ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/373326112

DeepFake Videos Detection by Using Recurrent Neural Network (RNN)

Conference Paper · July 2023

DOI: 10.1109/AICCIT57614.2023.10217956

CITATIONS
8

READS
509

6 authors, including:

Haider A. Al-Wzwazy
Queensland University of Technology
27 PUBLICATIONS **230** CITATIONS

SEE PROFILE

Ahmed Salih Al-Khaleefa
University of Misan
41 PUBLICATIONS **433** CITATIONS

SEE PROFILE

Murtadha Ali Alazzawi
Imam Alkadhim University College
30 PUBLICATIONS **391** CITATIONS

SEE PROFILE

All content following this page was uploaded by Haider A. Al-Wzwazy on 13 November 2024.

The user has requested enhancement of the downloaded file.

DeepFake Videos Detection by Using Recurrent Neural Network (RNN)

1st Ali Abdulzahra Mohsin Albazony
Computer Engineering Department, Faculty of Engineering
Shahid Chamran University of Ahvaz
Ahvaz, Iran
Alidash2000@gmail.com

2nd Haider A AL-wzwazy
Halfaya Oilfield Division
Missan Oil Company
Maysan, Iraq
haiderit@uomisan.edu.iq

3rd Ahmed Salih AL-Khaleefa
Department of Physics, Faculty of Education
University of Misan
Maysan, Iraq
alkhaleefa.as@uomisan.edu.iq

4th Murtadha A. Alazzawi
Department of Computer Techniques Engineering
Imam Al-Kadhum College (IKC)
Baghdad, Iraq
murtadhaali@alkadhum-col.edu.iq

5th Mohammed Almohamadi
Department of Computer Technical Engineering, College of Information Technology
Imam Ja'afar Al-Sadiq University
Al-Muthanna, Iraq
mohammed.sites89@gmail.com

6th Seyed Enayatallah ALAVI
Computer Engineering Department, Faculty of Engineering
Shahid Chamran University of Ahvaz
Ahvaz, Iran
se.alavi@scu.ac.ir

2023 Al-Sadiq International Conference on Communication and Information Technology (AICCIT) | 979-8-3503-4188-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/AICCIT57614.2023.10217956

***Abstract*— In the last few years, the increasing development of various tools to make fake videos from real videos has been raised. Thus, several models/approaches have been constructed to detect and reveals fake video. Consequently, this research is conducted to propose a new model based on combining Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and image preprocessing techniques to classify and find the fake video from the real video. To implement and evaluate the proposed model, a MATLAB simulator has been used. The deepFake Images dataset is used for evaluations. This dataset contains 135 real videos as well as 677 fake videos created using different tools on real videos. Two scenarios have been utilized to evaluate the performance of the proposed model which are; dimensions of the training data and different sizes of the training data. The results show that the proposed model has been able to provide better results than the previous model.**

Keywords—Deep learning, DeepFake Detection, convolutional Neural Network (CNN), Recurrent Neural Network (RNN)

I. INTRODUCTION

This template, modified to be able to create videos that are indistinguishable from immaculate footage, may now be achieved by anyone, even those without a lot of experience or processing power, by advancements in video manipulation techniques over the past several years. These techniques, for instance, can be used to insert someone into a movie or modify their lip movements to say whatever they want. They can also be used to do convincing face swaps in videos utilizing free software tools. These videos are "deepFake," or fake, videos. The term "DeepFake," which combines the terms "Deep Learning" and "Fake," refers to artificial intelligence-produced videos in which one person in a video or image is replaced by another [1]. Nowadays, reality distortion is a major issue since more and more phony images and movies circulate online every day and are harder and harder to spot. scenarios in which these convincing phony videos are used to stoke political unrest, extort money, or launch terrorist attacks. DeepFake technology will therefore have certain benefits but also many drawbacks [2]. Nothing is more hazardous than people taking movies like this for real in a time when the truth is eroding [1].

The face is the aspect of deepFake that is targeted the most. As a result, various algorithms and strategies can be employed to find altered faces. These facial adjustments might be either Expression- or Identity-based. In the first, a person's expressions are instantly conveyed to a different person. Identity manipulation involves switching the faces of two individuals. By exchanging the face of a prominent person, this kind of manipulation can be used to spread misleading information among people [3]. In this study, a method employing image preprocessing techniques and neural networks, such as CNN or LSTM neural networks, is proposed. Finally, a distinction between real and phony images can be made. This is accomplished by utilizing the proper deep neural networks to categorize the photos as real or false after training the model on a dataset.

II. RELATED WORK

The initial purpose of deep neural network processing units (also known as deep neural networks for fraudulent video detection) was to render deep neural networks. The deep neural network of false video identification, however, has gained popularity in the last 10 years for general-purpose high-performance computation, including medical image processing [4]. This is most likely a result of the great performance, low cost, and increased programmability of these devices. Using biological signals, Umur et al. [3] introduced the generic DeepFake detector FakeCatcher (internal representations of image generators and synthesizers). They employed a straightforward, three-layer Convolutional Neural Network (CNN) classifier. The authors utilized 3000 movies for training and testing. They did not, however, go into great depth about how they prepared their data. MesoNet network, a CNN model, was suggested by Afchar et al., [5] to automatically identify hyper-realistic fake movies produced with DeepFake and Face2Face. The Meso-4 and MesoInception-4 network topologies, which concentrate on the mesoscopic characteristics of an image, were adopted by the authors.

Inconsistencies in image transformation (i.e., scaling, rotation, and shearing) result from the construction of DeepFakes, according to Yuezun and Siwei's [6] CNN architecture. Their model focuses on the affine face-warping artifacts as the distinguishing characteristic between authentic and fraudulent photos. Their technique looks for resolution

979-8-3503-4188-1/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: Queensland University of Technology. Downloaded on November 13,2024 at 01:02:07 UTC from IEEE Xplore. Restrictions apply.

irregularities caused by face warping by comparing the DeepFake face region with that of the surrounding pixels. A novel deep-learning strategy was put forth by Huy et al., [7] to identify fake photos and videos. The authors concentrated on totally computer-generated picture spoofing, face swapping, facial reenactments, and replay attacks. A system that extracts visual and temporal features from faces seen in a movie was proposed by Montserrat et al., [8]. They use a CNN and RNN architectural combination to find DeepFake videos. To detect deepFakes, Rana and Sung [9] suggested the DeepFakeStack ensemble model (A stack of various DL models). The open-source DL models used in the ensemble are XceptionNet, InceptionV3, InceptionResNetV2, MobileNet, ResNet101, DenseNet121, and DenseNet169.

Using ShallowNet, VGG-16, and Xception pre-trained DL models, Kim et al., [10] suggested a classifier that separates target individuals from a pool of similar individuals. Their system's primary goal is to assess how well the three DL models perform in terms of classification. A convolutional vision transformer is suggested by [11] to identify DeepFakes. A CNN and a vision transformer are the two halves of the convolutional vision transformer (ViT). While ViT uses learned features as input and classifies them using an attention model, CNN extracts learnable features. The DeepFake Detection Challenge (DFDC) dataset was used to train the authors' model, which had an accuracy rate of 91.5 percent, an AUC value of 0.91, and a loss value of 0.32. A CNN module added to the ViT architecture has typically, using this strategy, produced competitive results in the DFDC dataset. For three main explanations, the authors generalize their modelology [12]. 1) The suggested model can use CNN and a transformer architecture with a transformer attention mechanism to learn local and global picture attributes. 2) This model emphasizes the pre-processing of data during training and classification. 3) To identify DeepFakes produced in various contexts, environments, and orientations, it is suggested that the model be trained on a wide range of face photos utilizing the greatest dataset currently accessible.

An LSTM-based residual network (CLRNet) is built in this research [13] that uses a video's sequence of consecutive images as input to learn temporal information that aids in the detection of artificial artifacts that appear between frames of fraudulent videos. Additionally, a transfer learning-based strategy is suggested to generalize various deepFake techniques [14]. The suggested model with higher generality outperforms five of the most sophisticated deepFake detection models previously described in detecting various deepFake models using a single model, according to thorough trials using the FaceForensics++ dataset.

III. Proposed Model

Many tools are available to generate DeepFake, but few tools are available to detect DeepFake. The major objective of this research is to create the best model for detecting phony deep movies with the greatest degree of accuracy. Three sub-goals can be created from the main aim mentioned above. Prepare the data set, which combines data from various distributions, in the first phase so that the findings, when employing movies from various distributions, will be consistent. Reducing the discrepancy between attained accuracy and validation accuracy is the second subcategory. Although the test accuracy was in the same distribution as the CNN and RNN were trained, the suggested DeepFake detection systems using CNN and RNN previously managed to obtain an average accuracy higher than 90%. Such strategies cannot produce the same precision when used with various distributions. The third sub-goal is to create a real-time-based platform where the user can upload or select a video that contains the people on it. It tells if this video is a deepFake or not [15]. Our goal is for the model to return optimal values for the evaluation criteria, whether it's fake or not, but ideally, we want this movie to return more optimal results. This research uses a hybrid dataset that contains an equal number of real and fake videos in our research on DeepFake detection [16]. This dataset contains the "Deep Fake Images" dataset. In this dataset, 135 real videos are placed, and 677 fake videos are also created by altering various fake video creation tools.

The experimental steps of the problem are generally shown as follows:

Step 1: To get better results, we must first gather a dataset of authentic and fraudulent movies. This dataset should be labeled as such, and we must ensure that the videos come from various distribution channels.

Step 2: We must first remove the frames from the videos before removing the faces area from the frames.

Step 3: The CNN must get the data after we have extracted the faces from the frames to extract the features.

Step 4: The LSTM receives the CNN's output feature vector for training, then the LSTM generates the final output to verify correctness. The proposed model architecture is shown in Fig. 1. The Training Flow is represented by the black line, while the Prediction Flow is shown by the red line.

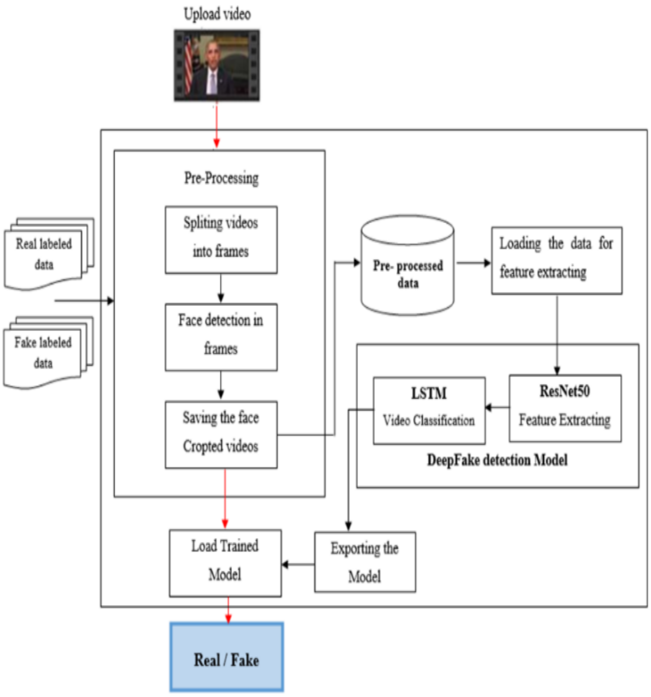

Fig. 1. The proposed DeepFake Video detector's flowchart.

IV. Evaluation criteria

In most research based on classification, like what we have done in this research, the benchmark criteria of Precision, Recall, Accuracy, and F-Measure are used for evaluation. These criteria have been utilized to evaluate the performance of the proposed model. As shown in Table I, these values have

Authorized licensed use limited to: Queensland University of Technology. Downloaded on November 13,2024 at 01:02:07 UTC from IEEE Xplore. Restrictions apply.

True-Positive (TP), True-Negative (TN), False-Positive (FP), and False-Negative (FN).

TABLE I. CONFUSION MATRIX

True value \ Predicted values	Real video	Fake video
Real video	TP	FP
Fake video	FN	TN

In the clutter matrix, the main diameter represents the number of items that are correctly classified, and the elements of the secondary diameter are the items that are not correctly classified.

A. Accuracy

The results presented in Fig. 2, and Fig. 3, prove that the proposed model has provided better results for all the measured intervals compared to the basic model in both scenarios. The improvement of the results in the proposed model shows that if the proposed model recognizes a video as real, it has a higher probability than the base model, that the video is real. On the other hand, if a video is identified as fake by the proposed model, it is more likely that the video is fake.

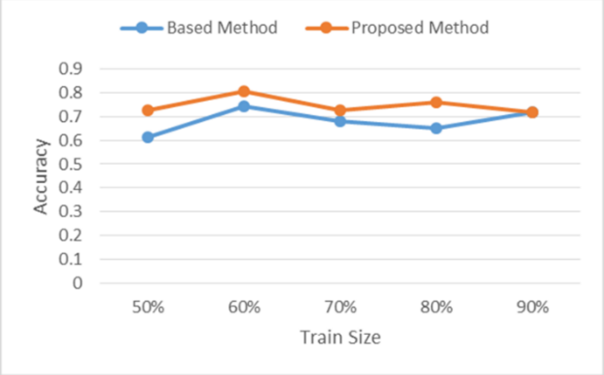

Fig. 2. Accuracy for various values of the train set size

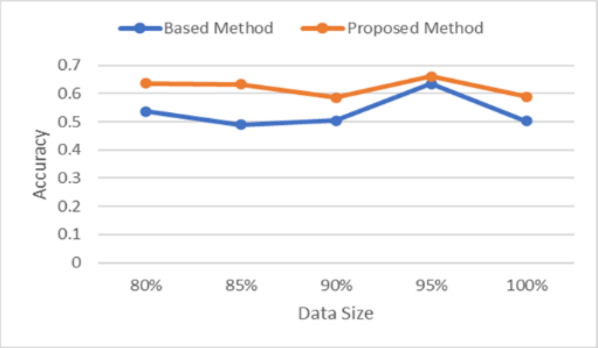

Fig. 3. Accuracy for various values of the data set size.

B. Precision

According to the results presented in Figures. above, the proposed model has been able to provide better results than the basic model in the first scenario. But in the second scenario, the two models have the same precisions. This shows that the real images discovered by the proposed model are more valid than the comparison model.

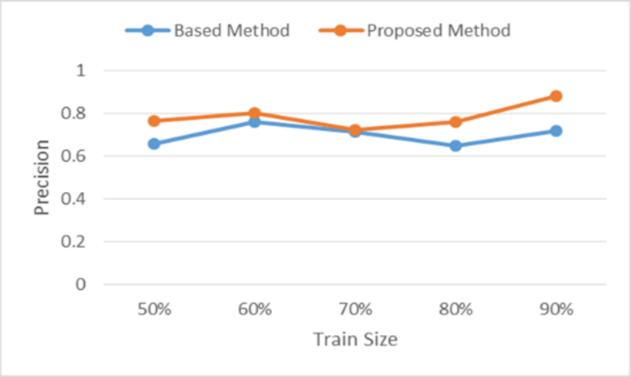

Fig. 4. Precision for various values of the train set size.

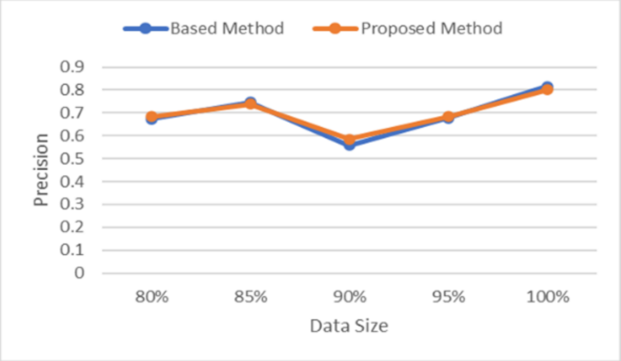

Fig. 5. Precision for various values of the data set size.

C. Recall

Fig. 4 shows that, except for the size of the training set, 90%, in the rest of the cases, this is the proposed model that provides better results than the basic model. But in the second scenario, the proposed model is better than the based model in all the data size ranges.

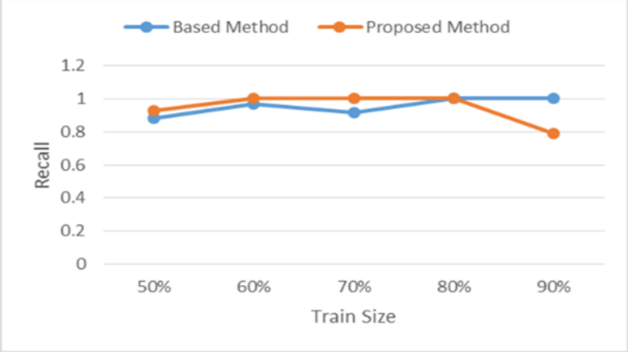

Fig. 6. Recall for various values of the train set size.

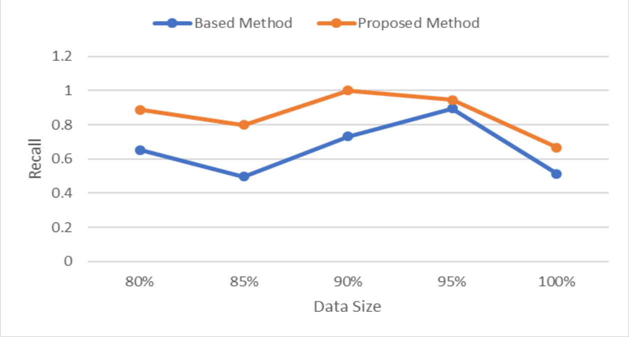

Fig. 7. Recall various values of the data set size.

Authorized licensed use limited to: Queensland University of Technology. Downloaded on November 13,2024 at 01:02:07 UTC from IEEE Xplore. Restrictions apply.

D. F-Measure

The last measure under consideration is F-Measure. The results obtained from the examination of this criterion can be seen in Fig. 5, Fig. 6, Fig. 7, Fig. 8 and Fig. 9. The results showed the proposed model was able to achieve more precision and recall than the compared model.

V. Discussion

The results obtained from the evaluations show that the proposed model for the evaluated scenario that included a change in the size of the training data set has been able to obtain better results than the basic model. Furthermore, the second scenario which investigates the impact of data size, shows the superiority of the proposed model too. To check the superiority of the proposed model, it should be noted that the use of RNN has been able to achieve better performance than the compared model due to its high power in feature extraction. The proposed model can extract the unique features of real or fake images by analyzing and performing various filtering on the images. This issue has helped the proposed model to achieve a high ability to distinguish real from fake images by increasing the power of exploring images. Table II shows a summary of the results and the degree of superiority of the proposed model compared to the basic model according to the evaluated criteria for both scenarios.

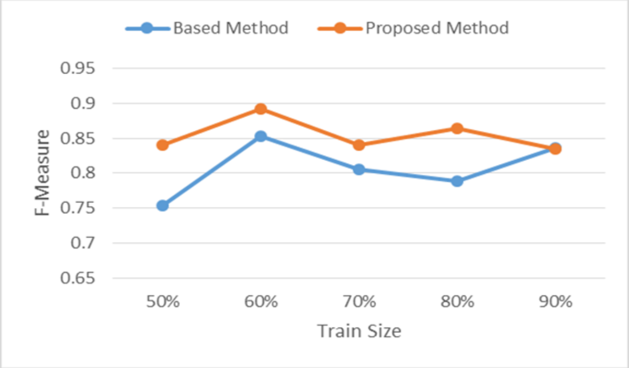

Fig. 8. F-Measure for various values of train set size.

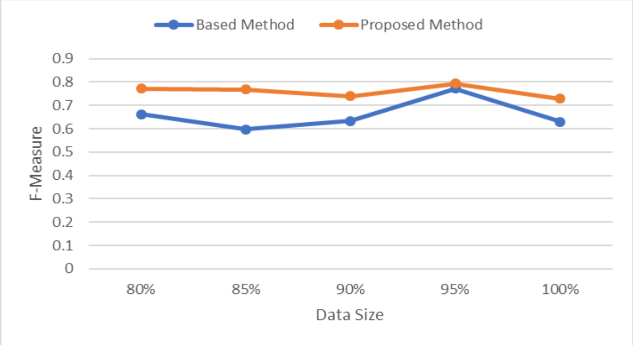

Fig. 9. F-Measure for various values of the data set size.

TABLE II. Summary of the Results

Criteria	First Scenario	Second Scenario
Accuracy	0% to 19%	-1% to 5%
Precision	5% to 23%	5% to 60%
Recall	-23% to 9%	4% to 29%
F-Measure	0% to 12%	3% to 29%

VI. Conclusion

Due to the increasing development of various tools to make fake videos from real videos. Thus, models to detect fake videos have been taken into consideration in different fields. These models try to distinguish fake videos from real videos by extracting unique features. This research has presented a model based on CNN and RNN, which tries to distinguish real videos from fake videos by extracting the features of the videos. To evaluate the proposed model, MATLAB has been used to implement the proposed model. The deepFake Images benchmark dataset is used for evaluations. This dataset contains a135 real videos as well as 677 fake videos created using different tools on real videos. For evaluation of the performance of the proposed model, two scenarios (dimensions of the training data and different sizes of the training data) have been used, and four performance metrics which are; accuracy, precision, recall, and F-Measure. The results show that the proposed model has been able to provide better results than the compared model. The accuracy has been improved by 9%, Precision, 23%, Recall, 9%, and F-Measure12% in the first scenario. While, the accuracy has been improved by 5%, Precision, by 60%, Recall, by 29%, and F-Measure by 29% in the second scenario.

References

[1] D. Güera and E. J. Delp, "DeepFake video detection using recurrent neural networks," in 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2018, pp. 1-6.

[2] A. S. Al-Khaleefa et al., "Feature adaptive and cyclic dynamic learning based on infinite term memory extreme learning machine," Applied Sciences, vol. 9, no. 5, 2019, Art no. 895.

[3] D. Wodajo and S. Atnafu, "DeepFake video detection using convolutional vision transformer," arXiv preprint arXiv:2102.11126, 2021.

[4] W. M. H. Azamuddin et al., "Quality of service (Qos) management for local area network (LAN) using traffic policy technique to secure congestion," Computers, vol. 9, no. 2, 2020, Art no. 39.

[5] D. Afchar et al., "Mesonet: A compact facial video forgery detection network," in 2018 IEEE International Workshop on Information Forensics and Security (WIFS), 2018, pp. 1–7.

[6] Y. Li and S. Lyu, "Exposing DeepFake Videos By Detecting Face Warping Artifacts," arXiv preprint arXiv:1811.00656v3, 2019.

[7] H. H. Nguyen, J. Yamagishi, and I. Echizen, "Capsule-forensics: Using Capsule Networks to Detect Forged Images and Videos," in ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 2307–2311.

[8] D. M. Montserrat et al., "DeepFakes Detection with Automatic Face Weighting," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020, pp. 2851–2859.

[9] M. S. Rana and A. H. Sung, "DeepFakeStack: A Deep Ensemble-based Learning Technique for DeepFake Detection," in 2020 7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom), 2020, pp. 70–75.

[10] J. Kim, S. Han, and S. S. Woo, "Classifying Genuine Face images from Disguised Face Images," in 2019 IEEE International Conference on Big Data (Big Data), 2019, pp. 6248–6250.

[11] D. Wodajo and S. Atnafu, "DeepFake video detection using convolutional vision transformer," arXiv preprint arXiv:2102.11126, 2021.

[12] A. S. Al-Khaleefa et al., "Knowledge preserving OSELM model for Wi-Fi-based indoor localization," Sensors, vol. 19, no. 10, 2019, Art no. 2397.

[13] S. Tariq, S. Lee, and S. S. Woo, "A convolutional LSTM based residual network for deepFake video detection," arXiv preprint arXiv:2009.07480, 2020.

[14] A. AL-Saffar, S. Awang, W. AL-Saiagh, S. Tiun, and A. S. Al-Khaleefa, "Deep learning algorithms for arabic handwriting

Authorized licensed use limited to: Queensland University of Technology. Downloaded on November 13,2024 at 01:02:07 UTC from IEEE Xplore. Restrictions apply.

recognition: A review," International Journal of Engineering & Technology, vol. 7, no. 3.20, 2018.

[15] A. S. Al-Khaleefa et al., "Mfa-oselm algorithm for wifi-based indoor positioning system," Information, vol. 10, no. 4, 2019, Art no. 146.

[16] W. Al-Saiagh, S. Tiun, A. Al-Saffar, S. Awang, and A. S. Al-Khaleefa, "Word sense disambiguation using hybrid swarm intelligence approach," PloS one, vol. 13, no. 12, 2018, Art no. e0208695.

Authorized licensed use limited to: Queensland University of Technology. Downloaded on November 13,2024 at 01:02:07 UTC from IEEE Xplore. Restrictions apply.

View publication stats