INTERNATIONAL JOURNAL
ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Voice-Authentication Model Based on Deep Learning for Cloud Environment

Ethar Abdul Wahhab Hachim [a], Methaq Talib Gaata [a,*], Thekra Abbas [a]

[a] *Department of Computer Science, College of Science, Mustansiriyah University, Baghdad, Iraq*
Corresponding author: [*]dr.methaq@uomustansiriyah.edu.iq

***Abstract*—Cloud computing is becoming an essential technology for many organizations that are dynamically scalable and employ virtualized resources as a service done over the Internet. The security and privacy of the data stored in the cloud is cloud providers' main target. Every person wants to keep his data safe and store it in a secure place. The user considers cloud storage the best option to keep his data confidential without losing it. Authentication in the trusted cloud environment allows making knowledgeable authorization decisions for access to the protected individual's data. Voice authentication, also known as voice biometrics, depends on an individual's unique voice patterns for identification to access personal and sensitive data. The essential principle for voice authentication is that every person's voice differs in tone, pitch, and volume, which is adequate to make it uniquely distinguishable. This paper uses voice metric as an identifier to determine the authorized customers that can access the data in a cloud environment without risk. The Convolution Neural Network (CNN) architecture is proposed for identifying and classifying authorized and unauthorized people based on voice features. In addition, the 3DES algorithm is used to protect the voice features during the transfer between the client and cloud sides. In the testing, the experimental results of the proposed model achieve a high level of accuracy, reaching about 98%, and encryption efficiency metrics prove the proposed model's robustness against intended attacks to obtain the data.**

***Keywords*—Cloud computing; authentication protocol; voice features; convolution neural network; cryptography.**

Manuscript received 18 Oct. 2022; revised 18 Mar. 2023; accepted 13 May. 2023. Date of publication 10 Sep. 2023.

I. Introduction

Enterprises are adapting to cloud computing at a rapid speed due to the benefits of cost-effectiveness, scalability, virtualization, and flexibility features. In the cloud environment, customers can access services based on their needs without knowing how they are delivered and where they are hosted. Cloud computing services are offered to the customers in three different models, and depending on the need of the consumer/enterprise, a suitable delivery model is deployed and adapted: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) [1]. Even though cloud computing services are growing and becoming more popular, the worry about the usage of cloud services is still an important issue. Trust and security concerns regarding the protection of critical data saved in the specific area of the cloud or the reliability of cloud software still impose restrictions on the full development of the cloud potential [2]. One of the important challenges in a cloud environment is authentication. Authentication is an important factor that plays a major role in cloud computing trust. It prevents shared information from unauthorized access. The various possible attacks on the cloud are prevented by applying different authentication mechanisms, which verify a user's identity when a user wishes to request services from cloud servers. Authentication and access control are more crucial than ever since cloud services and cloud storage may be accessible to anyone over the Internet. Nowadays, the conception of personal identity is becoming particularly important, and biometrics is a widespread approach for verification and utmost security. Authentication is a subdivision domain of security. Authentication is robustly related to availability, confidentiality, and integrity. They become commitments in the design of secure cloud systems [3], [4]. Biometric authentication is a reduced technology that can be employed in many frameworks to identify persons effectively instead of older approaches such as passwords [5]. Biometric-based authentication that depends on personal behavioral or biological characteristics is taking more attentiveness as an appropriate way to identify individuals. Biometrics is an authentication technique that can measure distinctive physical human characteristics, such as voice, fingerprint, iris, ear, and palm print. The main advantages of using biometrics are the high level of security to store sensitive data and the privacy preservation of the clients, diminishing the chances of forgery, and reduced cost-effectiveness [6].

In recent years, the convolutional neural network (CNN) has been broadly used for biometric recognition. It is a combination of procedures that a computer assesses training models to accomplish specific takes [7]. It is similar to the human brain's artificial neural networks, whereas neurons are named nodes that gather and classify the input as stated by architecture [8]. The convolutional neural network acts as a feature extractor by using the convolution and subsampling layers [9]. Integrating the artificial neural network and backpropagation algorithm simplifies the model complexity and reduces the parameters [10]. This work aims to design and apply a voice-authentication model based on the deep learning approach to establish trust among the customers and the cloud provider so that people will use cloud computing effectively and efficiently. This model will ensure that authorized persons can access their data based on the proposed biometric authentication protocol and protect the data security by the proposed encryption protocol to guarantee that others do not view it.

This paper has been organized as follows: A basic introduction to cloud computing, biometric authentication, and convolutional neural networks was provided in Section 1. Section 2 includes the related works. While Section 3 enlightens the theoretical foundation of voice preprocessing and feature extraction. Section 4 explains convolution neural networks. Section 5 shows the basic concepts of the cryptography system. Section 6 shows all steps of the proposed model. Section 7 contains the obtained results and the security evaluations. Finally, Section 8 shows the conclusion.

In 2019, Abed et al. [11] proposed a new bio-cryptographic way to get a more safe and utilizable cloud using obvious biometric modalities. Therefore, a set of fundamental investigations have been assumed and executed to detect the probable contribution of the proposed system. The earliest sequence of experiments focuses on examining how reliable the new bio-cryptographic system is in key generation from biometric modalities. The second set of experiments investigates the possibility of improving the security level of the generated key. The conclusive experiments examine the capability of generating various cryptographic key sizes during the vector of biometric features.

In 2020, Sudhakar and Gavrilova [12] worked on the AWS cloud platform, a distinctive cancelable biometric model that depends on machine learning. The cancel-able system's cancelable database, deep learning module, and biometric engine are all off-loaded to the cloud. The cloud-based parallel processing of a CNN makes it computationally faster than single systems with high accuracy. CNN was used on the cloud to find cross-fold biometric attributes, which were later transformed into cancelable templates through random projection. In the case of user verification, a unique MLP architecture was developed to achieve the suggested model.

In 2020, Hedaia et al. [13] proposed a new and different authentication method named Bio-CAPTCHA. They suggested using voice-based passwords randomly, which frequently changes every time the user attempts to implement the login process, reducing the opportunity for illegal access. The experimental results and hypothetical investigation emphasized the extremely high level of security requirement. The key contribution of their work is to establish a newfangled authentication method that mixes security and usability based on voice identity.

In 2020, Phipps et al. [14] explained the specifics an attack modeling applies and proposed a system to address the main threats while keeping the usability correlated with voice based on an additional voice-based authentication issue. Their proposed approach investigated current attacks, and usability restrictions and proposed a new technique of authentication for users of intelligent speakers and another digital voice system. The study has identified many attacks and, during threat modeling and outlines, a group of mitigations and a theoretical model that gives an extra factor based on voice for authentication based upon the new grouping of existing methodologies such as voice hiding, cryptosystem, GIS, and cloud machinery.

In 2021, Vinoth et al. [15] computed the data security using the biometric of client parameters and creating a biometric key. They use the client's fingerprint to create the biometric key by the QR code diversion. Clients are authenticated via transmitting the biometric key to clients. Then the personal data is transmitted to authenticated clients. The algorithm improves the processing time in the implementation stage and provides additional reliability and robust encryption methods.

In 2022, Mihailescu and Nita [16] presented a system consisting of three elements (biometric authentication, traditional authentication, and searchable encryption). They have established that the proposed way is very well for traditional and advanced network infrastructures. They computed the correctness of the processes to ensure that the cloud servers returned the proper documents. The highest difficulties met throughout this study were associated with the cloud environment. Due to the complication of the authentication protocol, those tools supplied them with inconceivable knowledge.

II. Material and Method

A. Voice Preprocessing and Feature Extraction

Extracting the optimum parametric representation of voice data is critical for improving recognition performance. The preprocessing of voice signals consider the essential step that needs before taking features from the signals [17]. Mel Frequency Cepstrum Coefficient (MFCC) is an approach for feature extraction from voice signals. In comparison, feature extraction is a procedure for determining the value of the vector that can be considered as a person's identity. The MFCC is one of the main methods used in different applications of the speech processing domain since it is considered quite well for representing the signals [18]. The MFCC feature extraction technique depends on human hearing perception. MFCC includes the following stages. Apply a pre-emphasis filter to increase the energy at a higher frequency of the voice signal, segment the voice signal into short frames with a length of 25ms and overlapping of 50%, and each frame will treat independently from the others.

Applying hamming window to all frames to minimize the discontinuities of the voice signal, get the spectrum and transfer all frames from the time domain to the frequency domain by applying Fast Fourier Transform and then warping the frequencies on a Mel scale. Followed by applying the Discrete Fourier transform to remove unnecessary data in the signal and produces a set of cepstral coefficients. Fourier analysis measurable factors are the ground of the method that is used to create feature vectors in most voice recognition frameworks in the digital processing of voice signals in a spectral domain [19].

Also, four other features were extracted by using Mean Frequency to detect the average across the whole frequency signal, Standard Division to measure the diversity in the voice dataset, amplitude to measure the sum of energy, and finally, find Zero-crossing feature to obtain more discriminatory features that will greatly contribute to the process of distinguishing data later [13].

B. *Convolution Neural Network (CNN)*

Deep learning strategies can detect the essential features without characterizing particular features and depending on the convolutional neural networks (CNNs) [20]. Previously Convolutional Neural Network (CNN) was constructed for object recognition and then it was used in image classification and segmentation, voice recognition, digital signals prediction, etc. Because of its autonomous functioning nature, it can be considered one of the most substantial deep learning tools for artificial intelligence and computer vision [21]. 1D CNNs beat 2D CNNs in situations with minimal categorized data and large signal variations from numerous sources. The main difference between 1D CNNs and 2D CNNs is that in both the function mappings and the kernels, one-dimension arrays exchange with two-dimension matrices. The CNN layers are trained to extract features used by the MLP layers in the classification phase by evaluating the raw one-dimension data as input [21].

There are three main layer types for building the network architecture. They are a convolutional layer, a pooling layer, and a fully connected layer. One of the most powerful features of convolutional neural networks is their use of shared weights, a group of connections that share the same weights instead of using different weights for each connection. The other feature is the local connection: each neuron does not contact all the neurons in the previous layer. Still, it only contacts a specific group of neurons to see if they contain the object's feature instead of contacting all cells. This produces strong responses to obtain local characteristics in a voice data Input. These two features exceedingly reduced the number of parameters in the network, and thus the training time will be reduced [10].

As a result, feature extraction and classification procedures are merged into a solitary process that may be tweaked to improve the efficiency of the classification. The most obvious benefit of 1D CNNs their low computational complexity, as the sole cost-effective technique is a series of 1D convolutions that are just linear weighted sums [21].

C. *Cryptography System*

One of the biggest concerns with cloud adoption today is data protection. Many cloud-based applications that people use in their everyday life (not only cloud storage applications but also banking applications, government applications, telecom sector applications, etc.) manage huge volumes of personal data. In many cases, applications enable users to upload data and share it with other users, thus making the data protection problem even more complex. Data may be interfered with by any suspicious entity [22].

Cryptography is a crucial mechanism to protect computer data and information that is communicated using cloud computing. Cryptography is an arty data transformation into an unreadable form, so only the intended recipient can understand and use it [23]. A cipher is a set of two algorithms, the encryption algorithm that transforms a plaintext into unreadable text called cipher text and the decryption algorithm that converts ciphertext to plaintext in the opposite direction of encryption. These are used to create the encoding and decoding procedures. In cryptography systems, there are two major types of encryption algorithms, symmetric algorithms/ asymmetric algorithms. This classification depends on the use of encryption keys. Symmetric cryptography (Secret or Private Key) employs only one key for encryption and decryption, providing stronger security and faster operation, but transferring a key between partners is impossible. They are classified into two major types, block cipher, and stream cipher. In block cipher algorithms, the data (plaintext) is encrypted block by block (set of bytes) by using a secret key such as Data Encryption Standard (DES), Triple Data Encryption Standard (3DES or TDEA), Advanced Encryption Standard (AES), International Data Encryption Algorithm (IDEA), Blowfish and Twofish [24].

On the other hand, stream cipher algorithms encrypt the data (plaintext) bit by bit (or byte by byte) at a time, such as Chacha, Salsa20, and Rivest Cipher 4 (RC4) [14]. Opposed to symmetric cryptography, asymmetric cryptography uses two keys; the first is a private key that is never communicated over the network, while the second is a shared public key such as Digital Signature Algorithm (DSA), Rivest Shamir Adleman (RSA), Diffie-Hellman and Elliptic Curve Cryptography (ECC). This cryptography is more difficult and time-consuming than symmetric cryptography[25].

D. *The Proposed Model*

There are two sides to the proposed model. These sides are named client-side and cloud-side. In order to attain the model objectives and meet the main goal, each side has its sub-stages, as shown in Fig. 1.

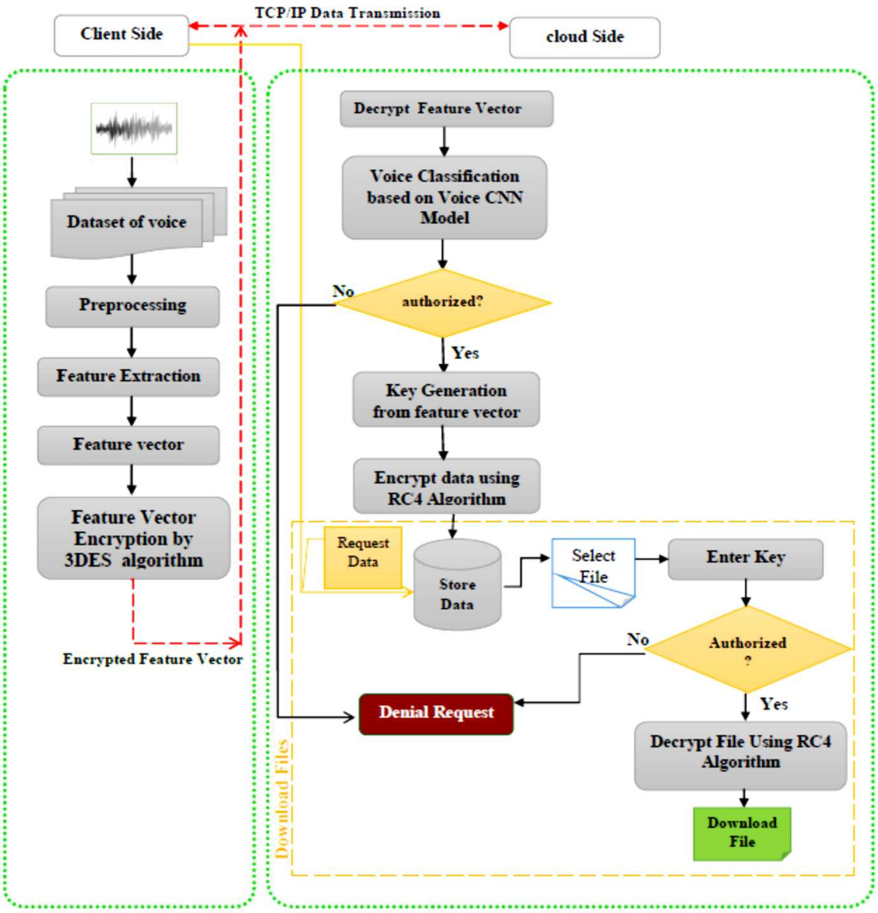

Fig. 1 The proposed model

On the client side, a set of operations includes reading the voice dataset and then implementing the preprocessing procedures of these voice data to extract the feature vectors from voice data equal to 16 features (12 MFCC features, Mean, SD, zero crossing, and amplitude). The extracted feature vector has two vital tasks, identify the client and can also be used as a secret key for an encryption algorithm to encrypt the data before storing it in the cloud.

In this proposed model, the voice dataset has been split into two parts (training, and testing) and then preprocessing to increase the data's quality and improve performance. Then feature extraction processes are applied to obtain a feature vector of 16 features. After that, the extracted feature vectors encrypt using the 3DES algorithm. Then the encrypted feature vector is sent together with the data (e.g., image) that the user wants to upload to the cloud by TCP/IP protocol.

On the cloud side, the first step is to receive the request, and this request includes the encrypted feature vector and the data to be uploaded to the cloud. After receiving the request, the encrypted feature vector is separated from the data and then decrypted for use in a Voice CNN model to authenticate individuals. If the person is authorized, the key has been created depending on the feature vector, and then using this key in RC4 algorithm to encrypt data and store it in the database. If the person is not authorized, he has been rejected. For downloading data from the cloud, wholly the operations are done in inverse, making it possible for the user to download the file he/she wants.

E. *Voice Dataset Description*

The voices dataset used in this model is called Prominent leader's speeches. It includes voice clips of five country leaders, and every leader has 1500 voice clips. It was downloaded from the Kaggle website. Table 1 explains the details of the dataset.

TABLE I
VOICE DATASET DESCRIPTION

Dataset's name	**Format of the file**	**Size of the file**	**Number of samples**	**Downloaded from**
Prominent leader's speeches	. Wav	292 MB	7500	Kaggle website

F. *Voice Data Splitting*

Data splitting is particularly important in data science, especially for creating models based on data. The process of dividing accessible datasets into two parts by using cross-validation way. The first set of data that used to build the predictive model, whereas the second was used to test and evaluate the proposed model. Part of the data (70%) is customized for the training stage, and the remaining data (30%) is customized for the testing stage.

G. *Voice* CNN Model

The 1D convolutional network architecture is proposed in this work. It is constructed of sixteen layers (the last two layers are Flatten layer, and the Dense layer is the output layer). The input layer is the passive layer, which receives the preprocessed one-dimension data represents the feature that extracted from the voice, whilst the output layer is a really fully connected layer (all neurons in this layer receive the input from entire neurons in the preceding layer) with a number of neurons that equal to the number of the groups.

Convolution layers, which contain kernels with the LeakyReLU activation function (with alpha=0.3) used to acquire feature maps, which equal the number of filters utilized. After the convolution is performed, the subsampling procedure takes a turn. The pooling layers are used after one or a small number of convolutional layers (in this proposed model, one pooling layer is used after one convolution layer) to decrease the amount of input whenever the network goes deeper. In every input size reduction followed by a pooling operation, the amount of network computing required is reduced. These procedures are repeated until the last layers are reached. A fully connected layer was used to match the neural network's output size to the desired output size. In the supervised learning setting, the output is then passed through a SoftMax function to produce a probability representation for the predictions for each class. The output from the final convolutional layer is usually flattened out, or the feature maps are subsampled to a size of one to use the fully connected layer, as shown in Fig.2.

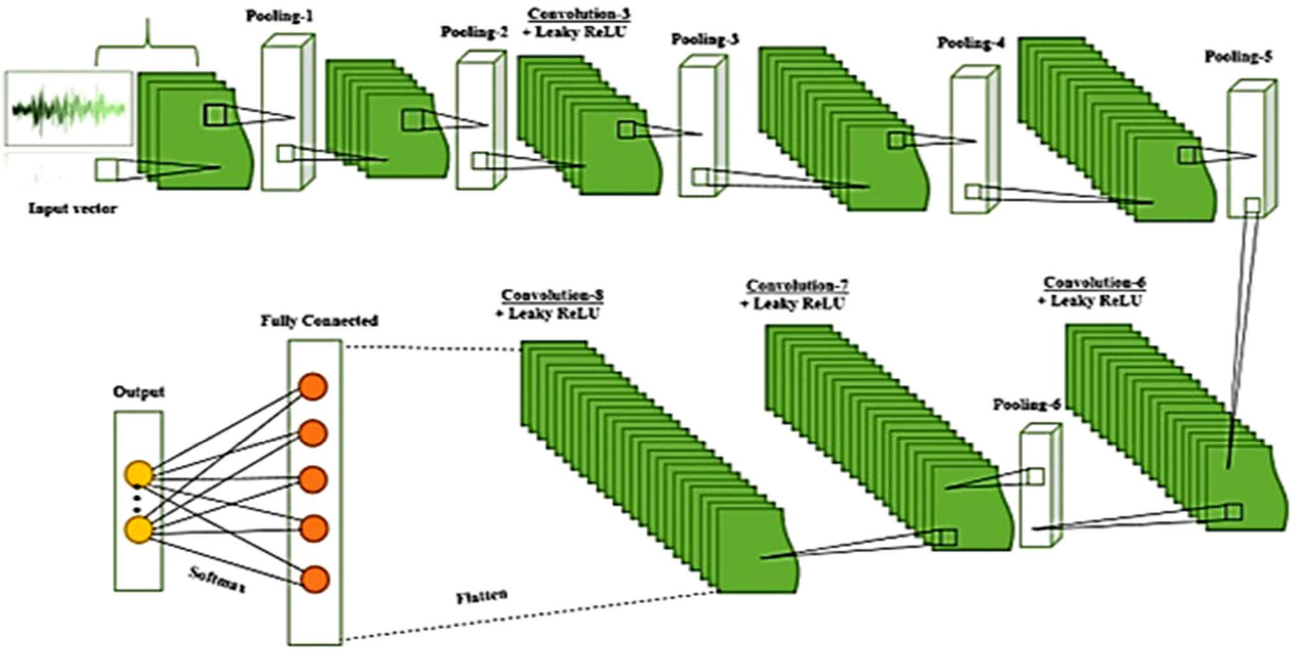

Fig. 2 The proposed Voice CNN model

This model includes two stages: training and testing. In training, the model learns to classify the clients depending on the feature vector of the voice signal received from the client side and reach high accuracy. In testing, the performance of the model is evaluated. One of the most important hyperparameters in this proposed model is the learning rate that controls updating the model weights. Selecting the proper learning rate is a challenge because if it is too low, the training will progress slowly with a little weight update. The proper learning rate for this proposed model is 0.001(often, the range is 0.0 - 1.0).

H. Data Encryption

To ensure the confidentiality and security of the data sent to the cloud, the encryption process in this proposed model was done on two levels using two encryption algorithms (3DES and RC4). Each one is according to the purpose for which it is used. After extracting the feature vector from voice data, these extracted features are protected by encrypting them using the 3DES algorithm before sending them into the cloud with the TCP/IP protocol. 3DES is a cipher block-type symmetric cryptography algorithm. Symmetric cryptography is a mechanism for encrypting and decrypting data that uses the same key. Cipher block cryptography is symmetric cryptography with a fixed bit size. In the case of 3DES, three keys are needed for encryption and decryption procedures, with the size of 64 bits for each key. After verifying the identity of the clients, they can now access the cloud and then the data (e.g., image) that is saved in the cloud after being encrypted using RC4 algorithm. RC4 is a type of stream cipher that uses the same key in encryption and decryption operations, which consist of two main algorithms. In the key-scheduling algorithm (KSA), the initialization stage is turned, and random permutations with a secret key are performed. While the keystream has been generated in Pseudo Random Key Generation Algorithm (PRGA), the plain message is XORed with the generated key stream. In this proposed model, the secret key generated from the same feature vector of the client increases the security level of the RC4 encryption algorithm.

I. Voice-Authentication Model Algorithm

Algorithm 1 presents details for the overall steps in the voice-authentication model. As mentioned earlier, in step 1, the input to the proposed model is the voice dataset. In steps 2 and 3, the preprocessing and feature extraction procedures were performed. Then 3DES encryption algorithm is applied to extract feature vectors before sending them to the cloud to authenticate the client uses the vector to generate the key that will be used later as a seed for RC4 encryption algorithm to encrypt the data (e.g., image). In the cloud side Voice CNN model executed as mentioned in step 7, the output is identifying the person and whether he has the authorization to access the data stored on the cloud. When a person can access the cloud, he can download or upload data to the cloud, as explained in step 8.

```
Input: voice dataset
Output: voice identification
Begin
1: Read voice data
2: Preprocessing for voice data
3: Extract the features (MFCC, Mean Frequency, Standard
      Division, Amplitude and Zero-Crossing Rate).
4: Encrypt the features voice by the 3DES algorithm.
5: Transmit the data (image) and the encrypted features
      vectors to the cloud side.
6: Decrypt features vectors by the 3DES algorithm.
7: Apply the Hold-Out-Cross-Validation technique to
      separate data.
        • Feed the features set with targets to CNN.
        • Check if the client is authorized, then:
        • Generate the encryption key based on feature
          vector.
        • Encrypt the data by the RC4 algorithm and
          save the data in the cloud then go to step 8.
      Else: Go to step 9.
8: Download the data (image) from cloud:
        • The client enters his/her biometric for
          verifying then:
        • If a client was authorized, then:
        • Downloading data file.
       Else: Go to step 9
9: End
```

III. Results And Discussion

The most important results attained from this model show the high level of accuracy of the system in both the authentication and the security of the data stored in the cloud concerning voice recognition. Certain parameters are used to evaluate the system's performance to determine its behavior. There are two groups of metrics, classifiers Performance Metrics, and Encryption Performance Metrics. For the Voice CNN model, the accuracy was 97.6%, precision was 98.2%, and recall was 97.8%, so the accuracy for the overall proposed model was about 98%, as shown in Table 2.

TABLE II
Accuracy of the proposed system [26]

Metric	Result
Accuracy	97.9%
Precision	98.2%
Recall	97.8%

As for the strong points of system encryption, a group of metrics evaluated encryption algorithm performance. The results, depending on the standards that were customized to measure the security of the encryption method, are very confident. For improved encryption/decryption efficiency, PSNR should be higher, and MSE should be lower between the original and encrypted image. In all situations, the MSE value is 0, indicating that the original image was entirely retrieved. PSNR, on the other hand, will be infinite in this scenario because the value is divided by zero, as presented in Table 3.

TABLE III
Encryption metrics

Metric	Result
PSNR [27]	Inf.
SSIM [28]	1.0
MSE [27]	0.0
RMSE [27]	0.0
NRMSE [29]	0.0
Entropy [30]	7.5663

IV. Conclusion

One of the most important issues related to using the cloud environment is that only authorized users have the right to access the services provided by the cloud. Traditional authentication methods like passwords and others can easily be accessible. Although these traditional methods have some advantages, there are many disadvantages, such as the hackers' ability to obtain the passwords, which can be readily used by the user, who may forget the password if he uses a complex one. For these and many other reasons, it can be considered the biometric authentication methods that enable users to use their biometrics features to verify their identity and acquire access to the cloud. This method can be the most preferred regarding authentication and security issues. Using biometric authentication techniques ameliorates the security level in a cloud environment. Voice-based authentication model for the cloud environment is designed in this paper.

We can conclude that biometric authentication techniques offer unique methods for authenticating clients in a cloud environment, depending on their observations. CNN-1D is the most dependable of the models to achieve the best outcomes regardless of dataset size or type, with an average accuracy of about 98% and reduces an average execution time of sec. Furthermore, give a robust platform to future studies of the cloud environment using other encryption algorithms and biometric traits such as face, hand palms, iris, etc.

References

[1] M. Aljanabi *et al.*, "Cloud Computing Issues, Challenges, and Needs: A Survey," *Int. J. Informatics Vis.*, vol. 5, no. 3, pp. 298–305, 2021, doi: 10.30630/JOIV.5.3.671.

[2] I. Indu, P. M. R. Anand, and V. Bhaskar, "Identity and access management in cloud environment: Mechanisms and challenges," *Eng. Sci. Technol. an Int. J.*, vol. 21, no. 4, pp. 574–588, 2018, doi: 10.1016/j.jestch.2018.05.010.

[3] L. Megouache, A. Zitouni, and M. Djoudi, "Ensuring user authentication and data integrity in multi-cloud environment," *Human-centric Comput. Inf. Sci.*, vol. 2020, no. 1, p. 10, doi: 10.1186/s13673-020-00224-yï.

[4] J. Mohamed, M. F. Md Fudzee, S. N. Ramli, M. N. Ismail, and - Defni, "Bridging Usability and Accessibility of User Authentication using Usable Accessed (UAce) for Online Payment Applications," *JOIV Int. J. Informatics Vis.*, vol. 5, no. 4, p. 366, Dec. 2021, doi: 10.30630/joiv.5.4.740.

[5] M. Albanese, A. De Benedictis, D. D. J. de Macedo, and F. Messina, "Security and trust in cloud application life-cycle management,"

Future Generation Computer Systems, vol. 111. 2020, doi: 10.1016/j.future.2020.01.025.

[6] D. Raja, T. Bhalodia, and R. Buch, "Review On Biometric Two-Way Authentication In Cloud Computing," 2020.

[7] N. Sharma, R. Sharma, and N. Jindal, "Machine Learning and Deep Learning Applications-A Vision," *Glob. Transitions Proc.*, vol. 2, no. 1, pp. 24–28, 2021, doi: 10.1016/j.gltp.2021.01.004.

[8] M. Mahmud, M. S. Kaiser, A. Hussain, and S. Vassanelli, "Applications of Deep Learning and Reinforcement Learning to Biological Data," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 29, no. 6, pp. 2063–2079, 2018, doi: 10.1109/TNNLS.2018.2790388.

[9] M. M. Taye, "Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions," *Computation*, vol. 11, no. 3, p. 52, Mar. 2023, doi: 10.3390/computation11030052.

[10] S. T. Ahmed and S. M. Kadhem, "Using Machine Learning via Deep Learning Algorithms to Diagnose the Lung Disease Based on Chest Imaging: A Survey," *Int. J. Interact. Mob. Technol.*, vol. 15, no. 16, pp. 95–112, 2021, doi: 10.3991/ijim.v15i16.24191.

[11] L. Abed, N. Clarke, B. Ghita, and A. Alruban, "Securing cloud storage by transparent biometric cryptography," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11359 LNCS, pp. 97–108, 2019, doi: 10.1007/978-3-030-12942-2_9.

[12] T. Sudhakar and M. Gavrilova, "Cancelable Biometrics Using Deep Learning as a Cloud Service," *IEEE Access*, vol. 8, pp. 112932–112943, 2020, doi: 10.1109/ACCESS.2020.3003869.

[13] Omer Ahmed Hedaia, Ahmed Shawish, Essam H. Houssein, and Hala Zayed, "Bio-CAPTCHA Voice-Based Authentication Technique for Better Security and Usability in Cloud Computing ," *Int. J. Serv. Sci. Manag. Eng. Technol*, vol. 11, no. 2, pp. 59–79, 2020.

[14] A. Phipps, K. Chiazzane, and V. Vassilev, "Your password is music to my ears: Cloud based authentication using sound," *Proc. Conflu. 2021 11th Int. Conf. Cloud Comput. Data Sci. Eng.*, pp. 227–232, 2021, doi: 10.1109/Confluence51648.2021.9377126.

[15] K. M. Vinoth, K. Venkatachalam, P. Prabu, A. Almutairi, and M. Abouhawwash, "Secure biometric authentication with de-duplication on distributed cloud storage," *PeerJ Comput. Sci.*, vol. 7, pp. 1–20, 2021, doi: 10.7717/PEERJ-CS.569.

[16] M. I. Mihailescu and S. L. Nita, "A Searchable Encryption Scheme with Biometric Authentication and Authorization for Cloud Environments," *Cryptography*, vol. 6, no. 1, pp. 1–22, 2022, doi: 10.3390/cryptography6010008.

[17] W. Jia, M. Sun, J. Lian, and S. Hou, "Feature dimensionality reduction: a review," *Complex Intell. Syst.*, vol. 8, no. 3, pp. 2663–2693, 2022, doi: 10.1007/s40747-021-00637-x.

[18] D. Anggraeni, W. S. M. Sanjaya, M. Y. S. Nurasyidiek, and M. Munawwaroh, "The Implementation of Speech Recognition using Mel-Frequency Cepstrum Coefficients (MFCC) and Support Vector Machine (SVM) method based on Python to Control Robot Arm," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 288, no. 1, 2018, doi: 10.1088/1757-899X/288/1/012042.

[19] A. B. Abdusalomov, F. Safarov, M. Rakhimov, B. Turaev, and T. K. Whangbo, "Improved Feature Parameter Extraction from Speech Signals Using Machine Learning Algorithm," *Sensors*, vol. 22, no. 21, Nov. 2022, doi: 10.3390/s22218122.

[20] A. A. Hezam, S. A. Mostafa, Z. Baharum, A. Alanda, and M. Z. Salikon, "Combining Deep Learning Models for Enhancing the Detection of Botnet Attacks in Multiple Sensors Internet of Things Networks," *JOIV Int. J. Informatics Vis.*, vol. 5, no. 4, p. 380, Dec. 2021, doi: 10.30630/joiv.5.4.733.

[21] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mech. Syst. Signal Process.*, vol. 151, p. 107398, 2021, doi: 10.1016/j.ymssp.2020.107398.

[22] K. Limniotis, "Cryptography as the means to protect fundamental human rights," *Cryptography*, vol. 5, no. 4. MDPI, Dec. 2021, doi: 10.3390/cryptography5040034.

[23] M. T. Gençoğlu and M. T. Gençoğlu, "Importance of Cryptography in Information Security," *ISOR J. Comput. Eng.*, vol. 21, no. 1, pp. 65–68, 2019, doi: 10.9790/0661-2101026568.

[24] C. Atika Sari, E. H. Rachmawanto, and C. A. Haryanto, "Cryptography Triple Data Encryption Standard (3DES) for Digital Image Security," *Sci. J. Informatics*, vol. 5, no. 2, pp. 105–117, 2018, doi: 10.15294/sji.v5i2.14844.

[25] L. K. Yee and C. C. Wen, "Secret channel using video steganography," *Int. J. Informatics Vis.*, vol. 1, no. 4–2, pp. 240–245, 2017, doi: 10.30630/joiv.1.4-2.71.

[26] Anna Berger and Sergey A. Guda, "Threshold optimization for F measure of macro- averaged precision and recall," *Pattern Recognit.*, vol. 102, pp. 107–250, 2020.

[27] M. T. Gaata, M. T. Younis, J. N. Hasoon, and S. A. Mostafa, "Hessenberg factorization and firework algorithms for optimized data hiding in digital images," *J. Intell. Syst.*, vol. 31, no. 1, pp. 440–453, 2022, doi: 10.1515/jisys-2022-0029.

[28] B. J. Saleh, A. Y. F. Saedi, A. T. Q. Al-Aqbi, and L. A. Salman, "Optimum median filter based on crow optimization algorithm," *Baghdad Sci. J.*, vol. 18, no. 3, pp. 614–627, 2021, doi: 10.21123/BSJ.2021.18.3.0614.

[29] Logs Dongyu, Sixuan Wu, and Mingcai Hou, "Fully connected deep network: An improved method to predict TOC of shale reservoirs from well logs," *Mar. Pet. Geol.*, vol. 132, pp. 105–205, 2021.

[30] A. Ben-Naim, "Entropy and Time," *Entropy*, vol. 22, no. 4, 2020, doi: 10.3390/E22040430.