Neural Kalman Filter Application to Speech Enhancement

Amir Moghtadaei Rad[1*]

Abstract– Speech enhancement is introduced as a requirement to increase the quality of communication systems' operations. There is a broad range of improvements for speech recognition systems in aviation, military, telecommunication, and cellular environments. Also, speech quality confirmation can be important in decrement of audience boredom in noisy environments. In this paper, speech quality enhancement and its intelligibility by extended Kalman filter is introduced. It is obvious that for speech detection the right estimation is needed, but an important subject is that the linear filter is unable to estimate nonlinear systems while most real systems like voice systems have nonlinear architecture. Hence, according to the extended Kalman filter method, modelling and estimation of voice signal with nonlinearity assumption that leads to speech enhancement, is executed and its results are shown.

Keywords: Speech Enhancement Extended Kalman Filter, Neural Network.

1. Introduction

Since 1960 the Kalman filter was introduced, always researchers tried to use this filter for linear systems dynamics estimation, however, the main problem was that most real systems have nonlinear architecture while the linear Kalman filter is unable to estimate those systems. For instance, voice systems are nonlinear systems that because of voice nonlinearity architecture and linear Kalman filter restriction could not be estimated until an extended Kalman filter was introduced.

This filter creates optimized estimations of nonlinear system states with nonlinear system modelling using a neural network[1-3].

Speech enhancement aims to improve quality and intelligibility. Quality refers to the amount of noise free in speech and intelligibility refers to the percentage number of words understand in the sentence. Speech enhancement involves noise estimation as crucial part. Many researchers represent different ideas for nonlinear system methods and each one has its own advantages and disadvantages, but a few nonlinear methods are introduced for voice enhancement. In this paper, we introduce a simple algorithm to identify nonlinear voice systems by neural network and then with an extended Kalman filter (EKF) estimation algorithm, improve noisy voice signals. Moreover, at the end of the paper; clean, noisy, and enhanced signals are compared. In conclusion, a comparative analysis demonstrates significant improvement in the proposed method.

1* Corresponding Author: Department of Computer and Electrical Engineering, Has.C.,Islamic Azad University, Alborz, Iran.

Email: Amir.Moghtadaei@iau.ac.ir

Received: 2025.02.06; Accepted: 2025.05.14

2. Speech Identification

In the nonlinear system identification field, voice is also comprised, because nonlinear system dynamics are variable at each time and unpredictable certainly, therefore linear system identification methods are not implemented. The only suitable way for this modelling is by applying neural networks.

Because of learning ability, neural networks and fuzzy neural networks are the only methods in nonlinear systems identification and prediction. These networks can parallel by main system and learn system behaviour intelligently after some iteration. Then, they can operate like the main system and they can be replaced with a modelled system. This extra ability makes possibility to model a lot of nonlinear and difficult real systems dynamics with neural networks[4],[5].

2.1 Voice State Space Representation

Figure 1 showsx_kSignalsare the noisy voice production nonlinear system states andv_kProcess noise is the system input;the output is y_k Which is a noisy signal destroyedbyn_kMeasurement noise.

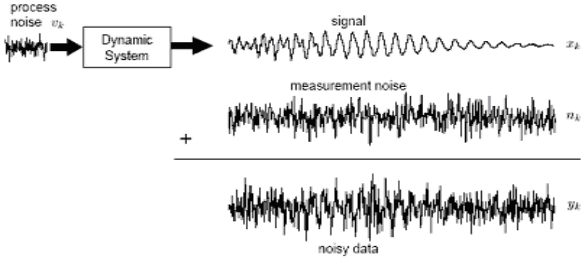


Fig. 1: Data is processed with an AR model and destroyed by added measurement noise.

The system state space representation is as below[6]:

$$X_k = F(X_{k-1}, W) + B.V_k \quad (1)$$

$$\begin{bmatrix} x_k \\ x_{k-1} \\ \vdots \\ x_{k-M+1} \end{bmatrix} = \begin{bmatrix} f(x_{k-1}, \dots, x_{k-M}, w) \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \ddots & 0 & \ddots \\ 0 & 0 & 1 & 0 \end{bmatrix} . \begin{bmatrix} x_{k-1} \\ \vdots \\ x_{k-M} \end{bmatrix} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} . v_k \quad (2)$$

And system output is calculated as below:

$$Y_k = C \ . \ X_k + n_k \quad (3)$$

$$y_k = [1 \quad 0 \quad \dots \quad 0] \ . \begin{bmatrix} x_k \\ x_{k-1} \\ \vdots \\ x_{k-M+1} \end{bmatrix} + n_k \quad (4)$$

2.2 Neural Autoregressive Identification

In this system identification method, is supposed that only system output is known and system input is unknown hence, only output is accessible for identification.

In this communication system, output, that is clean voice, is considered as the network target in the mentioned method, and output feedback is considered as system input:

$$y(t) = f(y(t-1), y(t-2), \dots, y(t-n_y)) \quad (5)$$

Then, the network begins to train as its block representation is shown in Figure 2:

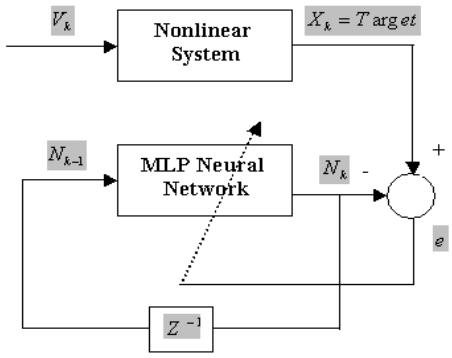


Fig.2: System Identification Loop

For system identification, an MLP neural network is paralleled with the system, and applied system input is used as network input. Also, the system output is defined as a network target signal.

Then with the recurrent error back propagation (EBP) method, the neural network is trained whereas in each iteration, network output is compared with the target, and error is an important item of feedback to the neural network. Since x_k Vector length, that is a voice signal, is 39000, thus at least 39000 iterations are needed for the neural network [7-9].

2.3 Applied Neural Network

To build matrix$A$That is applied in the extended Kalman filter (EKF) programis needed to linearize nonlinear function. F which $F(x_k, w)$ Isthe network output. This linearization is performed with Jacobin matrix calculation. The Jacobin matrix array consistsof input difference divided by output difference. Therefore,matrix ADimension is related to the number of inputs.

In this paper, 5 inputs are assigned to the network; thus, the matrix.ADimension is 5.

The network scheme as a feed-forward network, with 5 inputs, 5 hidden layers with tanh fire function, and 1 output with linear fire function that is defined as an order 5 nonlinear model (Figure 3)[10]:

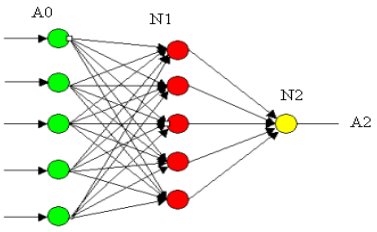


Fig. 3: MLP Neural Network Structure

In the neural network training condition, the important point is that the MLP network is trained with the EBP method.

In this training method, after the first and second network layer calculations, network output is obtained and compared with the target signal; then, the calculated error is feedbacked to the network and is affected by updating weight and bias neural network vector. The network begins a new recurrent iteration whereas posterior time output (y_{k-1})is applied as a prior input (x_k).

Also forgetting factor is defined as an option item that if selected by the operator, weight, and bias updating vector are changed, however, this subject does not change the learning process and only the update equation is varied slightly.

Neural network block-diagram representation by applying this method is as Figure 4:[11]

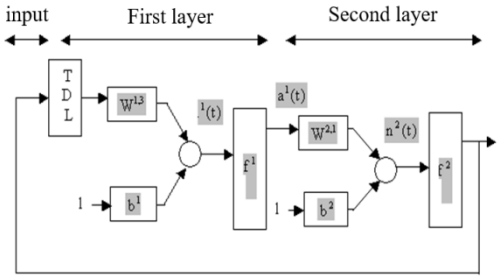
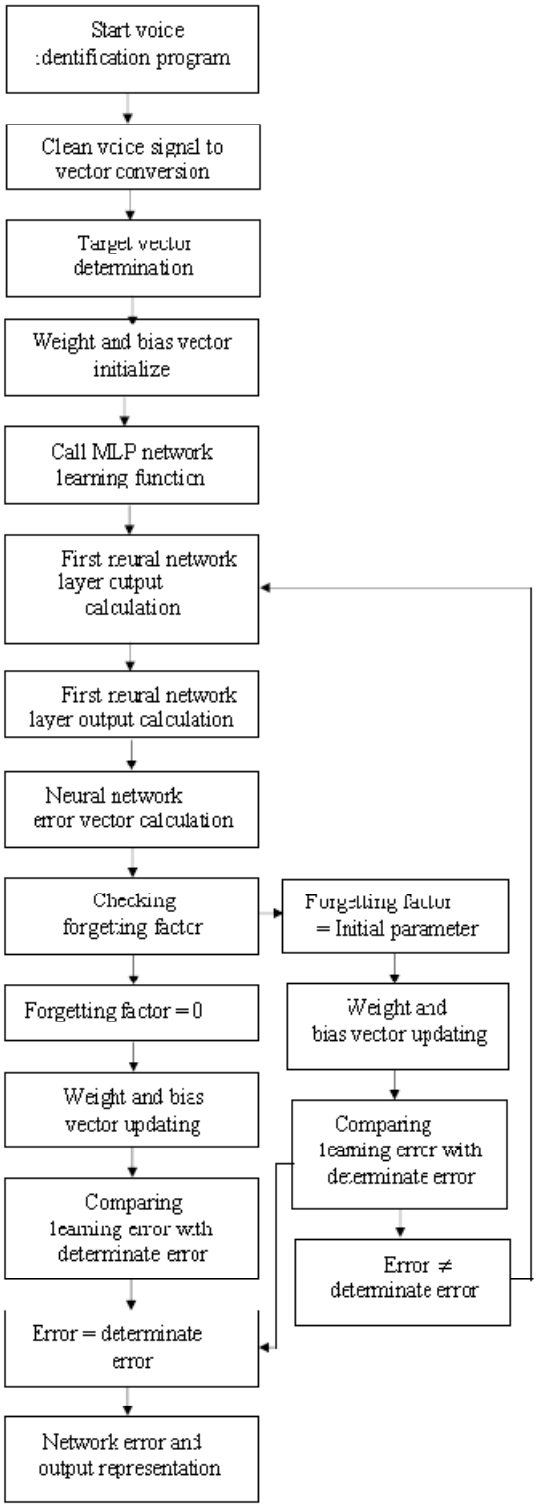


Fig.4: Neural network block diagram representation that trained with EBP method

2.4 Voice Identification Flowchart

In this flowchart, voice identification steps that consist of initializing, calculating, updating, and comparing network training parameters are shown separately.



3. Voice EKF Estimation

According to the mentioned voice identification steps, which lead to modelling and system dynamics identification, nonlinear model dynamics should be estimated. But here because of the nonlinear model, EKF should be used instead of the Kalman filter. EKF is an approximate method for nonlinear models that estimates these models as a time-variant linear model.

This point is important to mention, according to the invariant Kalman filter property, it is highly needed that system dynamics be known. On the other hand, for signal estimating with all kinds of Kalman filters, model estimation (modelling) is an important item.

In the following section, EKF equations are represented[12]:

3.1 Initialize

$$\hat{x}(0) = E(x_0)$$
$$P_0 = E[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T] \quad (6)$$

These equations show that the first input vector and then noise variance are initialized.

3.2 Time Update Equations

Then, time update equations are introduced.A_kVector is the most important segment and it is obtained by derivation of FFunction, which is an MLP neural network. To manipulate this vector, a Jacobin matrix is needed with arrays in which the irnumerators are the difference between two prior and posterior outputs and their enumerator are the difference between two present and previous inputs. These array numbers also depend on neuron numbers in the network input layer. In this paper, the number of neurons in the first layer is 5,so the matrix.A_kDimension is 5*5.

This point is important.A_kThe matrix is changed by index k, which varies from 1 to 39000, and it is not constant like the matrix. A Inthe Kalman filter time update equation.[13]

$$A_k = \frac{\partial F(X,W)}{\partial X} | x = \hat{x}_k \quad (7)$$

$$A_k = \begin{bmatrix} \frac{\partial f(\hat{x}_k, w)^T}{\partial x} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \ddots & 0 & \ddots \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{bmatrix} \quad (8)$$

$$\hat{x}_k^- = F(\hat{x}_{k-1}, w)$$
$$P_k^- = AP_{k-1}A^T + B\sigma_v^2 B^T \quad (9)$$

3.3 Measurement Update Equations

In the measurement update equation too, gain K and signal estimation and noise of variance are updated.

$$K = P_k^- C^T (C P_k^- C^T + \sigma_n^2)^{-1}$$
$$\hat{x}_k = \hat{x}_k^- + K(y_k - C\hat{x}_k^-) \quad (10)$$
$$P_k = (I - KC).P_k^-$$

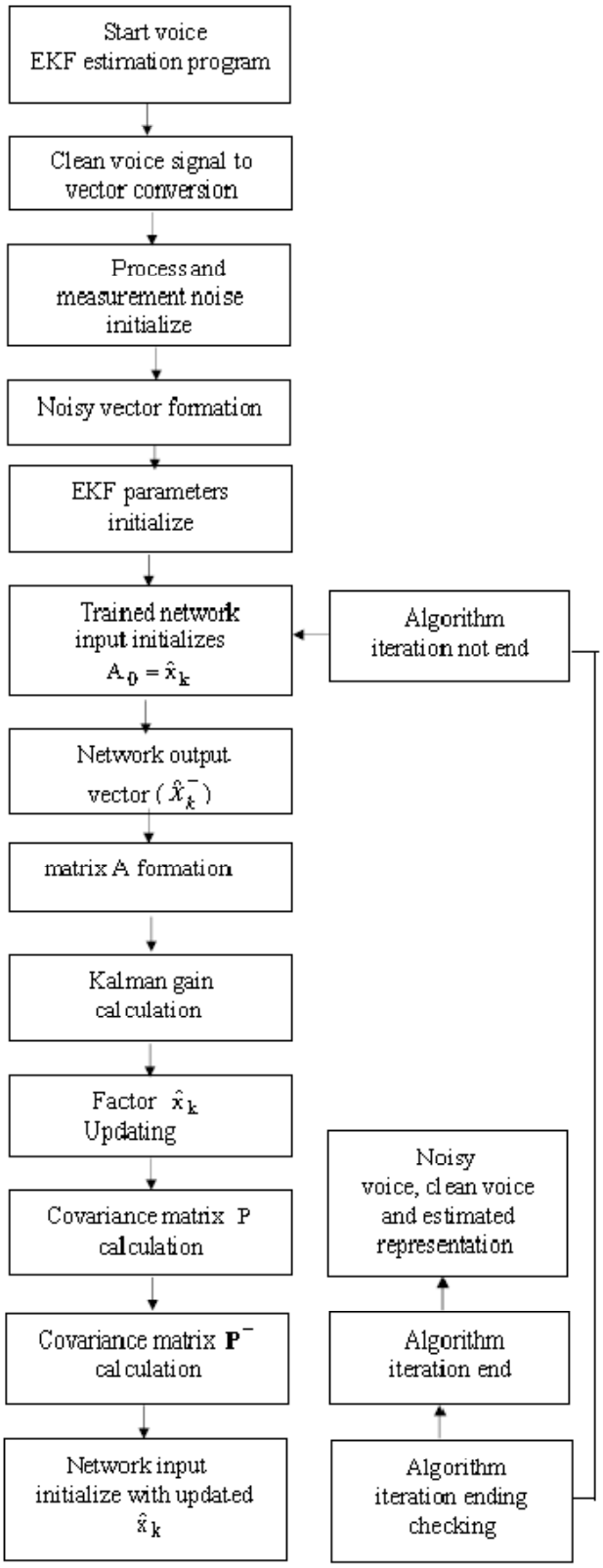
As an example, the condition of obtaining a matrixA_6That is the first matrix.AIn program circle is as below:

$$A_6 = \begin{bmatrix} \frac{\hat{x}_7^- - \hat{x}_6^-}{\hat{x}_2^- - \hat{x}_1^-} & \frac{\hat{x}_7^- - \hat{x}_6^-}{\hat{x}_3^- - \hat{x}_2^-} & \frac{\hat{x}_7^- - \hat{x}_6^-}{\hat{x}_4^- - \hat{x}_3^-} & \frac{\hat{x}_7^- - \hat{x}_6^-}{\hat{x}_5^- - \hat{x}_4^-} & \frac{\hat{x}_7^- - \hat{x}_6^-}{\hat{x}_6^- - \hat{x}_5^-} \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Then by using a matrixA_6in the measurement update equation, $\hat{x}_6$Is calculated; these calculations are iterated 39000 times, so iterations continue until$\hat{x}_{39006}$Iscalculated and the results are saved in program memory; they call back at the end of the program and compare with the clean voice.

3.4Voice EKF Estimation Flowchart

In this flowchart, clean voice estimation steps that consist of initializing, calculating, updating, and comparing parameters are shown separately.



4 Results

First, neural network output and clean signal, which

train network, are represented. Figures No. 5 to No. 8 are simulated with MATLAB 7.1 software.

In Figures 5and 6 voice identification is demonstrated with a neural network.

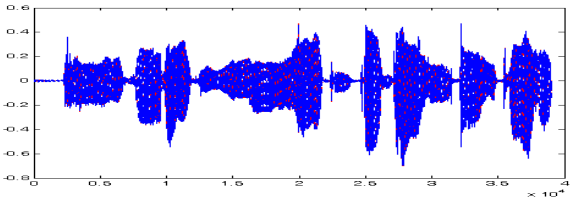


Fig. 5: Comparing network output (blue graph) with target signal (red graph)

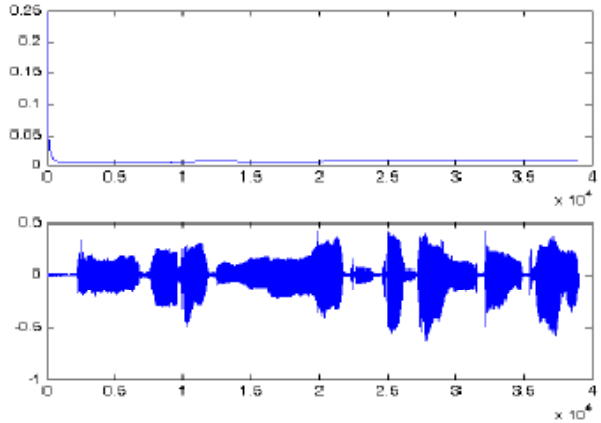


Fig. 6: network training error (first graph) and network output (second graph)representation

As specified in Figure 6, nonlinear voice is modelled well and the network output signal and target signal, which is clean voice, have obviously overlapped. Also, network training error, which is the sum of squared error and is obtained from the difference between the target signal and network output, converges to zero. That is the reason for network convergence.

In addition, we represent the estimated voice signal that is obtained by voice modelling and then voice estimation with EKF (Figure 7).

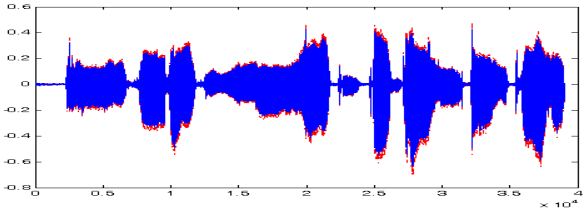


Fig. 7: Compare clean voice signal (red graph) with estimated voice signal (blue graph)

Finally, figure 8 represents 3 graphs. First, a clean voice signal is considered a neural network target for training. The second is, a noisy signal which is the conclusion of adding white noise with known variance to the clean voice signal, and the third is, the estimated voice signal that is obtained by EKF.

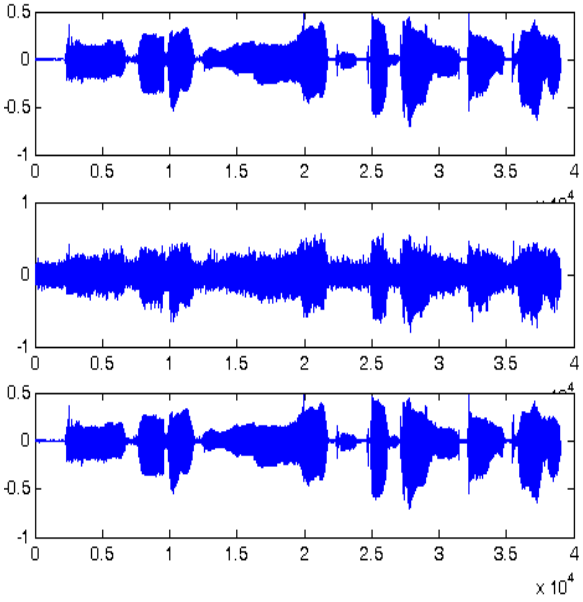


Fig. 8: Representation and comparison of clean voice signal (first graph), noisy signal (second graph), and estimated voice signal (third graph)

5 Conclusions

In conclusion, it is clear that neural networks in many cases in which classic control cannot be used, act well for system identification or modelling and can be applied with EKF to estimate nonlinear system dynamics as voice signals.

Also, it is shown that EKF is able to give certain estimations of nonlinear system dynamics, which are identified by neural networks and it is useful for researchers in control and communication fields.

References

[1] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," in IEEE Transactions on Speech and Audio Processing, vol. 9, no. 5, pp. 504-512, July 2001, doi: 10.1109/89.928915.

[2] I. Cohen and B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," in IEEE Signal Processing Letters, vol. 9, no. 1, pp. 12-15, Jan. 2002, doi: 10.1109/97.988717.

[3] I. Cohen, "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging," in IEEE Transactions on Speech and Audio Processing, vol. 11, no. 5, pp. 466-475, Sept. 2003, doi: 10.1109/TSA.2003.811544.

[4] P. C. Loizou, "Speech enhancement based on perceptually motivated bayesian estimators of the magnitude spectrum," in IEEE Transactions on Speech and [5] Audio Processing, vol. 13,

no. 5, pp. 857-869, Sept. 2005, doi: 10.1109/TSA.2005.851929.

[6] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," in IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 32, no. 6, pp. 1109-1121, December 1984, doi: 10.1109/TASSP.1984.1164453.

[7] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," in IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 33, no. 2, pp. 443-445, April 1985, doi: 10.1109/TASSP.1985.1164550.

[8] H. G. Hirsch and C. Ehrlicher, "Noise estimation techniques for robust speech recognition," 1995 International Conference on Acoustics, Speech, and Signal Processing, Detroit, MI, USA, 1995, pp. 153-156 vol.1, doi: 10.1109/ICASSP.1995.479387.

[9] R. Jaiswal, "Speech Activity Detection under Adverse Noisy Conditions at Low SNRs," 2021 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatre, India, 2021, pp. 97-101, doi: 10.1109/ICCES51350.2021.9488934.

[10] C. Medina, R. Coelho and L. Zão, "Impulsive Noise Detection for Speech Enhancement in HHT Domain," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 2244-2253, 2021, doi: 10.1109/TASLP.2021.3093392.

[11] R. Kumar and P. V. Subbaiah, "Enhancement of noisy speech using sub-band harmonic regeneration and speech presence uncertainty estimator," 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 2016, pp. 456-460, doi: 10.1109/RTEICT.2016.7807862.

[12] K. N. SunilKumar and Shivashankar, "A review on security and privacy issues in wireless sensor networks," 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 2017, pp. 1979-1984, doi: 10.1109/RTEICT.2017.8256945.

[13] R. Gatti and S. Hivashankar, "Study of different resource allocation scheduling policy in advanced LTE with carrier aggregation," 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 2017, pp. 2257-2261, doi: 10.1109/RTEICT.2017.8257002.