Survey of Speech Steganography in Audio Signal Based on Deep Learning Techniques

Hind I.Mohammed[1, a)] Saja Salim Mohammed[2,b)] Saad Albawi[3,c)] Thuraya S. Abdulrasool[4,d)]

[1,4] University of Diyala /Al-Muqdad college of education- Department of Mathematics.
[2] University of Diyala/ College of Physical Education and Sport Sciences- Department of Theoretical Sciences.
[3]Department of Computer Engineering, College of Engineering , University of Diyala, Diyala , Iraq

[a)]Corresponding Author: hindim@uodiyala.edu.iq
[b)] saja.salim@uodiyala.edu.iq
[C)]saadalbawi@uodiyala.edu.iq
[d)] Thurayasadoon@gmail.com

Abstract. One of the most significant areas of computer science, steganography deals with the imperceptible and undetectable concealment of information. Whereas, steganography analysis finds hidden information in observable data without knowledge of method. This is especially true after tying the mechanism of concealing or protecting data with deep learning (DL) algorithms due to their significance, efficiency, and accuracy. Deep learning has recently produced cutting-edge achievements in a variety of sectors. The Convolutional Neural Network (CNN), the Deep Neural Networks (DNN), and the Support Vector Machine (SVM) are three of the most significant deep learning algorithms employed by researchers in the study of steganography. However, introduced the audio steganalysis based approaches of deep learning. The present study, offer a thorough analysis of prior research on voice masking in audio data using DL techniques.

Keywords: Steganography, Deep Learning (DL) algorithms, the Convolutional Neural Network (CNN), the deep neural networks (DNN), and Support Vector Machine (SVM).

INTRODUCTION

Steganography is the science of concealing messages within other messages ("steganos" - concealed or covered, "graphein" - writing)[1], [2]. It is typically used to deliver concealed "secret" communications to recipients who are aware of their presence while concealing their very existence from other ignorant parties who only view the "public" or "carrier" message[3] [4].Steganography is a technique for concealing potentially publically accessible data in a host media. While earlier research concentrated on unimodal configurations (e.g., hiding images in images or sounds in audio)[5].Recent years have seen deep learn DLing surpass previous best achievements in a variety of areas. However, only a handful of DL based approaches to audio steganography have been presented so far[6].The concealment of information within audio is a relatively unexplored area of study. Earlier audio steganographic endeavors relied on signal processing and audio coding[7] [8][9] [10], but recent advances rely on DL[11]. There are a number of early approaches that encode the concealed signal within the audio's LSB (perceptually least significant bits)[10].DL or deep neural network (DNN) refers to multiple-layered artificial neural networks (ANNs) [13][14]. In the previous years, it has been regarded as one of the most potent instruments, and it has gained a great deal of notoriety in the academic community due to its capacity to manage vast quantities of data. Recently, in a number of disciplines, interest in deeper concealed layers has begun to outperform conventional techniques [15] [16].CNN is one of the most widely used DNN. It gets its name from the linear arithmetic technique called convolution between matrices, and it is a type of Artificial Neural Network [6][17].CNN has several levels: Convolutional, nonlinear, pooling, and fully linked layers are among them [18]. Nonlinear and aggregate layers don't have parameters, but convolutional and fully linked layers do. CNN does a great job with ML tasks.[19] [20].Three different steganographic methods are used to test these steganalysis methods. (GMMs), (DBNs), and Recurrent Neural Networks (RNNs)[8]. SVM are a supervised learning algorithm ML[21] that are used as a classifier. Accurately predict a sample's label value from a testing database used only the training database that is the goal of a model. The study input the training data and label into a linear hyperplan to train SVMs. The study goal of partitioning the training data into two groups, one for each label. In the event that this is not possible, we project the data onto a plane in which it is [10]. Convey information in everyday life voice is one of the common ways. Thus, use sound as a shorthand vector it is of practical importance for this study. In general, the current method ML can be divided into three steps[10]: 1) "remaining account; 2) Extract feature. and 3) classification. Among these steps, designing appropriate features plays a crucial role" [9]."Contrastive divergence trains the first RBM's visible layer with training data. Whereas, the visible layer is hidden layer from the last RBM trained and second RBM is trained with various deviation. The RBM will then offer DBN initial weights. Technically, all concealed layers have the same units. This feed-forward network is trained through backpropagation"[22] [11].

Review of the Literature

A.Janicki et al. (2014) [23]: to identify audio signs built Transteg used MFCCs and GMMs. Acoustic analysis begins with MFCC parameters. Whereas, to generate two models for unblemished and clear signals the study utilized GMMs. Each signal's MFCC parameters' model

probability calculated. Probabilities evaluated to classify the signal with the highest model probability. Five speech corpora—including TIMIT—have been examined with this approach.

Kocal, O. H et al. (2016) [24]: employed a combination of method (SFFS) and SVMs by used TIMIT corpus. At the same time, the study applied delay vector variance (DVV) features derived from surrogate data to detect the presence of a stego-signal. The current study applied nine watermarking including StegHide and Hide 4PGP. The greatest detection possibility is hundred percent, and the lowermost is seventy-seven percent.

H. Ghasemzadeh and M. K. Arjmandi (2014)[25]: proposed a new steg analysis method and a combination of Reversed-Mel cestrum by employed audio Steg analysis and SVMs performed. Steganographic techniques, Hide4PGP and StegHide have been tested by the current method.

Paulin et al.(2016) [6]: first used DBN for speech steganalysis. DBN classified the suspicious signal after calculating MFCC. The neural network was employed as a classifier, not for adaptive features extraction, hence detection accuracy improved.

Paulin, C et al. (2017) [26]: compared four speech/audio steganalysis methods. Mel-Frequency Cepstral Coefficients (MFCCs) analyze audio files acoustically. Steganalyzers include SVMs, GMMs, DBNs, and RNNs. Three steganographic methods test these steganalysis methods. They used StegHide, Hide4PGP, and FreqSteg to steganograph the Noizeus corpus. GMMs had the highest classification rate without errors, followed by SVMs, DBNs, and RNNs.

Hamzeh G et al. (2016) [27]: applied (Steghide@1.563%) findings showed that the analyzed contained 97.8% and 94.4% in the targeted and universal scenarios, respectively. The current finds got better results than the previous D2-MFCC- method by 17,3% and 20% rates.

Li, S. et al. (2017)[28]: pointed out that QIM steganography changes the correlation properties of split vector quantization (VQ) codewords of LPC filter coefficients. We build the Quantization Codeword Correlation Network (QCCN) using split VQ codewords from adjacent speech frames. Pruning the QCCN model strengthens the correlation network. The study derived steganalysis-sensitive feature vectors by calculating the correlation properties of pruned correlation network vertices. Finally, design a high-performance SVM classifier detector. Experimental results showed that the proposed QCCN steganalysis approach may detect QIM steganography in encoded speech stream when applied to low-bit-rate speech codecs as G.723.1 and G.729.

Chen et al. (2017)[29]: CNN used to make a system for end-to-end audio steganography. Taking into account the benefits of Long Short-Term Memory (LSTM) when dealing with time series.

Wang et al. (2017) [30]: introduced a new CNN-based MP3 steganalysis approach that outperforms the EECS algorithm.

Lin et al. (2018) [31]: suggested LSTM version RNN-SM to discover CNV-QIM steganography in VoIP.

Kreuk, F. et al. (2019) [3]: employed deep neural networks to steganograph voice data and proved that vision-based models are unsuitable for speech. They suggested a new model that includes the short-time Fourier transform and inverse-short-time Fourier transform as differentiable layers in the network, placing a crucial constraint on network outputs. statistically and qualitatively compared the proposed method to DL on numerous speech datasets. revealed that several decoders or a conditional decoder might be used to conceal numerous messages in a carrier. Finally, tested this model under channel distortions. Qualitative investigations imply that human listeners cannot detect carrier alterations and that decoded communications are extremely comprehensible.

Das Abhishek et al. (2021) [32]: The objective was to encode and decode multiple secret images within a single cover image of the same resolution using deep neural networks.

Bhangale, K. B., & Mohanaprasad, K. (2021)[33]: explored the feature extraction methods and classifiers created by ML that are used in voice processing and recognition tasks. On the Berlin EmoDB database, speech emotion recognition application is validated to check the effectiveness of various machine learning approaches. It also details the diverse application domains and difficulties of ML for voice processing.

PERFORMANCE ANALYSIS

In recent years, valuable research efforts in data steganography have revealed the easiest and

distinctive way of human-computer interaction using the latest methods based on DL and its important algorithm. Challenges in ML for speech processing. Some previous studies indicated that DL models for image masking are less suitable for audio data, as information masking in sound is a relatively unexplored research field. The researchers used more than one audio data set and took the most efficient model in terms of accuracy and result.

They used the SVM algorithm using watermarking techniques with better results compared to the results of other ML algorithms. But when applying the DBN algorithm, they were used only as a classifier and not in the feature extraction process. Whereas, using the three algorithms DBN, GMM and SVM together on the same data set, GMM was the best performer to obtain an ideal classification rate. At the same time, SVM, as GMM used to improve the feature dimensions' values, and the final decision is taken by SVM to discover the audio files.

CONCLUSION

This paper succeeded in presenting the most prominent algorithms that were applied to hide audio data and reveal the highest results in data hiding. For the SVM algorithm and the CNN algorithm, there is a need to apply more than one data set to choose the highest accuracy. At the present time, the process of creating or preparing a data set by the researcher is important and distinctive to highlight the efforts of researchers in terms of applying their research to two sets of audio data. First prepared by them, and second is available on the Internet and for free. Then the study compared the results to create a kind of challenges to link the mechanism of hiding or fading data with DL algorithms. As well as, better production of a research path specialized in hiding data used the latest and best deep and ML algorithms. A wide variety of data masking methods have been proposed over the past years, and currently CNN widely used in data masking and protection, and it is unique in that it extracts features automatically. Research in this field is still very active, which relies on DL algorithms, with continuous updating of previous studies.

Acknowledgments

We extend my thanks and appreciation to the University of Diyala, which is the supporter of discreet scientific research.

References

1. Shumeet Baluja, ‘Hiding images in plain sight: Deep steganography,’ in Advances in Neural Information Processing Systems,2017, pp. 2069–2079.” .
2. Jiren Zhu, Russell Kaplan, Justin Johnson, and Li Fei-Fei, ‘Hidden: Hiding data with deep networks,’ in European Conference on Computer Vision. Springer, 2018, pp. 682–697.
3. Kreuk, F., Adi, Y., Raj, B., Singh, R., & Keshet, J. (2019). Hide and speak: Towards deep neural networks for speech steganography. arXiv preprint arXiv:1902.03083.
4. Zhang, C., Lin, C., Benz, P., Chen, K., Zhang, W., & Kweon, I. S. (2021). A brief survey on deep learning based data hiding. arXiv: 2103.01607.
5. M. Geleta, C. Puntí, K. McGuinness, J. Pons, C. Canton, and X. Giro-I-Nieto, “Pixinwav: Residual Steganography for Hiding Pixels in Audio,” ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, vol. 2022-May. pp. 2485–2489, 2022, doi: 10.1109/ICASSP43922.2022.9746191.
6. Catherine Paulin, Sid-Ahmed Selouani, and Eric Hervet. 2016. Audio steganalysis using deep belief networks. International Journal of Speech Technology 19, 3 (2016), 585–591.
7. Mohammed Salem Atoum, Subariah Ibrahimn, Ghazali Sulong, Akram Zeki, and Adamu Abubakar. Exploring the challenges of mp3 audio steganography. In 2013 International Conference on Advanced Computer Science Applications and Technologies, pages 156–161. IEEE,.”
8. N. Cvejic. Algorithms for audio watermarking and steganography, 2004. Department of Electrical and Information Engineering, Information Processing Laboratory, University of Oulu. 2.
9. Rully Adrian Santosa and Paul Bao. Audio-to-image wavelet transform based audio steganography. In 47th International Symposium ELMAR, 2005., pages 209–212. IEEE, 2005. 2, 4.

10. Kadir Tekeli and Rifat Asliyan. A comparison of echo hiding methods. The Eurasia Proceedings of Science Technology Engineering and Mathematics, 1:397–403, 2017.
11. Felix Kreuk, Yossi Adi, Bhiksha Raj, Rita Singh, and Joseph Keshet. Hide and speak: Towards deep neural networks for speech steganography. Proc. Interspeech 2020, pages 4656– 4660, 2020. 1, 2, 3, 4, 6.
12. Dalal N Hmood, Khamael A Khudhiar, and Mohammad S Altaei. A new steganographic method for embedded image in audio file. International Journal of Computer Science and Security (IJCSS), 6(2):135–141, 2012. 2.
13. W. Zhang, "Dynamic Hand Gesture Recognition Based on 3D Convolutional Neural Network Models," 2019 IEEE 16th Int. Conf. Networking, Sens. Control, pp. 224–229, 2019.
14. Turki, A. I., & Hasson, S. T. (2023). Study Estimating hourly traffic flow using Artificial Neural Network: A M25 motorway case. Samarra Journal of Pure and Applied Science, 5(1), 47-59.
15. M. Mustafa, "A study on Arabic sign language recognition for differently abled using advanced machine learning classifiers," Journal of Ambient Intelligence and Humanized Computing. 2020, doi: 10.1007/s12652-020-01790-w.
16. H. I. Mohammed, J. Waleed, and S. Albawi, 'An Inclusive Survey of Machine Learning based Hand Gestures Recognition Systems in Recent Applications,' IOP Conf. Ser. Mater. Sci. Eng., vol. 1076, no. 1, p. 012047, 2021.
17. S. Albawi, T. A. M. Mohammed, and S. Alzawi, "Understanding of a Convolutional Neural Network," Ieee. 2017, [Online]. Available: https://wiki.tum.de/display/lfdv/Layers+of+a+Convolutional+Neural+Network.
18. Mohammed, Hind Ibrahim, and Jumana Waleed. 'Hand gesture recognition using a convolutional neural network for arabic sign language.' AIP Conference Proceedings. Vol. 2475. No. 1. AIP Publishing LLC, 2023.
19. G. Li et al., "Hand gesture recognition based on convolution neural network," Cluster Comput., 2017, doi: 10.1007/s10586-017-1435-x.
20. A. K. and A. V. Kumud Alok, Anurag Mehra, "Hand Sign Recognition using Convolutional Neural Network," Int. Res. J. Eng. Technol., vol. 07, no. 01, pp. 1680–1682, [Online]. Available: https://www.irjet.net.
21. Merzah, B. M. (2021). Actual Needs Criteria for Assessing Data Classification Platforms. Samarra Journal of Pure and Applied Science, 3(1), 125-138." .
22. Paulin, Catherine, Sid-Ahmed Selouani, and Éric Hervet. 'A comparative study of audio/speech steganalysis techniques.' 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE). IEEE, 2017.
23. A. Janicki, W. Mazurczyk, and K. Szczypiorski, 'Steganalysis of transcoding steganography,' annals of telecommunications-annales des tel´ ecommunications ´ , vol. 69, no. 7-8, pp. 449–460, 2014.
24. Kocal, O. H., YÜRÜKLÜ, E., & DİLAVEROĞLU, E. (2016). Speech steganalysis based on the delay vector variance method. Turkish Journal of Electrical Engineering and Computer Sciences, 24(5), 4129-4141.
25. H. Ghasemzadeh and M. K. Arjmandi, 'Reversed-mel cepstrum based audio steganalysis,' in Computer and Knowledge Engineering (ICCKE), 2014 4th International eConference on. IEEE, 2014, pp. 679–684.
26. Paulin, C., Selouani, S. A., & Hervet, É. (2017, April). A comparative study of audio/speech steganalysis techniques. In 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE) (pp. 1-4). IEEE.
27. Ghasemzadeh, Hamzeh, Mehdi Tajik Khass, and Meisam Khalil Arjmandi. 'Audio steganalysis based on reversed psychoacoustic model of human hearing.' Digital signal processing 51 (2016): 133-141.
28. Li, S., Jia, Y., & Kuo, C. C. J. (2017). Steganalysis of QIM steganography in low-bit-rate speech signals. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25(5), 1011-1022.
29. Bolin Chen, Weiqi Luo, and Haodong Li. 2017. Audio steganalysis with convolutional neural network. In Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security. ACM, 85–90.

30. Kun Yang, Xiaowei Yi, Xianfeng Zhao, and Linna Zhou. 2017. Adaptive MP3 Steganography Using Equal Length Entropy Codes Substitution. In International Workshop on Digital Watermarking. Springer, 202–216.
31. Zinan Lin, Yongfeng Huang, and JilongWang. 2018. RNN-SM: Fast Steganalysis of VoIP Streams Using Recurrent Neural Network. IEEE Transactions on Information Forensics and Security 13, 7 (2018), 1854–1868.
32. Das, Abhishek, Japsimar Singh Wahi, Mansi Anand, and Yugant Rana. 'Multi-image steganography using deep neural networks.' arXiv preprint arXiv:2101.00350 (2021).
33. Bhangale, K. B., & Mohanaprasad, K. (2021). A review on speech processing using machine learning paradigm. International Journal of Speech Technology, 24, 367-388.