2.43, and 2.29 to 2.61, and improved PESQ test score by 0.21 (8.04%), 0.18 (6.89%), and 0.32 (12.26%), respectively. Finally, the SDR scores use to measure the distortion in the reconstructed speech suggested that the proposed model successfully attenuated the noise signals with better speech quality and intelligibility. For illustration, the SDR test score with noisy speech is increased from −0.47 dB to 6.52 dB with UFLSTM and improved SDR by 6.99 dB. In addition, SDR test scores with GAN, CNN, and DNN are improved from 6.17 dB, 6.29 dB, and 6.06 dB to 6.52 dB and improved SDR test scores by 0.35 dB, 0.23 dB, and 0.46 dB, respectively. The average STOI scores with DeepResGRU, DeepXi, and LSTM-KF are improved from 86%, 86.8%, and 84.1% to 87.2%, thereby obtained 1.2%, 0.40%, and 3.1% improvements over the recent deep learning models, respectively. On other hand, the average PESQ scores with DeepResGRU, DeepXi, and LSTM-KF are improved from 2.49, 2.43, and

TABLE 3. Average test scores in two unseen noise types (factory2 and cafeteria).

Methods	STOI				ESTOI				PESQ				SDR			
	0dB	5dB	10dB	Avrg	0dB	5dB	10dB	Avrg	0dB	5dB	10dB	Avrg	0dB	5dB	10dB	Avrg
Noisy (UnP)	57.7	64.6	74	65.4	25.5	33.4	45.2	34.7	1.36	1.52	1.82	1.57	-4.3	-1.3	2.72	-0.96
Att-LSTM [25]	77.9	84.2	88.6	83.6	55.2	64.5	74.6	64.8	2.05	2.26	2.62	2.31	3.07	5.41	8.65	5.71
LSTM [60]	76.9	83.3	88.2	82.8	54.9	64.6	73.9	64.5	2.02	2.25	2.61	2.29	3.05	5.35	8.77	5.72
UFLSTM	81.4	86.5	91.7	86.5	59.2	68.6	77.7	68.5	2.28	2.45	2.78	2.51	3.31	5.71	9.05	6.02

TABLE 4. Comparison with SOTA deep learning models.

Methods	STOI (in percentage)				ESTOI (in percentage)				PESQ				SDR (in dB)			
SNRs	0dB	5dB	10dB	Avrg	0dB	5dB	10dB	Avrg	0dB	5dB	10dB	Avrg	0dB	5dB	10dB	Avrg
Noisy (UnP)	58.6	65.3	74.5	66.2	26.1	33.9	45.7	35.2	1.45	1.65	1.89	1.66	-3.70	-0.80	3.10	-0.47
Att-LSTM [25]	78.8	84.9	89.0	84.3	55.8	65.0	75.0	65.3	2.14	2.38	2.69	2.40	3.73	5.86	9.03	6.21
LSTM [60]	77.8	84.0	88.6	83.5	55.5	65.1	74.4	65.0	2.11	2.37	2.68	2.39	3.71	5.80	9.15	6.22
CNN [61]	80.0	84.8	87.7	84.2	56.2	66.0	75.0	65.7	2.17	2.41	2.70	2.43	3.75	5.90	9.21	6.29
GAN [62]	77.7	84.6	89.2	83.8	55.9	65.9	75.8	65.9	2.13	2.39	2.67	2.40	3.69	5.84	8.98	6.17
DNN [63]	75.4	80.2	83.1	79.6	52.0	63.1	72.5	62.5	2.06	2.26	2.54	2.29	3.64	5.72	8.84	6.06
DNN-TGSA [64]	76.4	81.2	84.1	80.6	51.0	62.1	73.5	60.5	2.00	2.31	2.58	2.30	3.68	5.70	8.81	5.96
DeepResGRU [21]	81.8	85.5	90.7	86.0	57.3	65.4	77.6	66.7	2.29	2.49	2.71	2.49	3.73	5.89	9.29	6.30
LSTM-KF [65]	80.3	82.8	89.3	84.1	56.1	64.7	75.4	65.4	2.13	2.36	2.67	2.38	3.43	5.33	9.16	5.97
DeepXi [66]	81.2	84.7	90.2	86.8	56.9	65.1	77.3	66.3	2.21	2.40	2.68	2.43	3.86	5.66	9.24	6.25
UFLSTM	82.3	87.2	92.0	87.2	59.8	69.1	78.1	69.0	2.38	2.58	2.86	2.61	3.97	6.16	9.43	6.52

TABLE 5. Statistical analysis of average STOI and PESQ test scores at 95% confidence interval with $F_{Critical}$ is 3.09 and $P_{Critical}$ is 0.05.

SE Models	STOI				PESQ			
	0dB		5dB		0dB		5dB	
	PValue	FValue	PValue	FValue	PValue	FValue	PValue	FValue
Proposed → Noisy (UnP)	<0.0001	126.1	<0.0001	131.1	<0.0001	128.2	<0.0001	125.5
Proposed → Att-LSTM [25]	<0.0001	89.8	<0.0005	27.56	<0.0001	75.71	<0.0001	58.22
Proposed → LSTM [60]	<0.0001	95.13	<0.0005	26.22	<0.0001	85.12	<0.0001	80.91
Proposed → CNN [61]	<0.0016	19.71	<0.0005	27.03	<0.0002	37.94	<0.0021	18.23
Proposed → GAN [62]	<0.0001	89.43	<0.0005	28.11	<0.0001	65.09	<0.0001	62.03
proposed → DNN [63]	<0.0001	65.59	<0.0001	44.35	<0.0001	45.82	<0.0022	15.37
proposed → DNN-TGSA [64]	<0.0001	54.22	<0.0001	48.64	<0.0001	51.13	<0.0015	20.01

2.38 to 2.61%, thereby obtained 4.59%, 6.89%, and 8.81% improvements over the recent deep learning models, respectively. The SDR is improved by 0.22 dB, 0.27 dB, and 0.55 dB over DeepResGRU, DeepXi, and LSTM-KF, respectively. All the deep learning models for SE present better SDR test scores, indicating that deep learning models are successful in speech denoising. The SDR test scores are further improved by the proposed model. In comparison, the proposed model largely improved PESQ and STOI scores.

D. STATISTICAL ANALYSIS

The test scores suggested that the proposed SCSE method performed better at all SNR levels. However, to verify the effectiveness of the results, we conducted one-way **AN**alysis-of-**VA**riance (ANOVA) statistical tests. All the statistical tests are conducted at 95% confidence interval. The differences between test scores are deemed statistically significant if the probability is smaller than 0.05 ($p < 0.05$) and f_{Value} is greater than the critical value of f-distribution $f_{Value} > f_{Critical}$. Table 5 indicates the statistical tests at 95% confidence interval with $f_{Critical}$ equal to 3.09. It is clear from the tables that P_{Values} of the proposed model are smaller than 0.05 and $f_{Critical}$ is greater than 3.09. The statistical analysis suggested that the results achieved with the proposed models are statistical significant at all input SNRs.

E. COMPLEXITY ANALYSIS

Figure 5 shows the quantity of trainable parameters in the proposed LSTM model. With the U-Shaped strategy, the trainable parameters are reduced and our proposed UFLSTM is using 30% fewer parameters than conventional LSTM. Table 6 shows the general performance of the four training DL models in terms of time consumption. It can be stated that the implementation of 5-layered UFLSTM model is considerably faster than other three DL models. The training time can be increased with depth of the models; still, we find optimum results with depth of the proposed model. The time consumption was measured on the Intel Core i5-6500T Processor with NVIDIA GeForce GTX 1050. In addition, to measure inference time for a model, we have calculated the total number of computations the model in terms of FLOPs (Floating Point Operations). The FLOPs provide the complexity of the model, as given in Table 6.

F. SPECTRO-TEMPORAL ANALYSIS

To visualize the spectral regions of a speech processed by DL models, we show spectro-temporal plots. The clean

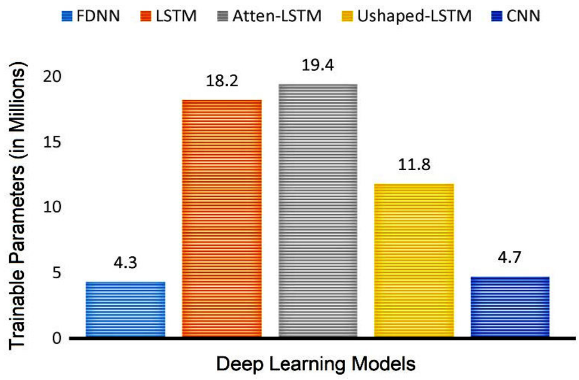

FIGURE 5. Model complexity in terms of trainable parameters.

TABLE 6. Complexity of the models.

DL Models	Training Time	FLOPs (Training)	FLOPs (Testing)
LSTM	19276 sec	0.262 × (3.1e16)	0.230 × (7.1e10)
Att-LSTM	20382 sec	0.243 × (3.1e16)	0.222 × (7.1e10)
UFLSTM	13265 sec	0.222 × (3.1e16)	0.213 × (7.1e10)

speech (Fig. 6A) is degraded by the street noise at 0dB SNR to generate the noisy speech (Fig. 6B). Principally, this is a challenging noisy condition since the noise signals emerge from various sources to degrade the target speech. The DNN enhanced speech is depicted in Fig. 6C, where background noise is notably reduced. The LSTM enhanced speech (Fig. 6D) shows less residual noise and distortion as compared to DNN. Figure 6E illustrates the enhanced speech produced by the Atten-LSTM where additional low distortion and residual noise is visible. The enhanced speech produced by the UFLSTM is plotted in Fig. 6F. It can be observed that less residual noise and speech distortion is present in the spectrogram of speech produced by the proposed model.

G. ABLATION STUDY

We conduct ablation studies to comprehend the proposed SE system. We have evaluated the full model (U-Shaped-FLSTM with Skips and energy redistribution) with (a) model using U-Shaped-FLSTM without skips and no energy redistribution (denoted by Proposed-1); (b) model using U-shaped FLSTM with skips but no energy redistribution (denoted by Proposed-2); (c) model using U-Shaped-FLSTM without skips but with energy redistribution (denoted by Proposed-3). To examine four models (Full model, Proposed-1, Proposed-2, and Proposed-3), we have used separate noise types selected from the DEMAND [55] database and trained each model on different noise type, also different utterances are used to train the models, that is, the training data for all three models is different for fair comparison. Table 7 shows the results (averaged over all SNRs) of various models for SE. The preprocessing with the energy redistribution performed better since energy levels of the weak speech segments are increased prior to the magnitude spectrum estimation such that low energy segments became distinctive to the noise segments.

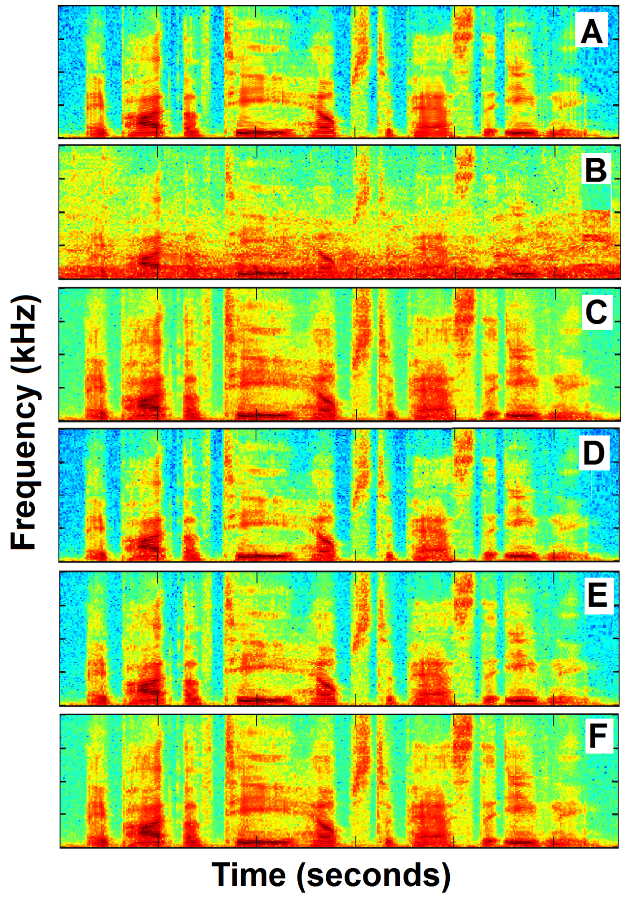

FIGURE 6. Spectro-temporal analysis. (A) Clean speech, (B) Noisy speech by adding babble noise at 0 dB SNR, (C) FDNN, (D) LSTM, (E) Atten-LSTM, and (F) Proposed UFLSTM.

TABLE 7. Ablation analysis of the proposed model.

Proposed Models	STOI (%)	ESTOI (%)	PESQ	SDR (dB)
Noisy Speech	70.16	40.11	1.77	-0.47
Proposed-1	81.23	63.14	2.28	5.04
Proposed-2	81.66	64.23	2.31	5.11
Proposed-3	82.33	66.47	2.35	5.19
Full Model	85.72	68.45	2.47	5.33

The full model has performed the best since preprocessing and architectural changes; we have observed significant improvements in the STOI, ESTOI, PESQ, and SDR with full model over noisy speech and other models. Figure 7 shows the STOI, ESTOI and SDR improvements for four proposed models. It is clear that with the energy redistribution and architectural changes, the proposed LSTM outscored at all SNRs.

H. CROSS CORPUS ANALYSIS

We have performed an experimental study to examine the cross-corpus generalization of neural models. The datasets including DNS [51], TIMIT [68], LibriSpeech [41], and VoiceBank [69] are examined for the speech quality and intelligibility. A speech dataset usually composed of different

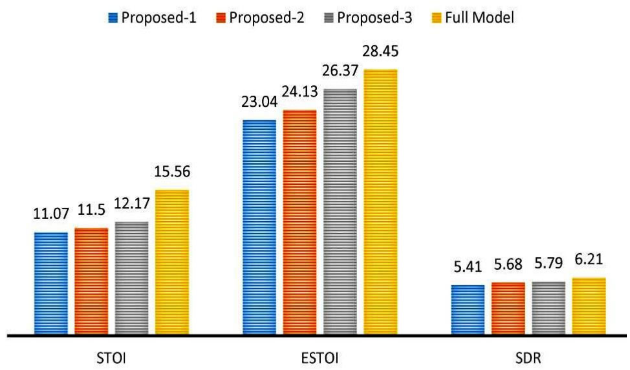

FIGURE 7. The STOI, ESTOI and SDR improvements for four proposed models.

TABLE 8. Cross-corpus analysis of the proposed model.

Datasets	STOI [57]	ESTOI [58]	PESQ [59]	SDR [60]
LibriSpeech [41]	83.72	66.45	2.37	5.24
VoiceBank [53]	83.61	65.29	2.25	5.03
DNS [51]	84.01	65.62	2.53	5.6
TIMIT [52]	83.58	64.55	2.26	5.06
Average	83.73	65.48	2.35	5.23

TABLE 9. Cross-training target analysis.

Datasets	IRM			IAM		
	STOI	PESQ	SDR	STOI	PESQ	SDR
LibriSpeech [41]	83.72	2.37	5.24	83.98	2.46	5.31
VoiceBank [53]	83.61	2.25	5.03	83.8	2.32	5.14
DNS [51]	84.01	2.53	5.6	84.32	2.64	5.74
TIMIT [52]	83.58	2.26	5.06	83.75	2.32	5.24
Average	83.73	2.35	5.23	83.96	2.44	5.36

utterances uttered by various speakers. The recording of the speech utterances is carried out in controlled environments for clean recordings and are appropriate for speech applications. The utterances are recorded in different controlled environments for different dataset which can lead to changed components in the speech utterances. For instance, an utterance recorded with different microphones by the same person can be very different in quality. To examine the influence of different speech datasets on the performance of the proposed model, we present Table 8 which shows the average PESQ and STOI values over all noise types and SNRs. The cross-corpus results indicate that proposed deep model performed better when trained with the DNS dataset as compared the LibriSpeech, VoiceBank, and TIMIT datasets.

I. CROSS TRAINING-TARGET ANALYSIS

In a set of experiments, we have examined the performance of two training-targets that is IRM and IAM. Table 9 shows the results of IRM and IAM in terms of STOI, PESQ, and SDR using four training datasets. In many noisy conditions, IAM training-target reasonably outperformed the IRM in terms of STOI, PESQ, and SDR, respectively. The average results indicate that the proposed model with IAM as training-target outperformed with DNS dataset. The average STOI, PESQ, and SDR with IAM are improved over IRM.

VI. SUMMARY AND CONCLUSION

The proposed deep learning-based SE system in this paper estimated the magnitude spectra from the noisy speech spectra using a U-Shaped-FLSTM framework with skips, attention process, and energy redistribution. The proposed system is compared with several deep learning SE models for performance comparison. Assessments are conducted under various noisy conditions for three SNR levels. The PESQ and SDR test scores indicated that the proposed method achieved improvements of (1.02–1.05) and (7.87–6.08) dB, respectively for airport and babble noisy conditions with reference to the noise contaminated speech. Similarly, STOI and ESTOI test scores indicated that the proposed method maintains the intelligibility under all situations, and STOI and ESTOI achieved large gains of (21.2–23.6)% and (37.13–36)%, respectively. The statistical analysis also indicated that the proposed method improved the quality without degrading intelligibility. The U-shaped LSTM model reduced the computational complexity of the usual LSTM by reducing the trainable parameters. Also, the replacement of the forgetting gate with attention- gate has improved the model performance. Regardless those DNNs outscore in SE application with their complex network models, yet require computationally less complex and efficient models for better performance. No future information was used by the proposed LSTM model which made it a causal speech enhancement system, appropriate for the real-time speech processing. Our proposed model has demonstrated reduced computational complexity in terms of the trainable parameters and outperformed the SOTA models used for comparison in this study. It is concluded that implementation of 5-layer U-Shaped-FLSTM model is considerably faster than other RNN models. The proposed DL model has improved the quality without degrading the speech intelligibility in adverse noisy conditions. Note that the proposed U-Shaped-FLSTM and related DL models showed repeated performance in terms of PESQ, STOI and SDR test scores in reference to the test scores of noisy speech, suggesting the potentials of DL for speech enhancement task. It is concluded from the spectrogram analysis that residual noise and speech distortions are reduced in the speech generated by the proposed DL model for SE. The cross-corpus results indicate that proposed deep model performed better when trained with the DNS dataset as compared the LibriSpeech, VoiceBank, and TIMIT datasets. In many noisy conditions, IAM training-target reasonably outperformed the IRM in terms of STOI, PESQ, and SDR, respectively. The average results indicate that the proposed model with IAM as training-target outperformed with DNS dataset. Based on the STOI, ESTOI, PESQ, and SDR results, the following inferences can be made.(1) Under various seen and unseen noisy conditions, the PESQ test scores stipulate that the proposed SE system achieved the highest improvements in quality compared to other deep learning models under all noisy conditions. (2) The STOI and ESTOI test scores over various seen and unseen noisy conditions indicate that the proposed SE system maintained the intelligibility to a greater extent as compared to all SE models under all noisy conditions. (3) The SDR test scores indicate that the proposed SE system minimized the distortion to a greater extent as compared to other models under various seen and unseen noisy conditions.

Our future research will focus on further improving the speech quality and intelligibility by using computationally less complex models in strong noisy conditions. In addition more robust features can further improve the performance.

REFERENCES

[1] T. Zhang, H. Wang, Y. Geng, X. Zhao, and L. Kong, "A speech separation algorithm based on the comb-filter effect," *Appl. Acoust.*, vol. 203, Feb. 2023, Art. no. 109197.
[2] N. Abajaddi, B. Mounir, I. Mounir, and A. Farchi, "Speech spectral subtraction in modulation domain," in *Proc. ICI2C*, vol. 357, 2022, p. 239.
[3] R. K. Jaiswal, S. R. Yeduri, and L. R. Cenkeramaddi, "Single-channel speech enhancement using implicit Wiener filter for high-quality speech communication," *Int. J. Speech Technol.*, vol. 25, no. 3, pp. 745–758, Sep. 2022.
[4] M. Kim and J. W. Shin, "Improved speech enhancement considering speech PSD uncertainty," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 30, pp. 1939–1951, 2022.
[5] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 6, pp. 1109–1121, Dec. 1984.
[6] M. I. Khattak, N. Saleem, J. Gao, E. Verdu, and J. P. Fuente, "Regularized sparse features for noisy speech enhancement using deep neural networks," *Comput. Electr. Eng.*, vol. 100, May 2022, Art. no. 107887.
[7] N. Saleem, E. Mustafa, A. Nawaz, and A. Khan, "Ideal binary masking for reducing convolutive noise," *Int. J. Speech Technol.*, vol. 18, no. 4, pp. 547–554, Dec. 2015.
[8] S. Zhao, B. Ma, K. N. Watcharasupat, and W.-S. Gan, "FRCRN: Boosting feature representation using frequency recurrence for monaural speech enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 9281–9285.
[9] F. Bao and W. H. Abdulla, "A new ratio mask representation for CASA-based speech enhancement," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 1, pp. 7–19, Jan. 2019.
[10] N. Saleem, M. I. Khattak, M. Al-Hasan, and A. B. Qazi, "On learning spectral masking for single channel speech enhancement using feedforward and recurrent neural networks," *IEEE Access*, vol. 8, pp. 160581–160595, 2020.
[11] H. Reddy, A. Kar, and J. Østergaard, "Performance analysis of low complexity fully connected neural networks for monaural speech enhancement," *Appl. Acoust.*, vol. 190, Mar. 2022, Art. no. 108627.
[12] I. Ullah, I. Hussain, K. Rehman, P. Wróblewski, W. Lewicki, and B. P. Kavin, "Exploiting the moth–flame optimization algorithm for optimal load management of the university campus: A viable approach in the academia sector," *Energies*, vol. 15, no. 10, p. 3741, May 2022.
[13] Y. Xu, J. Du, L.-R. Dai, and C.-H. Lee, "An experimental study on speech enhancement based on deep neural networks," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 65–68, Jan. 2014.
[14] M. Hasannezhad, H. Yu, W.-P. Zhu, and B. Champagne, "PACDNN: A phase-aware composite deep neural network for speech enhancement," *Speech Commun.*, vol. 136, pp. 1–13, Jan. 2022.
[15] Z. Ouyang, H. Yu, W.-P. Zhu, and B. Champagne, "A fully convolutional neural network for complex spectrogram processing in speech enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 5756–5760.
[16] T.-A. Hsieh, H.-M. Wang, X. Lu, and Y. Tsao, "WaveCRN: An efficient convolutional recurrent neural network for end-to-end speech enhancement," *IEEE Signal Process. Lett.*, vol. 27, pp. 2149–2153, 2020.
[17] W. Yu, J. Zhou, H. Wang, and L. Tao, "SETransformer: Speech enhancement transformer," *Cogn. Comput.*, vol. 14, no. 3, pp. 1152–1158, May 2022.
[18] V. Srinivasarao, "An efficient recurrent Rats function network (RRFN) based speech enhancement through noise reduction," *Multimedia Tools Appl.*, vol. 81, pp. 30599–30614, Apr. 2022.
[19] D. Takeuchi, K. Yatabe, Y. Koizumi, Y. Oikawa, and N. Harada, "Real-time speech enhancement using equilibriated RNN," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 851–855.
[20] Y. Wang, J. Han, T. Zhang, and D. Qing, "Speech enhancement from fused features based on deep neural network and gated recurrent unit network," *EURASIP J. Adv. Signal Process.*, vol. 2021, no. 1, pp. 1–19, Dec. 2021.
[21] N. Saleem, J. Gao, M. I. Khattak, H. T. Rauf, S. Kadry, and M. Shafi, "DeepResGRU: Residual gated recurrent neural network-augmented Kalman filtering for speech enhancement and recognition," *Knowl.-Based Syst.*, vol. 238, Feb. 2022, Art. no. 107914.
[22] S. Kanageswari, D. Gladis, I. Hussain, S. S. Alshamrani, and A. Alshehri, "Effective diagnosis of lung cancer via various data-mining techniques," *Intell. Autom. Soft Comput.*, vol. 36, no. 1, pp. 415–428, 2023.
[23] Y. Wang, G. Yu, J. Wang, H. Wang, and Q. Zhang, "Improved relativistic cycle-consistent GAN with dilated residual network and multi-attention for speech enhancement," *IEEE Access*, vol. 8, pp. 183272–183285, 2020.
[24] S. Routray and Q. Mao, "Phase sensitive masking-based single channel speech enhancement using conditional generative adversarial network," *Comput. Speech Lang.*, vol. 71, Jan. 2022, Art. no. 101270.
[25] R. Liang, F. Kong, Y. Xie, G. Tang, and J. Cheng, "Real-time speech enhancement algorithm based on attention LSTM," *IEEE Access*, vol. 8, pp. 48464–48476, 2020.
[26] J. Oruh, S. Viriri, and A. Adegun, "Long short-term memory recurrent neural network for automatic speech recognition," *IEEE Access*, vol. 10, pp. 30069–30079, 2022.
[27] R. LeBlanc and S. A. Selouani, "A two-stage deep neuroevolutionary technique for self-adaptive speech enhancement," *IEEE Access*, vol. 10, pp. 5083–5102, 2022.
[28] T. Lan, W. Ye, Y. Lyu, J. Zhang, and Q. Liu, "Embedding encoder–decoder with attention mechanism for monaural speech enhancement," *IEEE Access*, vol. 8, pp. 96677–96685, 2020.
[29] D. Shan, X. Zhang, C. Zhang, and L. Li, "A novel encoder–decoder model via NS-LSTM used for bone-conducted speech enhancement," *IEEE Access*, vol. 6, pp. 62638–62644, 2018.
[30] A. Nicolson and K. K. Paliwal, "Masked multi-head self-attention for causal speech enhancement," *Speech Commun.*, vol. 125, pp. 80–96, Dec. 2020.
[31] A. Pandey and D. Wang, "Dual-path self-attention RNN for real-time speech enhancement," 2020, *arXiv:2010.12713*.
[32] L. Sun, J. Du, L.-R. Dai, and C.-H. Lee, "Multiple-target deep learning for LSTM-RNN based speech enhancement," in *Proc. Hands-Free Speech Commun. Microphone Arrays (HSCMA)*, 2017, pp. 136–140.
[33] T. Gao, J. Du, L.-R. Dai, and C.-H. Lee, "Densely connected progressive learning for LSTM-based speech enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 5054–5058.
[34] M. Strake, B. Defraene, K. Fluyt, W. Tirry, and T. Fingscheidt, "Speech enhancement by LSTM-based noise suppression followed by CNN-based speech restoration," *EURASIP J. Adv. Signal Process.*, vol. 2020, no. 1, pp. 1–26, Dec. 2020.
[35] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[36] S. Zaman, L. U. Khan, I. Hussain, and L. Mihet-Popa, "Fast computation of highly oscillatory ODE problems: Applications in high-frequency communication circuits," *Symmetry*, vol. 14, no. 1, p. 115, Jan. 2022.
[37] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—I," *Inf. Sci.*, vol. 8, no. 3, pp. 199–249, 1975.
[38] H. Hagras, "Type-2 fuzzy logic controllers: A way forward for fuzzy systems in real world environments," in *Proc. IEEE World Congr. Comput. Intell.* Berlin, Germany: Springer, Jun. 2008 pp. 181–200.
[39] P. C. de Lima Silva, H. J. Sadaei, R. Ballini, and F. G. Guimaraes, "Probabilistic forecasting with fuzzy time series," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 8, pp. 1771–1784, Aug. 2020.
[40] R. Navarro-Almanza, R. Juárez-Ramírez, G. Licea, and J. R. Castro, "Automated ontology extraction from unstructured texts using deep learning," in *Intuitionistic and Type-2 Fuzzy Logic Enhancements in Neural and Optimization Algorithms: Theory and Applications*. Cham, Switzerland: Springer, 2020, pp. 727–755.
[41] C. Li, J. Yi, H. Wang, G. Zhang, and J. Li, "Interval data driven construction of shadowed sets with application to linguistic word modelling," *Inf. Sci.*, vol. 507, pp. 503–521, Jan. 2020.

[42] N. Anh, S. Suresh, M. Pratama, and N. Srikanth, "Interval prediction of wave energy characteristics using meta-cognitive interval type-2 fuzzy inference system," *Knowl.-Based Syst.*, vol. 169, pp. 28–38, Apr. 2019.

[43] C. Li, C. Zhou, W. Peng, Y. Lv, and X. Luo, "Accurate prediction of short-term photovoltaic power generation via a novel double-input-rule-modules stacked deep fuzzy method," *Energy*, vol. 212, Dec. 2020, Art. no. 118700.

[44] Y.-Y. Lin, J.-Y. Chang, N. R. Pal, and C.-T. Lin, "A mutually recurrent interval type-2 neural fuzzy system (MRIT2NFS) with self-evolving structure and parameters," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 3, pp. 492–509, Jun. 2013.

[45] F. Weninger, H. Erdogan, S. Watanabe, E. Vincent, J. L. Roux, J. R. Hershey, and B. Schuller, "Speech enhancement with LSTM recurrent neural networks and its application to noise-robust ASR," in *Proc. Int. Conf. Latent Variable Anal. Signal Separat.* Cham, Switzerland: Springer, 2015, pp. 91–99.

[46] Y. Chen, C. Li, and J. Yang, "Design and application of Nagar–Bardini structure-based interval type-2 fuzzy logic systems optimized with the combination of backpropagation algorithms and recursive least square algorithms," *Expert Syst. Appl.*, vol. 211, Jan. 2023, Art. no. 118596.

[47] D. Wu, "On the fundamental differences between interval type-2 and type-1 fuzzy logic controllers," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 5, pp. 832–848, Oct. 2012.

[48] D. Wu and J. M. Mendel, "A vector similarity measure for interval type-2 fuzzy sets," in *Proc. IEEE Int. Fuzzy Syst. Conf.*, Jun. 2007, pp. 1–6.

[49] J. M. Mendel, R. I. John, and F. Liu, "Interval type-2 fuzzy logic systems made simple," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 6, pp. 808–821, Dec. 2006.

[50] W. Chen and Y. Zou, "Group decision making under generalized fuzzy soft sets and limited cognition of decision makers," *Eng. Appl. Artif. Intell.*, vol. 87, Jan. 2020, Art. no. 103344.

[51] W. Peng, C. Zhou, C. Li, X. Deng, and G. Zhang, "Double-input rule modules stacked deep interval type-2 fuzzy model with application to time series forecasting," *Int. J. Fuzzy Syst.*, vol. 23, no. 5, pp. 1326–1346, Jul. 2021.

[52] Y. Chen, "Study on centroid type-reduction of interval type-2 fuzzy logic systems based on noniterative algorithms," *Complexity*, vol. 2019, pp. 1–12, Apr. 2019.

[53] T. Lavanya, T. Nagarajan, and P. Vijayalakshmi, "Multi-level single-channel speech enhancement using a unified framework for estimating magnitude and phase spectra," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 28, pp. 1315–1327, 2020.

[54] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: An ASR corpus based on public domain audio books," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 5206–5210.

[55] J. Thiemann, N. Ito, and E. Vincent, "The diverse environments multi-channel acoustic noise database (DEMAND): A database of multichannel environmental noise recordings," in *Proc. Meetings Acoust.*, vol. 19, no. 1. Melville, NY, USA: Acoustical Society of America, 2013, Art. no. 035081.

[56] F. Ahmad, A. Ahmad, I. Hussain, G. Muhammad, Z. Uddin, and S. A. AlQahtani, "Proactive caching in D2D assisted multitier cellular network," *Sensors*, vol. 22, no. 14, p. 5078, Jul. 2022.

[57] J. Jensen and C. H. Taal, "An algorithm for predicting the intelligibility of speech masked by modulated noise maskers," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 11, pp. 2009–2022, Nov. 2016.

[58] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ)—A new method for speech quality assessment of telephone networks and codecs," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, May 2001, pp. 749–752.

[59] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 1, pp. 229–238, Jan. 2008.

[60] J. Chen, Y. Wang, S. E. Yoho, D. Wang, and E. W. Healy, "Large-scale training to increase speech intelligibility for hearing-impaired listeners in novel noises," *J. Acoust. Soc. Amer.*, vol. 139, no. 5, pp. 2604–2612, 2016.

[61] T. Kounovsky and J. Malek, "Single channel speech enhancement using convolutional neural network," in *Proc. IEEE Int. Workshop Electron., Control, Meas., Signals Appl. Mechatronics (ECMSM)*, May 2017, pp. 1–5.

[62] N. Shah, H. A. Patil, and M. H. Soni, "Time-frequency mask-based speech enhancement using convolutional generative adversarial network," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Nov. 2018, pp. 1246–1251.

[63] Y. Wang, A. Narayanan, and D. Wang, "On training targets for supervised speech separation," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 22, no. 12, pp. 1849–1858, Dec. 2014.

[64] J. Kim, M. El-Khamy, and J. Lee, "T-GSA: Transformer with Gaussian-weighted self-attention for speech enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 6649–6653.

[65] H. Yu, W.-P. Zhu, and B. Champagne, "Speech enhancement using a DNN-augmented colored-noise Kalman filter," *Speech Commun.*, vol. 125, pp. 142–151, Dec. 2020.

[66] Q. Zhang, A. Nicolson, M. Wang, K. K. Paliwal, and C. Wang, "DeepMMSE: A deep learning approach to MMSE-based noise power spectral density estimation," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 28, pp. 1404–1415, 2020.

[67] Y. Ju, W. Rao, X. Yan, Y. Fu, S. Lv, L. Cheng, Y. Wang, L. Xie, and S. Shang, "TEA-PSE: Tencent-ethereal-audio-lab personalized speech enhancement system for ICASSP 2022 DNS challenge," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 9291–9295.

[68] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, "DARPA TIMIT acoustic-phonetic continous speech corpus CD-ROM. NIST speech disc 1-1.1," NASA STI/Recon Tech. Rep. n, 93, 1993, p. 27403.

[69] C. Veaux, J. Yamagishi, and S. King, "The voice bank corpus: Design, collection and data analysis of a large regional accent speech database," in *Proc. Int. Conf. Oriental COCOSDA Conf. Asian Spoken Lang. Res. Eval. (O-COCOSDA/CASLRE)*, Nov. 2013, pp. 1–4.

[70] S. Dowerah, R. Serizel, D. Jouvet, M. Mohammadamini, and D. Matrouf, "Joint optimization of diffusion probabilistic-based multichannel speech enhancement with far-field speaker verification," in *Proc. IEEE SLT*, Jan. 2023, pp. 1–9.

[71] W. Xie, X. Xiang, X. Zhang, and G. Liu, "A pre-separation and all-neural beamformer framework for multi-channel speech separation," *Symmetry*, vol. 15, no. 2, p. 261, Jan. 2023.

[72] K. Smagulova, O. Krestinskaya, and A. P. James, "A memristor-based long short term memory circuit," *Analog Integr. Circuits Signal Process.*, vol. 95, no. 3, pp. 467–472, 2018.

[73] I. Hussain, I. Ullah, W. Ali, G. Muhammad, and Z. Ali, "Exploiting lion optimization algorithm for sustainable energy management system in industrial applications," *Sustain. Energy Technol. Assessments*, vol. 52, Aug. 2022, Art. no. 102237.

NASIR SALEEM received the B.S. degree in telecommunication engineering from the University of Engineering and Technology, Peshawar, Pakistan, in 2008, the M.S. degree in electrical engineering from CECOS University, Peshawar, in 2012, and the Ph.D. degree in electrical engineering, with specialization in digital speech processing and deep learning, from the University of Engineering and Technology, in 2021. From 2008 to 2012, he was a Senior Lecturer with the Institute of Engineering Technology, Gomal University, where he was involved in teaching and research. He is currently an Assistant Professor with the Department of Electrical Engineering, Faculty of Engineering and Technology, Gomal University. He has published more than 50 research articles in well-known venues, such as Elsevier, Springer, and IEEE, so far. His current research interests include human–machine interaction, speech enhancement, speech and video processing, and machine learning applications. He is also involved in academic activities such as reviewing articles from several well-known venues, including IEEE/ACM Transactions on Audio Speech and Language Processing, IEEE Transactions on Human-Machine Systems, IEEE Transactions on Artificial Intelligence, IEEE Signal Processing, IEEE Multimedia, and IEEE Access.

MUHAMMAD IRFAN KHATTAK received the B.S. degree in electrical engineering from the University of Engineering and Technology (UET), Peshawar, Pakistan, in 2004, and the Ph.D. degree in engineering from Loughborough University, U.K., in 2010. Since 2004, he has been involved in teaching and research with UET Peshawar. He is currently an Associate Professor with the Department of Electrical Engineering, Faculty of Computer and Electrical Engineering, UET Peshawar. He has published more than 100 research articles in well-known venues so far, such as Elsevier, Springer, and IEEE. His current research interests include on-body communications, speech enhancement, speech processing, antenna design, and machine learning applications.

SALMAN A. ALQAHTANI (Member, IEEE) is currently a Professor with the Department of Computer Engineering, King Saud University, Riyadh. His interests include performance evaluation and analysis of high-speed packet switched networks, system model and simulations, and integration of heterogeneous wireless networks. Mainly his focus is on the design and optimization of 5G MAC layers, closed form mathematical performance analysis, energy efficiency, and resource allocation and sharing strategies. He has authored two scientific books and authored/coauthored around 76 journal and conference papers on the topics of his research interests, since 2004. His current research interests include the area of 5G networks, broadband wireless communications, radio resource management for 4G and beyond networks (call admission control, packet scheduling, and radio resource sharing techniques), cognitive and cooperative wireless networking, small cell and heterogeneous networks, self-organizing networks, SDN/NFV, 5G network slicing, smart grid, intelligent IoT solutions for smart cities, dynamic spectrum access, coexistence issues on heterogeneous networks in 5G, industry 4.0 issues, the Internet of Everything, mobile edge and fog computing, and cyber sovereignty. He serves as a reviewer for several national and international journals.

ATIF JAN received the B.S. and M.Sc. degrees in electrical engineering and the Ph.D. degree from the University of Engineering and Technology (UET), Peshawar, Pakistan, in 2011, 2015, and 2022, respectively. He is involved in teaching and research with the University of Engineering and Technology. He is currently a Lecturer with the Department of Electrical Engineering, Faculty of Computer and Electrical Engineering, UET Peshawar. His current research interests include video processing, anomaly detection, speech processing, and machine learning applications.

IRSHAD HUSSAIN received the M.Sc. degree in electrical engineering from the University of Engineering and Technology Peshawar (UET Peshawar) Peshawar, Pakistan. He worked as a Research Associate with the Center for Intelligent Systems and Network Research, from 2014 to 2016. He is currently a Visiting Researcher with the College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia. He is also teaching and involved in various research and development activities with the Faculty of Electrical and Computer Engineering, UET Peshawar., Pakistan. He has published more than 30 research papers in national/international journals and conferences. His research interests include fault tolerant systems, optimal control of multi-agent systems, control of unmanned vehicles, distributed systems, smart water management systems, green IoT, the IoET, the IoMT, consensus control of MAS and smart grid network and cyber-physical control of networked/distributed systems especially the Internet of Things, the Internet of Vehicles, and cyber secure smart power systems. He is offering his services as a Reviewer in many reputed international journals, such as IEEE Transactions on Sustainable Energy, IEEE Transaction on Industrial Informatics, *Applied Sciences*, *Entropy*, *International Transactions on Electrical Energy Systems* (Wiley), *Journal of Network and Computer Applications*, and IEEE Transactions on Systems, Man, and Cybernetics. He is a guest editor of many reputed journals.

MUHAMMAD NAEEM KHAN received the B.Sc., M.Sc., and Ph.D. degrees in engineering. He is currently an Assistant Professor with the Department of Electrical Engineering, University of Engineering and Technology, Peshawar, where he is involved in teaching core engineering courses and conducting research. He has more than 13 years of teaching experience and published several papers in well reputed international journals and conferences.

MOSTAFA DAHSHAN (Senior Member, IEEE) received the Bachelor of Computer Engineering degree from Cairo University, Egypt, in 1999, the M.Sc. degree in telecom systems and the Ph.D. degree in electrical and computer engineering from the University of Oklahoma, USA, in 2002 and 2006, respectively. From 2002 and 2005, he worked as a Graduate Research Assistant at the University of Oklahoma, administering the Interoperability Laboratory network and assisting students, where he was worked as an IT Specialist, developing web applications and providing IT service and support, from 2005 to 2007. In 2007, he joined King Saud University, Saudi Arabia, as an Assistant Professor of computer engineering. In recognition of his academic performance, he was promoted as an Associate Professor, in 2017. In October 2018, he joined the School of Computing and Mathematics, Charles Sturt University, as a Lecturer in computing. He has taught several subjects in the areas of computer networks, network security, and network programming. He has published high-quality journals and conferences papers, supervised theses and capstone projects, participated in academic accreditation committees, and developed several solutions to enhance the academic productivity and teaching experience. He is a Member and a Certified Professional of the Australian Computer Society.