Automatic Speech Recognition for Qur'an Verses using Traditional Technique

Hamzah A. Alsayadi[1, *], Mohammed Hadwan[2]

[1]Computer Science Department, Faculty of Sciences, Ibb University, Yemen
[2] Department of Information Technology, College of Computer, Qassim University, Buraydah 51452, Saudi Arabia
[2]Department of Computer Science, College of Applied Sciences, Taiz University, Taiz 6803, Yemen
Email: hamzah.sayadi@cis.asu.edu.eg; m.hadwan@qu.edu.sa

Abstract

Deep learning is the one of approaches of machine learning that uses algorithms for building a model based on complex unstructured data. The Muslims Holy Qur'an book is written using Arabic diacritized text. In this paper, a traditional method to build a robust Qur'an versus recognition is proposed. The MFCC is used to extract features. These features are adapted using minimum phone error (MPE) as a discriminative model. The acoustic model was built using the deep neural network (DNN) model. We present an n-gram language model (LM). The dataset of Qur'an verses is used for training and evaluating the proposed model, consisting of 10 hours of .wav recitations performed by 60 reciters. The Experimental results showed that the proposed DNN model achieved a significantly low character error rate (CER) of 4.09% and a word error rate (WER) of 8.46%.

Keywords: Quran verses; Deep neural network (DNN); Arabic ASR.

1 Introduction

Nearly 300 million people are native speakers of the Arabic language, which is one of a Semitic language family that includes other well-known languages such as Hebrew and Aramaic [1, 2]. The Arabic alphabet consists of 28 letters, all of which represent consonants and are written from right to left. The three-letter root system is the most distinctive feature of Semitic languages. The pattern system is another important aspect of the Arabic language when a pattern is imposed on the fundamental root. For example, wrote and writer in the Arabic language is Katab "كتب" and Kateb "كاتب" which have the same three root letters, for more details about Semitic languages refer to [1]. The Arabic script is a modified abjad [3] in which letters represent short consonants and long vowels, but short vowels and consonant length are not commonly shown in writing. Diacritics "Tashkeel" is an optional character that can be used to denote missing vowels and consonant length. The Arabic script contains various diacritics, which are critical in achieving typographic text usability standards such as homogeneity, clarity, and readability. Any modification to the letter's diacritic of the word can radically alter its meaning [4, 5].

The Holy Qur'an is Islam's fundamental religious book divided into 114 chapters, each of which consists of verses. Besides its religious importance for Muslims, it is largely recognized as the best work in Arabic literature and has had a tremendous influence on the Arabic language [6]. Qur'an recitation "Qira'at" is a daily practice of Muslims as a part of their faith. The Qur'an was passed down from generation to generation through recitation. The Qur'an has seven canonical qira'at [7]. The correct recitation of the Holy Qur'an depends mainly on another discipline known as Tajweed. Tajweed [8] specifies the correct way of reciting the Qur'an, how to correctly pronounce each individual syllable, the pause places in Qur'an verses, in addition to elisions, where long or short pronunciation is

Doi: https://doi.org/10.54216/JAIM.010202
Received: March 11, 2022 Accepted: June 17, 2022

needed, where letters should be kept separate, and where they should be sounded together, and so on. All Muslims around the world, Arabic native speakers, and non-Arabic speakers are reciting and listening to Qur'an in the Arabic language.

Automatic Speech Recognition (ASR) is a task used to convert speech waves or signals to its mapping sequence of words or units using a determined algorithm. These sequences are represented like human transcription. In addition, it is a technology that makes disabled people communicate with society. It can able to make life easier and very promising [9, 10]. ASR is the use of computers to recognize and process a person's speech. In research and industry, there has been a substantial growth in interest in ASR mostly directed toward the English language. For the Arabic language, little attention is paid to exploring ASR where several attempts can be found in the literature such as in [11, 12, 13, 14]. Published review papers of ASR for the Arabic language can be found in [13], [15, 16, 17].

In this paper, we used DNN to build ASR Qur'an reciters based on some of the Qur'an Verses. We present an acoustic model using the deep neural network (DNN) method. This model is adapted using the minimum phone error (MPE) method. Our experiments show the effectiveness of the proposed model for recognition of the reciter's voice and the detection of the versus text.

This paper is organized as follows, Section 2 presents the related works. In Section 3, the implemented system is described. In Section 4, the experimental setup and the used dataset are introduced. Then, Section 5 is devoted to the results and discussion. Finally, the conclusion of this research is in Section 6.

2 Literature Review

In this section, the related literature to the proposed model is presented. The attention paid to the related ASR method focused to recognized the Qur'an reciters speech recognition. Jawad [18] used two classifiers, (i) K nearest neighbors (KNN), and (ii) artificial neural network (ANN) to recognize the Holy Qur'an reciters. MFCC is used to analyze the audio dataset. Pitch was used as a feature to train KNN and ANN.

Nahar et al. in [19] have used a recognition model to identify the "Qira'ah" (type of reading) from the related Holy Qur'an audio wave. The proposed model was created in 3 stages: (i) the extraction and labeling of MFCC feature from an acoustic signal, (ii) training the SVM learning model with the identified features, and (iii) detecting "Qira'ah". With a success rate of 96 percent, the experimental findings demonstrated the power of the introduced SVM-based recognition model.

Lataifeh et al. [20] compared the performance of classical vs. deep-based classifiers. Moreover, a comparison between the accuracy of the automatically proposed method in contradiction of human expert listeners' in recognizing reliable reciters from imitators. Results showed that the accuracy of selected classical and deep-based classifiers reached 98.6% of accuracy compared to 61% of human experts. Arabic diversified dataset is introduced lately by Lataifeh and Elnagar [21] to have a unified dataset that can be used to assess the introduced method and models for Qur'anic research. Ammar Mohammed et al. in [22] provided a technique for Qur'an reciters rules Recognition to detect the Medd rule and Ghunnah using phoneme duration. The used dataset was gathered from 10 Qur'anic reciters in order to compute the Medd and Ghunnah durations in the right recitation. The developed approach was then utilized to identify the Medd and Ghunnah as Tajweed norms in Quran recitation.

In Gunawan et al. [23], for Qur'anic reciter identification, the features of Mel frequency cepstral coefficients (MFCC) were extracted from the recorded audio, and after training a Gaussian mixture model (GMM), Gaussian Supervectors (GSVs) were formed using model parameters such as the mean vector and the main diagonal of the covariance matrix. This model can be applied to protocol classification, feature learning, anomalous protocol identification, and unknown protocol classification. The researchers in [24] used a support vector machine (SVM) and threshold scoring system to recognize different Tajweed rules automatically. 70- dimensional filter banks were used for feature extraction. A new dataset collected by the authors was used for the experiments and very promising results were obtained.

In [25], the authors used features such as MFCC and Pitch for learning process to recognize the Qur'an reciters. Several machine learning algorithms are used as Random Forest, Na¨ıve Bayes and J48. The obtained results show the ability of proposed model to detect Qur'an reciter based on the used dataset. The best recognition accuracy was 88% when used the Na¨ıve Bayes.

Doi: https://doi.org/10.54216/JAIM.010202
Received: March 11, 2022 Accepted: June 17, 2022

3 System Description

The implemented ASR system comprises different components including feature extraction, language model, and acoustic model. These components are built and processed separately.

3.1 Feature Extraction

We use Mel-frequency cepstral coefficient (MFCC) steps for feature extraction. In the feature extraction stage, the input audio waveform represents the input signal, these waves are converted into vectors with a fixed size that are merged. The MFCC is important because it simulates human ear behavior [26].

MFCC-based features are extracted and presented for data preparation, and training steps. The presented models use the standard 13-dimensional cepstral mean-variance normalized (CMVN) features for building the acoustic models as presented in [27]. Each frame includes the neighboring ±4 frames. Then, we used linear discriminative analysis (LDA) transformation to transform frames with 40 dimensions.

3.2 Language Model

The language model is built using the training and collected data. The LM is built using the CMUCLMTK toolkit based on the 3-g method.

3.3 Acoustic Model

A DNN model comprises an input layer, an output layer, and two or more layers of hidden units. Each hidden unit is used to associate all inputs from the previous layer to the scalar state using the logistic function and sent into the next layer. DNNs are discriminatively trained using the backpropagation of cost function derivatives. This backpropagation is utilized to determine the conflict between the original outputs and resulted from DNN training [28]. DNN can be trained on a large training set by calculating the derivatives on a small part of the training set "minibatch" compared to the whole training set. Then, it updates the weights to the gradient. The trained neural networks in DNN are used to recognize speech. It includes the dimension of the input spectral features as the input layer, N hidden layers, and one output layer. The output layer dimension is equal to the number of utterances the system is designed to identify. The frame-level DNN posteriors from the output layer must be combined by simply averaging over the test utterance.

Neural networks recently have been considered one of the state-of-the-art of acoustic modeling. So, we use the DNN technique to build the acoustic model based on MFCC feature vectors. MFCC with LDA adaptation represents the acoustic features for the DNN model. In a DNN model, a DNN produces the state posteriors.

We apply MPE to decrease an approximation of the expected errors over the training data for enhancing the accuracy of the model parameters. At the model decoding stage, DNN features are calculated from state vectors that are obtained from the DNN model for each utterance. [29-42] In addition, MPE is performed for the new training data (DNN features + MFCC with LDA adaptation). Therefore, the loss function for MPE is calculated depending on the phone error the hypothesis and the corresponding reference at the word level. Thus, the DNN-MPE model will be created as a new model.

4 Experimental Setup and Dataset

This section reports the preliminary experiments for the traditional approach based on Qur'an Verses Dataset. The proposed model is trained and evaluated using a collected Qur'an verses dataset. LM is trained and evaluated on collected data from several websites. For evaluation purposes, the traditional character error rate (CER) and word error rate (WER) are used to report the accuracy of recognition.

4.1 Qur'an Verses Dataset

For the Qur'an dataset used in this research, 10 hours of mp3 Qur'an verses recited by 60 reciters were collected from the A2Youth.com website and its associated transcripts. The collected dataset is used for training and

Doi: https://doi.org/10.54216/JAIM.010202
Received: March 11, 2022 Accepted: June 17, 2022

evaluating the proposed models. Dataset is distributed into three parts: (i) 70% for the training set, (ii) 10% for the development set, and (iii) 20% for the testing set.

4.2 Experimental setup

We trained two DNN acoustic models:

• DNN was trained on SD-GMM feature vectors and alignment lattices that were obtained from GMM model.
• DNN-MPE was trained on DNN feature vectors that were obtained from DNN model.

Firstly, we trained the DNN model based on MFCC with LDA adaptation features. This model was trained with four hidden layers, and 1,024 neurons per layer. We used -1,0,1 -1,0,1, -3,0,3 -3,0,3 as splicing indexes. The first index makes the frst layer deal with three consecutive frames of observation. While -3,0,3 makes most hidden layers deal with three frames in the previous layer. Then, we use the DNN model to generate the new feature vectors (DNN features) + MFCC with LDA adaptation as input features for DNN-MPE. For training DNN-MPE model, we use MPE adaptation for DNN model in order to enhance the accuracy of DNN model. Then, we trained DNN-MPE over the DNN feature vectors.

5 Results and Discussion

The employed model is used for Qur'an versus recognition. Experimental results of this model on a collected verses dataset shows the outstanding performance of the proposed model. The used dataset comprises 60 reciters with 16 verses for each reciter.

We applied DNN technique with MPE adaptation method for overall training data. In this process, we employed two DNN models: DNN and DNN-MPE. Table 1 shows the results of DNN models with LM as CER and WER measures. The first result in Table 1 is the WER of the employed DNN model, which is trained using feature-level of MFCC + LDA adaptation.

Table 1: The results of the DNN models

Model	CER	WER
DNN	5.54	11.83
DNN-MPE	4.09	8.46

While the result of the DNN-MPE model is introduced in the second line of the table. This result is obtained after applying MPE adaptation methods, whereas DNN-MPE is trained using DNN features followed by MFCC + LDA.

In Table 1, we can see the MPE adaptation method over DNN improves the accuracy of the DNN model on testing data. We can note that the DNN model achieved 5.54% CER and 11.83% WER. After applying MPE as a proposed adaptation method on DNN and training DNN-MPE using DNN features in concatenation with MFCC + LDA features, CER is enhanced by 1.45% over the DNN model. While, WER is enhanced by 3.37% over the DNN model. In addition, the DNN-MPE model has result better than DNN model's results.

We can conclude that applying adaptation methods on DNN and using DNN-features reduce the results as WER. Thus, MPE affects the performance of DNN model.

The results are also visually presented in Figure 1. What stands out in Table 1 is the significant enhancement of the WER after applying the adaptation methods in the DNN model. It also shows the effect of feature-level of DNN on WER.

Doi: https://doi.org/10.54216/JAIM.010202
Received: March 11, 2022 Accepted: June 17, 2022

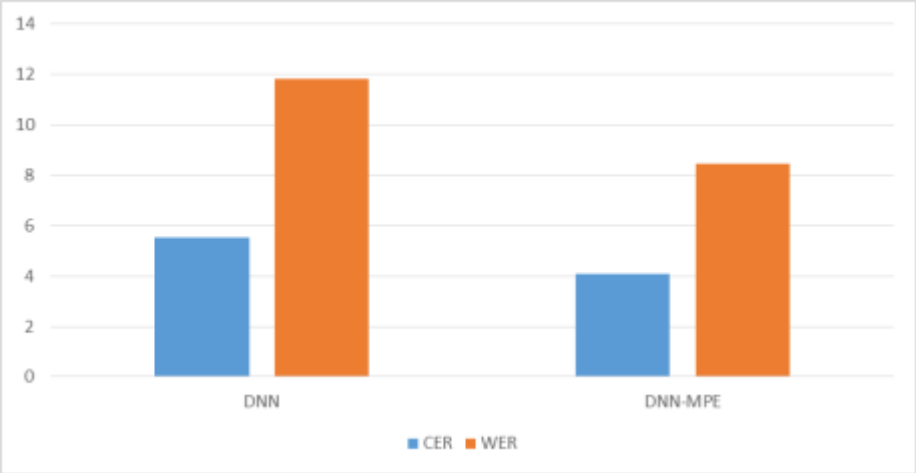

Figure 1: Results of DNN and DNN-MPE models

6 Conclusions

This paper proposed the traditional method for Qur'an verses recognition based on Qur'an verses dataset. This method is used for building the acoustic model. MFCC is used to extract the features. We built an n-gram language model using CMUCLMTK toolkit. We proposed DNN method to train and build acoustic model. This acoustic model is adapted using a discriminative model called MPE. After acoustic and discriminative models training, two models are obtained that are built separately. In experiments, we separately evaluated each model on Qur'an verses dataset. The results show that the accuracy of DNN-MPE model is better than DNN. It also shows the effects of discriminative models on accuracy.

References

[1] Geoffrey Khan, Michael P Streck, and Janet CE Watson. *The Semitic languages: An international handbook*, volume 36. Walter de Gruyter, 2011.

[2] Hamzah A Alsayadi and Abeer M ElKorany. Integrating semantic features for enhancing arabic named entity recognition. *International Journal of Advanced Computer Science and Applications*, 7(3), 2016.

[3] Norah Alsunaidi, Lobna Alzeer, Maha Alkatheiri, Alaa Habbabah, Marwah Alattas, Malak Aljabri, and Mona Altassan. Abjad: Towards interactive learning approach to arabic reading based on speech recognition. *Procedia Computer Science*, 142:198–205, 2018.

[4] Mohamed Hssini and Azzeddine Lazrek. Design of arabic diacritical marks. *arXiv preprint arXiv:1107.4734*, 2011.

[5] Hamzah A. Alsayadi, Abdelaziz A. Abdelhamid, Islam Hegazy, Bandar Alotaibi, and Zaki T. Fayed. Deep investigation of the recent advances in dialectal arabic speech recognition. *IEEE Access*, 10:57063–57079, 2022.

[6] JA Devenny. Arberry, aj," the koran interpreted"(book review). *Theological Studies*, 17:440, 1956.

[7] Mohamed Abdelmonem Elsayed Khalil and Nor Hafzi Yusof. [the differences of the quranic qiraat in tafsir imam al-tabari and its effects on the hukm of fqh] ikhtilaf al-qira'at al-qur'aniah f tafsir at-tabari wa asruhu ala al-ahkam al-fqhiyyah: Dirasat tahliliah. *Jurnal Islam Dan Masyarakat Kontemporari*, 16(1):111–126, 2018.

[8] Ahmad Hanifuddin Ishaq and Ruston Nawawi. Ilmu tajwid dan implikasinya terhadap ilmu qira'ah. *QAF*, 1(1):15–37, 2017.

[9] Hamzah Alsayadi, Abdelaziz Abdelhamid, Islam Hegazy, and Zaki Taha. Data augmentation for arabic speech recognition based on end-to-end deep learning. *International Journal of Intelligent Computing and Information Sciences*, 21(2):50–64, 2021.

[10] Abdelaziz A Abdelhamid, Waleed H Abdulla, and Bruce A MacDonald. Roboasr: A dynamic speech recognition system for service robots. In *International Conference on Social Robotics*, pages 485–495. Springer, 2012.

Doi: https://doi.org/10.54216/JAIM.010202

Received: March 11, 2022 Accepted: June 17, 2022

[11] Imad K Tantawi, Mohammad AM Abushariah, and Bassam H Hammo. A deep learning approach for automatic speech recognition of the holy qur'an recitations. ¯ *International Journal of Speech Technology*, 24(4):1017–1032, 2021.
[12] Hassan Tabbal, W El Falou, and B Monla. Analysis and implementation of a" quranic" verses delimitation system in audio fles using speech recognition techniques. In *2006 2nd international conference on information & communication technologies*, volume 2, pages 2979–2984. IEEE, 2006.
[13] Nazik O'mar Balula, Mohsen Rashwan, and Shrief Abdou. Automatic speech recognition (asr) systems for learning arabic language and al-quran recitation: A review. 2021.
[14] Faza Thiraf and Dessi Puji Lestari. Hybrid hmm-blstm-based acoustic modeling for automatic speech recognition on quran recitation. In *2018 International Conference on Asian Language Processing (IALP)*, pages 203–208. IEEE, 2018.
[15] Abdelaziz A Abdelhamid, Hamzah A Alsayadi, Islam Hegazy, and Zaki T Fayed. End-to-end arabic speech recognition: A review. In *Proceedings of the 19th Conference of Language Engineering (ESOLEC'19), Alexandria, Egypt*, pages 26–30, 2020.
[16] SR Shareef and YF Irhayim. A review: isolated arabic words recognition using artificial intelligent techniques. In *Journal of Physics: Conference Series*, volume 1897, page 012026. IOP Publishing, 2021.
[17] Noor Jamaliah Ibrahim, Mohd Yamani Idna Idris, MY Zulkifli Mohd Yusoff, and Asma Anuar. The problems, issues and future challenges of automatic speech recognition for quranic verse recitation: A review. *Al-Bayan: Journal of Qur'an and Hadith Studies*, 13(2):168–196, 2015.
[18] Jawad H Alkhateeb. A machine learning approach for recognizing the holy quran reciter. *International Journal of Advanced Computer Science and Applications*, 11(7), 2020.
[19] Khalid MO Nahar, M Ra'ed, A Moy'awiah, and M Malek. An efficient holy quran recitation recognizer based on svm learning model. *Jordanian Journal of Computers and Information Technology (JJCIT)*, 6(04), 2020.
[20] Mohammed Lataifeh, Ashraf Elnagar, Ismail Shahin, and Ali Bou Nassif. Arabic audio clips: Identification and discrimination of authentic cantillations from imitations. *Neurocomputing*, 418:162–177, 2020.
[21] Mohammed Lataifeh and Ashraf Elnagar. Ar-dad: Arabic diversified audio dataset. *Data in Brief*, 33:106503, 2020.
[22] Ammar Mohammed, Mohd Shahrizal Bin Sunar, Md Salam, and Sah Hj. Recognition of holy quran recitation rules using phoneme duration. In *International Conference of Reliable Information and Communication Technology*, pages 343–352. Springer, 2017.
[23] Teddy Surya Gunawan, Nur Atikah Muhamat Saleh, and Mira Kartiwi. Development of quranic reciter identification system using mfcc and gmm classifier. *International Journal of Electrical & Computer Engineering (2088-8708)*, 8(1), 2018.
[24] Ali M Alagrami and Maged M Eljazzar. Smartajweed automatic recognition of arabic quranic recitation rules. *arXiv preprint arXiv:2101.04200*, 2020.
[25] Rehan Ullah Khan, A Qamar, and Mohammed Hadwan. Quranic reciter recognition: a machine learning approach. *Advances in Science, Technology and Engineering Systems Journal*, 4(6):173–176, 2019.
[26] Hamzah A Alsayadi, Abdelaziz A Abdelhamid, Islam Hegazy, and Zaki T Fayed. Non-diacritized arabic speech recognition based on cnn-lstm and attention-based models. *Journal of Intelligent & Fuzzy Systems*, (Preprint):1–13, 2021.
[27] Hamzah A Alsayadi, Abdelaziz A Abdelhamid, Islam Hegazy, and Zaki T Fayed. Arabic speech recognition using end-to-end deep learning. *IET Signal Processing*, 15(8):521–534, 2021.
[28] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by backpropagation errors. *nature*, 323(6088):533–536, 1986.
[29] El-kenawy, El-Sayed M., Marwa M. Eid, and Abdelhameed Ibrahim. "Anemia estimation for covid-19 patients using a machine learning model." *Journal of Computer Science and Information Systems* 17, no. 11 (2021): 2535-1451.
[30] Hussien, Hussien Rezk, El-Sayed M. El-Kenawy, and Ali I. El-Desouky. "EEG channel selection using a modified grey wolf optimizer." *European Journal of Electrical Engineering and Computer Science* 5, no. 1 (2021): 17-24.
[31] Salamai, Abdullah Ali, El-Sayed M. El-kenawy, and Ibrahim Abdelhameed. "Dynamic voting classifier for risk identification in supply chain 4.0." *CMC-COMPUTERS MATERIALS & CONTINUA* 69, no. 3 (2021): 3749-3766.

Doi: https://doi.org/10.54216/JAIM.010202
Received: March 11, 2022 Accepted: June 17, 2022

[32] Eid, Marwa M., El-Sayed M. El-kenawy, and Abdelhameed Ibrahim. "A binary sine cosine-modified whale optimization algorithm for feature selection." In *2021 National Computing Colleges Conference (NCCC)*, pp. 1-6. IEEE, 2021.

[33] Eid, Marwa M., and M. El-Sayed. "El-kenawy, and Abdelhameed Ibrahim." An Advanced Patient Health Monitoring System."." *Journal of Computer Science and Information Systems* 17.

[34] El-kenawy, El-Sayed M., Marwa M. Eid, and Abdelhameed Ibrahim. "Automatic identification from noisy microscopic images." *Journal of Computer Science and Information Systems* 17, no. 11 (2021).

[35] Alharbi, Manal SF, and El-Sayed M. El-kenawy. "Optimize machine learning programming algorithms for sentiment analysis in social media." *International Journal of Computer Applications* 174, no. 25 (2021): 38-43.

[36] Alharbi, Manal SF, and El-Sayed M. El-kenawy. "Recommendation System for Analyzing the Preference Data of the Multimedia Software Tools in Education." (2021).

[37] El-kenawy, E. S. M. T. "Solar radiation machine learning production depend on training neural networks with ant colony optimization algorithms." *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)* 7, no. 5 (2018): 1-4.

[38] El-Sayed Towfek, M. "El-kenawy. Trust Model for Dependable File Exchange in Cloud Computing." *International Journal of Computer Applications* 180, no. 49 (2018): 22-27.

[39] El-Kenawy, El-Sayed M., Abdelhameed Ibrahim, Nadjem Bailek, Kada Bouchouicha, Muhammed A. Hassan, Basharat Jamil, and Nadhir Al-Ansari. "Hybrid ensemble-learning approach for renewable energy resources evaluation in Algeria." *Computers, Materials & Continua* 71, no. 3 (2022): 5837-5854.

[40] El-kenawy, El-Sayed M., Abdelhameed Ibrahim, Nadjem Bailek, Kada Bouchouicha, Muhammed A. Hassan, Mehdi Jamei, and Nadhir Al-Ansari. "Sunshine duration measurements and predictions in Saharan Algeria region: an improved ensemble learning approach." *Theoretical and Applied Climatology* 147, no. 3 (2022): 1015-1031.

[41] Ibrahim, Abdelhameed, Seyedali Mirjalili, Mohammed El-Said, Sherif SM Ghoneim, Mosleh M. Al-Harthi, Tarek F. Ibrahim, and El-Sayed M. El-Kenawy. "Wind speed ensemble forecasting based on deep learning using adaptive dynamic optimization algorithm." *IEEE Access* 9 (2021): 125787-125804.

[42] El-kenawy, El-Sayed M., Hattan F. Abutarboush, Ali Wagdy Mohamed, and Abdelhameed Ibrahim. "Advance artificial intelligence technique for designing double T-shaped monopole antenna." *CMC-COMPUTERS MATERIALS & CONTINUA* 69, no. 3 (2021): 2983-2995.

Doi: https://doi.org/10.54216/JAIM.010202

Received: March 11, 2022 Accepted: June 17, 2022