Robustness of Electrical Network Frequency Signals as a Fingerprint for Digital Media Authentication

Nihal Poredi[a], Deeraj Nagothu[a], Yu Chen[a], Xiaohua Li[a], Alexander Aved[b], Erika Ardiles-Cruz[b], Erik Blasch[b]
[a]Dept. of Electrical & Computer Engineering, Binghamton University, SUNY, Binghamton, NY 13902, USA
[b]The U.S. Air Force Research Laboratory, Rome, NY 13441, USA
{nporedi1, dnagoth1, ychen, xli}@binghamton.edu, {alexander.aved, erika.ardiles-cruz, erik.blasch}@us.af.mil

***Abstract*—Leveraging modern Artificial Intelligence (AI) technology, Deepfake attacks manipulate audio/video streams (AVS) to mimic any targeted person or scenario. Deepfake attacks are highly disturbing, and the misinformation can mislead the public, raising further challenges in policy, technical, social, and legal aspects. Electrical Network Frequency (ENF) signals embedded in AVS data are promising to be utilized as fingerprints to authenticate digital media and timely detect deepfaked audio or video. Meanwhile, the success of ENF-based deepfake detection approaches will be forfeited if attackers can create false ENF fingerprints to fool the detector. In this paper, a thorough experimental study validates the robustness of ENF signals as a fingerprint for digital media authentication. Taking statistical, supervised learning, and deep learning approaches, this work shows that it is infeasible to forecast the future ENF signals based on historical records. While strict theoretical proof is yet to be done, this work experimentally verifies ENF signals as a reliable fingerprint to authenticate digital media.**

***Index Terms*—Electrical Network Frequency (ENF) Signals, Embedded Fingerprints, Digital Media Authentication.**

I. Introduction

Internet of Video Things (IoVT) is increasingly being deployed, and is playing critical roles, in 5G and beyond communication enabled applications [7], such as, public safety surveillance in smart cities [22]. There is a growing trend where more and more meetings and activities are being migrated from the physical world to cyberspace, and people have become more willing to communicate with each other over video calls. Specifically, the video conferencing market is projected to reach $19.73 billion by 2030 [21]. Naturally, an increasing amount of audio/video streams (AVS) data is transmitted over the Internet, and it is compelling to secure the data against malicious actors. One of the most important pillars of data security is data authentication. It is imperative that the data being transmitted is not tampered with in transit against media attacks like DeepFakes, which alter the perception of captured media [27], and spread disinformation that hurts the trust foundation of media in our society [23].

With more IoVT devices being deployed at the edge every day, it is ideal that security measures are available on-site to protect the integrity of input data. Visual layer attacks, including frame forgery attacks, replay attacks and DeepFake attacks; have a common basis, i.e. modifying the spatial or temporal nature of the pixels. As a result, visual layer attacks leave behind exploitable fingerprints that make detection possible. Similarly, audio tampering results in modification of underlying frequencies and signal phase continuity, which can be exploited to identify any forgeries. An environmental fingerprint-based detection technique is highly demanded to identify multimedia forgeries. Therefore, it is essential to have a reliable and unique fingerprint that varies in both spatial and temporal domains without relying on device operability.

Electrical Network Frequency (ENF) is the power supply frequency available through the electrical grid, with a nominal frequency value of 60 Hz in the United States and 50 Hz in most of the European and Asian countries [10]. The instantaneous fluctuations in the power supply frequency are caused by the supply and demand variations of continuous power generation, and result in the creation of unique frequency fluctuations, of which the instantaneous signal is referred to as an *ENF signal*. In the United States, the fluctuations are typically in the range ± 0.02 Hz, while for other locations, they depend on the varying loads on the generators.

For multimedia authentication systems, ENF has proven to be effective in detecting anomalies in both, the spatial and temporal domains. It has been found that ENF signals are embedded in multimedia recordings through audio recordings and surveillance recorders that may be directly connected to the power grid. Thus, it is noted that the multimedia ENF signatures result from interference of electromagnetic fields created by the power sources [11]. In addition, battery-operated devices can also potentially capture the ENF through the background hum generated by nearby power sources [6]. One peculiar and essential feature of the ENF signal is that the signal fluctuations are similar throughout the power grid interconnect. Therefore, an ENF signal estimated through a multimedia recorder has the highest correlation with the frequency fluctuations gathered from the connected power grid in the same temporal range.

ENF is used in forensics to authenticate the media recordings presented to law enforcement and surveillance media authorities [11], [16], [17]. Audio and video synchronization in multimedia projects can be obtained using ENF signal similarity in corresponding media [26]. Spatial signatures, like the region of recording identification, are carried out by comparing the embedded ENF signal traces, with reference databases from power grid interconnects collected using frequency distribution networks [14]. Strong ENF fluctuations can be exploited to identify the exact geographical location of a recording within a power grid interconnect using the correlation coefficient. The correlation decreases as the spatio-temporal distance increases from the region of recording [9].

Due to the vast number of applications of ENF for multimedia authentication, malicious actors are often interested in creating false ENF fluctuations, and rendering the ENF-based

978-1-6654-7189-3/22/$31.00 © 2022 IEEE

authentication system useless. Although existing work states that ENF is a continuous and random process signal, to the best of our knowledge, there is not a comprehensive study that has verified or proven the robustness of ENF signals as a fingerprint for multimedia authentication. In this paper, a thorough experimental study is conducted, using statistical, supervised learning, and deep learning approaches. Our experimental results show that it is infeasible to forecast the fluctuation pattern of ENF signals based on historical records. Although it is not a strict theoretical proof, this work verifies that the ENF signal is a reliable fingerprint in practice, to authenticate digital media.

The rest of this paper is structured as follows. Section II introduces the background knowledge. Section III presents the theory behind the forecasting methods used. The experimental results are presented in Section V and further discussed in Section VI. Finally, Section VII concludes this paper.

II. Background and Related Work

ENF is an emerging forensic fingerprint with robust potential against multimedia manipulations due to its spatio-temporal consistency. Although many factors determine an efficient estimation of ENF, there are always concerns regarding an adversary trying to bypass, or manipulate the ENF authentication scheme. The ENF signal features randomness and uniqueness throughout the power grid, making it difficult to manipulate it in real-time. Although the offline recordings are susceptible to a filtering mechanism, and negate the signal presence, there are authentication techniques which are based on real-time media capture and authenticity verification using ENF [16]. The presence of the ENF signal, captured in media recordings, along with systems developed for media authentications, are discussed in the following sections.

A. ENF in Multimedia Recordings

For indoor infrastructure IoVT devices like the smart surveillance system, the sensors are directly powered by the grid. ENF is embedded in both audio and video recordings through different media. In audio recordings, the signal is embedded through electromagnetic interference, or through the background hum [6]. Figure 1 represents the spectrogram of the power recordings. Along with the nominal frequency, the presence of ENF is also detected in higher frequency harmonics as a multiple of nominal frequency with similar fluctuations. In the case of video recordings, ENF is embedded through the illumination frequency projected by light sources running on the power grid. The camera sensors capture the photons depending on the type of sensor used, and the shutter mechanism deployed. The two most commonly manufactured imaging sensors are complementary metal-oxide-semiconductor (CMOS) and charge-coupled device (CCD) sensors.

ENF signals from video recordings also depend on the video frame rate. CCD sensors use the global shutter mechanism, where the number of samples collected does not satisfy the Nyquist Criterion, and thereby rely on the aliasing frequency [10]. However, because of the lower cost, most IoVT devices adopt CMOS sensors, which deploys a rolling shutter mechanism. The resulting samples present higher fidelity, from which a reliable ENF estimation can be carried out [25].

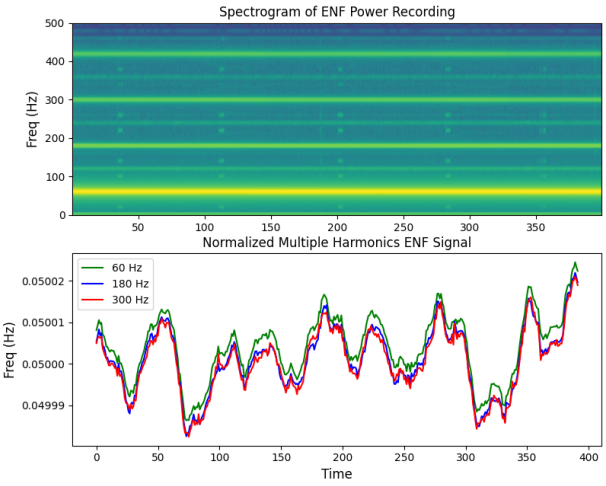

Fig. 1. Spectrogram of Power recording with its respective ENF estimates in multiple harmonic frequencies.

B. ENF-based Authentication System

Equipped with ENF estimation as a reliable authentication mechanism against multimedia manipulation attacks, real-time edge-based IoVT devices can carry out secure authentication of their content delivery. Frame duplication attacks and DeepFake attacks in both audio and video channels, are reliably detected using the ENF-based authentication on real-time applications [20]. Using the unique ENF fluctuations, the mismatch in the ENF from video frames and audio samples clearly indicate the attack localization and detection [20].

Online conferencing software algorithms are the most susceptible to DeepFake attacks, without any means of real-time authentication. Software like *Descript*, a text-to-speech generating software pre-trained on the target's voice, can create audio deepfake and mimic a targeted person in real-time [1]. Combined with a real-time face forgery application like the *DeepFaceLive* [24], a fake target can be generated and manipulated by the perpetrator. However, these deepfakes interrupt the underlying frequency pattern, and the disruption in the signal continuity in both spatial and temporal domain can be detected using ENF. Figure 2 illustrates the frequency noise added to media recordings by a Deepfake attack. The affected frequency bins also modify ENF fluctuations and result in an inconsistent ENF signal. For a detailed analysis and integration of ENF-based authentication system against Deepfake attacks, interested readers are referred to our previous publications for further reading [18], [19], [20].

In this work, we aim to verify the robustness of ENF signals as a fingerprint for digital media authentication. Using three widely adopted time-series forecasting technologies, namely classical statistical analysis, supervised learning, and deep learning, we investigate how (in)accurately, an attacker can predict the near future ENF fluctuations based on historical records.

III. Forecasting Methods Overview

The widely adopted time-series forecasting methods can be roughly classified into three categories, Statistical, Supervised

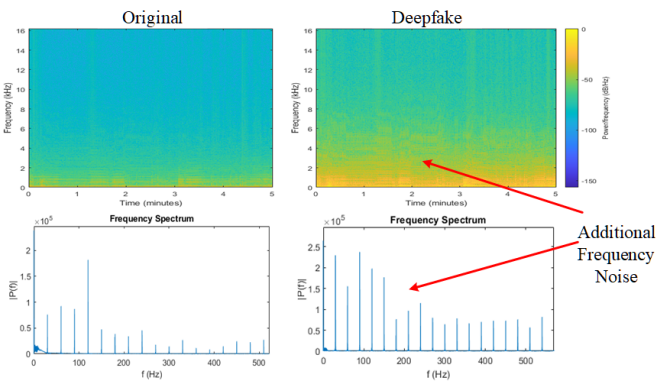

Fig. 2. Spatial Pixel Intensities and Frequency Spectrum affected by Deepfake forgery.

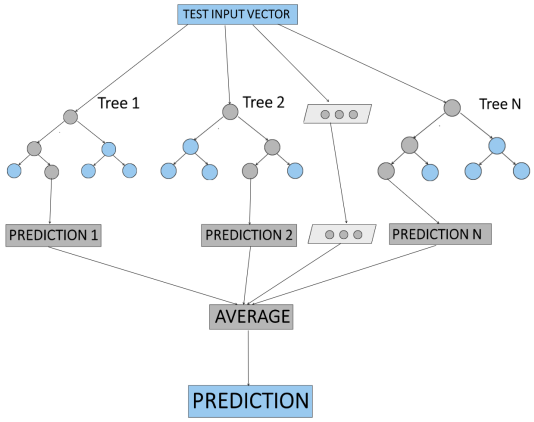

Fig. 3. Supervised Learning using the Random Forest algorithm.

Machine Learning, Unsupervised Deep Learning. The characteristics of each method are discussed in this section.

A. Classical Methods

These are statistical methods that exploit the univariate nature of ENF fluctuations within a time frame. The classical methods rely on temporal variations of ENF signals and use techniques such as moving averages and smoothing to predict future values. This paper examines two methods in particular, the Auto Regressive Integrated Moving Average (ARIMA) model and Exponential Smoothing.

1) ARIMA: The ARIMA model integrates the concepts of Auto Regression and Moving Average [4]. Auto Regression takes the future values as a function of one or more past values, while the Moving Average considers the average of all the values until the point of prediction. "Integrated" implies that the values are predicted by considering the difference of the adjacent values to ensure stationarity. However, as the ENF signal is a time series that is largely stationary around a fixed mean of 60 Hz, there is not a need to differentiate the values to account for hidden parameters that could make it non-stationary. Therefore, the model is essentially reduced to an ARMA model as per the equation below.

$$Y_t = \mu + \Sigma_{k=1}^{p} \phi_k (Y_{t-k} - \mu) + \epsilon_t + \Sigma_{i=1}^{q} \theta_i \epsilon_{t-i}$$

where p is the lag order, q is the order of the moving average, μ is the process mean, ϵ_t is a white noise process with mean 0, and ϕ and θ describe the weights of the past values and error terms respectively . The lag order p denotes the number of past values that were used to predict the current data point. The order of the moving average q represents the size of the moving average window. The resulting predictions were analyzed for accuracy on the basis of their Root Mean Square Error (RMSE) and correlation coefficient.

2) Exponential Smoothing: Exponential Smoothing makes an assumption that the current prediction is a weighted sum of past time series values [8], which in this case is a series of ENF values over time. The weights are such that they decrease in an exponential manner as the distance from the current prediction increases. The weights for the current prediction are selected such that the total RMSE of the past predictions is minimized according to the equation below.

$$\hat{Y_t} = \theta Y_{t-1} + (1-\theta)\hat{Y}_{t-1},$$
$$0 < \theta < 1$$

where $\hat{Y_t}$ is the current prediction, $\hat{Y}_{t-1}$ is the previous prediction, Y_{t-1} is the true previous value, and θ is the smoothing factor.

B. Supervised Learning-based Methods

Supervised learning is a process of training a model on a dataset, such that, it learns to map the input to the output according to the algorithm used. In this work, the training dataset is a collection of a fixed number of ENF values in time, and an output vector, which is the immediate next ENF value. The mapping from input to output is accomplished depending on the algorithm used. In this paper, we analyse ENF as a supervised learning problem using the algorithms of Linear Regression and Random Forest Regression.

1) Linear Regression: Linear Regression describes the relationship between two quantities by obtaining the optimal linear equation to minimize the least square error between them [2]. The representation of the relationship is through the linear equation that encapsulates a certain number of input values, and the solution for which is the prediction of the instance following the input set. For a single input x, the predicted value y is described by,

$$y = B_0 + B_1 x$$

where B_0 and B_1 are the weights or coefficients learnt by the model during training.

In this context, y and x were the predicted and the observed ENF values respectively, while the resulting fitted linear relationship, described the coefficients. The input set consisted of the last 12 ENF values, which were used to predict the following ENF value.

2) Random Forest Regression: Random Forest Regression (RFR) works by running several independent decision trees (multiple ML algorithms) in parallel to get a list of predictions [5], as shown in Fig. 3. The final prediction is then taken as the average of the list. The RFR technique is also referred to as *ensemble learning*. The advantage of averaging is realized in the estimation accuracy and the model's ability to avoid overfitting.

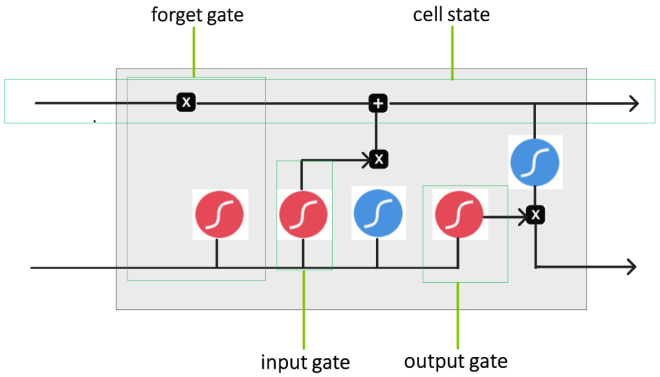

Fig. 4. LSTM cell architecture.

The predictions were ENF values obtained by considering the past 12 seconds of data.

C. Deep Learning using Long Short Term Memory Networks

Long Short Term Memory Networks (LSTMs) are a special type of Recurrent Neural Network (RNN), which store past values and use them to predict future values [13]. They have a chain-like structure with memory that enables them to learn long term dependencies [28]. As such, LSTM networks are ideal for sequence-based predictions and are used in many applications, from natural language processing [3] to radar [15]. An LSTM network is essentially a chain of several units having the same architecture. As illustrated in Fig 4, a typical LSTM unit includes an input gate, an output gate, and a forget gate, which regulate the flow of information in and out of the cell. The gates are built using *sigmoid* and *tanh* functions that perform various memory and decision making tasks. Here, the ENF sequence is viewed as a time series with several parameters, either known or unknown, affecting its patterns and seasonality. Therefore, it can be reduced to a time series prediction problem and could be solved using an LSTM model [29].

In this paper, two LSTM models were used to make two different types of ENF predictions. Both models comprised of an LSTM layer with 200 neurons, followed by a dense layer. However, they differed on the data that they were trained on as depicted in Fig. 5. The first model was trained on 2.5 hours of continuous ENF, recorded every second. This was done to understand if the model could identify short term fluctuations of the ENF signal. It was then tested on ENF data spanning five minutes. Thus, the predictions were over a forecast horizon of 300 seconds. The second model was trained on averaged ENF data. The data was produced by averaging ENF values every 30 minutes for 19 days. Thus, every day had 48 data points, and the entire dataset had 912 data points. This was done

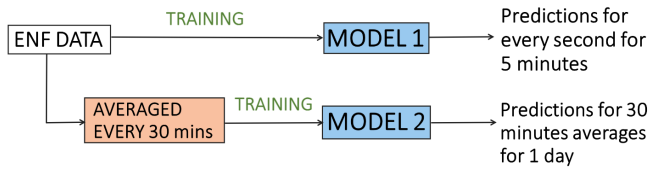

Fig. 5. LSTM training flow.

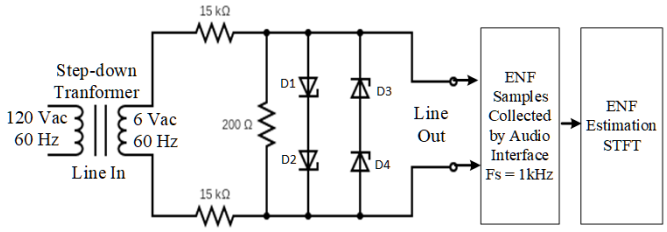

Fig. 6. ENF collection stages from power grid to frequency samples.

to understand if the model was able to identify the long term pattern of ENF fluctuations. The predictions were 48 average ENF values predicted for the 20-th day.

IV. Dataset Description

The ENF dataset used in this work is a continuous ENF series from the Eastern power grid interconnect, collected every second. The dataset spanned over several days and was used for both training the models and testing them for their prediction accuracy. The ENF data was collected using a step-down transformer and a voltage divider circuit at a sampling rate of 1000. The frequency samples were recorded as sound input and stored as audio recording. Figure 6 shows the ENF sample collection circuit used for our testbed [12], [17]. We used a non-parametric spectrum estimation technique called Short Time Fourier Transform (STFT) for a reliable estimation of ENF from its nominal frequency [10]. The collected power recordings have ENF embedded in multiple harmonic frequencies, with varying Signal-to-Noise (SNR) ratios. For ENF estimation, the frequency with the highest SNR is used as the nominal frequency, to capture accurate ENF fluctuations [20].

V. Experimental Results

ENF values were predicted using different forecasting methods described earlier, and the accuracy of each method was analyzed, on the basis of its Root Mean Squared Error (RMSE) value and Correlation Coefficient curve. The correlation value varies in the range $[-1, 1]$, where 1 represents highest similarity.

The ARIMA model was trained on one hour of training data, and was then evaluated on a testing dataset that was

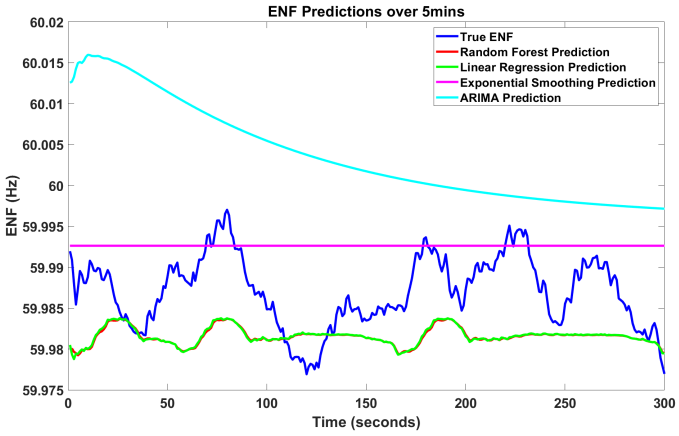

Fig. 7. ENF Predictions using Classical methods and Supervised Learning.

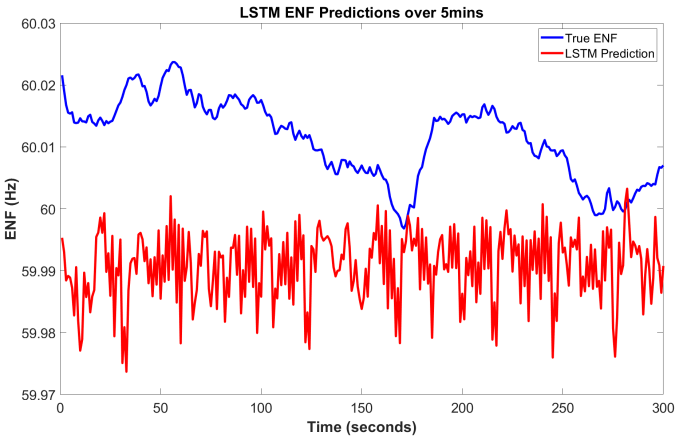

Fig. 8. ENF Prediction using LSTM.

a continuation of the training dataset. Up to five minutes of ENF was predicted and compared to the actual values. The lag order was chosen as 20, meaning that the last 20 ENF values were considered while making a prediction. This is reasonable considering the forecast horizon is five minutes. The order of moving average was taken as one. Fig. 7 shows that while the first few seconds of the prediction curve is related to the actual values, the predictions that go further into the horizon, converge to a mean value without exhibiting any fluctuations. Since the the ENF time series is stationary with a fixed mean, then the signal prediction would be stable.

The Exponential Smoothing model was also trained on one hour of training data and evaluated on a testing dataset of five minutes. The Holt Winter algorithm was used with a smoothing factor of 0.2. As shown in Fig. 7, the Exponential Smoothing model was also unsuccessful in identifying any short term fluctuation patterns in the ENF data. This confirms the fact that ENF fluctuations are truly random around a fixed mean of 60 Hz, and do not follow any easily predictable trend which would have otherwise been identified by the model.

The supervised learning (SL) models both showed similar results for predicting ENF values. They were trained on 3 hours worth of ENF data, and were used to predict ENF over a forecast horizon of 5 minutes. Although the SL models appear to be able to identify the fluctuations and short term trends better than ARIMA and Exponential Smoothing methods over the entirety of the forecast horizon, they fall short of performing the prediction with a strong enough correlation to the actual ENF curve as shown in Fig. 7.

Meanwhile, the LSTM model was observed to be relatively better at capturing the short term fluctuations in ENF data, although it was with a higher degree of variance than the original ENF as can be seen in Fig. 8. Moreover, the error is large enough for these predictions to be rendered ineffective in launching any type of forgery attack. The LSTM model results show some promise in terms of predicting ENF average trends over larger windows. While the averages are not predicted with fine accuracy, the long term trend is identified by the model. Figure 9 confirms that the average ENF predictions fall below the 0.8 correlation coefficient threshold, which was used to tell whether a video stream is faked [18]. This can be attributed to the absence of well defined short term patterns and seasonality

TABLE I
RMSE FOR DIFFERENT FORECASTING METHODS OVER VARIOUS TIME FRAMES.

Time-frame	30 s	1 min	2 min	5 min
RMSE-ARIMA	**0.0024**	**0.0035**	**0.005**	**0.006**
RMSE-ES	0.005	0.006	0.007	0.007
RMSE-RF	0.0070	0.0066	0.0072	0.0069
RMSE-LR	0.007	0.0066	0.0071	0.0069
RMSE-LSTM	0.026	0.028	0.027	0.023

in ENF data. While the patterns do exist, they appear to be psedo-random, and may be identified only after training over a larger dataset that could span over months.

VI. DISCUSSIONS

The different forecasting methods discussed in this paper were compared on the basis of their RMSE values and their correlation curves with respect to the actual ENF values.

Table I shows the various RMSE values observed over four different time frames, 30 seconds, one minute, two minutes and five minutes. It is clear that the RMSE value is the lowest for the ARIMA model, while it is the highest for the LSTM model. However, it is interesting to note that as the forecasting horizon increases, the RMSE increases for the statistical methods, while it actually decreases for the deep learning method. Thus, it is fair to deduce that while statistical methods or supervised learning techniques may be marginally accurate in the short term, it is the deep learning techniques that provide better promise in terms of consistency in the long term. Table I also confirms the fact that none of the forecasting methods used for time series prediction today, can predict ENF with a reasonable accuracy, owing to the relatively high RMSE values. In a successful ENF forgery scenario, the RMSE demands would be at least ten times lower than the lowest value obtained in these experiments.

Figures 9(a), (b), and (c) depict the correlation curves of three main methods with respect to the actual ENF curves. The Exponential Smoothing model was excluded in this analysis as the results were sub optimal, while the Linear Regression results were excluded due to their similar nature to the Random Forest results. The curves prove the fact that none of the methods

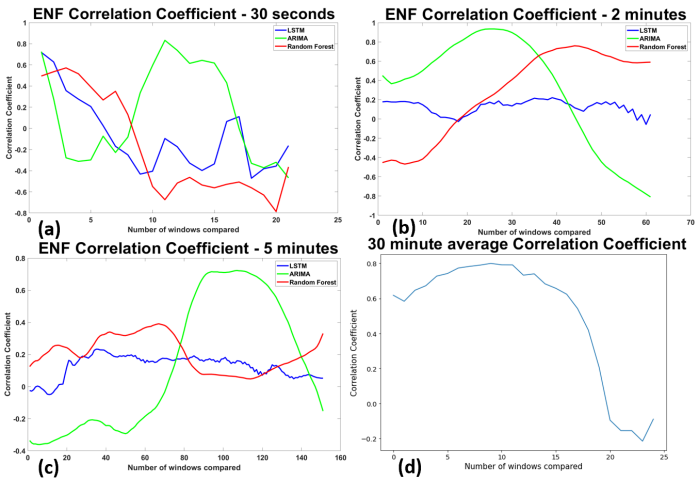

Fig. 9. Correlation curves of different methods over various time frames.

succeed in establishing a high enough correlation to perform ENF forgery. While ARIMA and Random Forest models do reach correlation values over the threshold, 0.8, they do so for a few seconds at a time and fail to exhibit any consistency over a longer period of time, and often dropping to negative correlation values. The LSTM model shows certain degree of consistency for longer forecasting horizons, but the correlation values hardly cross 0.3, which is very low for any realistic implementation in forgery. Figure 9(d) shows that the LSTM method can achieve a higher correlation coefficient while predicting average ENF values. However, the highest value is close to 0.6, which is a long way from being indicative of successful forgery of average values, predicting which moreover, are not a direct way of forging ENF.

VII. Conclusions

In this paper, a thorough experimental study shows that it is infeasible to predict the future ENF signals based on historical records, using standard AI methods. The results obtained by attempting to predict ENF fluctuations, indicate that none of the time series forecasting methods in use today, can successfully predict future ENF based on historical ENF data. While the classical and supervised learning methods converge to a constant value, the deep learning methods do not produce the level of correlation required for successful prediction. Moreover, the training time for the models, spans from several minutes to several hours, rendering them unsuitable for quick short-term predictions. Since these methods are the most popular in predicting time series, it is fair to conclude that ENF as a time series cannot be predicted due to its random nature. Hence, the hypothesis is disproved, resulting in the robustness of ENF as a fingerprint for digital media authentication. Being aware of the limitations of an experimental study, our ongoing work is focused on strict theoretical proof using signal processing theory and information theory.

Acknowledgement

This work is supported by the U.S. National Science Foundation (NSF) via grant CNS-2039342 and in part by the U.S. Air Force Office of Scientific Research (AFOSR) Summer Faculty Fellowship Program (SFFP) via contracts FA8750-15-3-6003, FA9550-15-001 and FA9550-20-F-0005. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the U. S. Air Force.

References

[1] "Descript | Create podcasts, videos, and transcripts." [Online]. Available: https://www.descript.com/

[2] C. G. Atkeson, A. W. Moore, and S. Schaal, "Locally weighted learning," *Lazy learning*, pp. 11–73, 1997.

[3] S. R. Bowman, G. Angeli, C. Potts, and C. D. Manning, "A large annotated corpus for learning natural language inference," *arXiv preprint arXiv:1508.05326*, 2015.

[4] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[5] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[6] J. Chai, F. Liu, Z. Yuan, R. W. Conners, and Y. Liu, "Source of enf in battery-powered digital recordings," in *Audio Engineering Society Convention 135*. Audio Engineering Society, 2013.

[7] C. W. Chen, "Internet of video things: Next-generation iot with visual sensors," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 6676–6685, 2020.

[8] E. S. Gardner Jr, "Exponential smoothing: The state of the art," *Journal of forecasting*, vol. 4, no. 1, pp. 1–28, 1985.

[9] R. Garg, A. Hajj-Ahmad, and M. Wu, "Feasibility study on intra-grid location estimation using power enf signals," *arXiv preprint arXiv:2105.00668*, 2021.

[10] R. Garg, A. L. Varna, A. Hajj-Ahmad, and M. Wu, ""seeing" enf: power-signature-based timestamp for digital multimedia via optical sensing and signal processing," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 9, pp. 1417–1432, 2013.

[11] C. Grigoras, "Applications of enf criterion in forensic audio, video, computer and telecommunication analysis," *Forensic science international*, vol. 167, no. 2-3, pp. 136–145, 2007.

[12] C. Grigoras, J. Smith, and C. Jenkins, "Advances in enf database configuration for forensic authentication of digital media," in *Audio Engineering Society Convention 131*. Audio Engineering Society, 2011.

[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[14] Y. Liu, S. You, W. Yao, Y. Cui, L. Wu, D. Zhou, J. Zhao, H. Liu, and Y. Liu, "A distribution level wide area monitoring system for the electric power grid–fnet/grideye," *IEEE Access*, vol. 5, pp. 2329–2338, 2017.

[15] U. K. Majumder, E. P. Blasch, and D. A. Garren, *Deep Learning for Radar and Communications Automatic Target Recognition*. Artech House, 2020.

[16] D. Nagothu, Y. Chen, A. Aved, and E. Blasch, "Authenticating video feeds using electric network frequency estimation at the edge," *EAI Endorsed Transactions on Security and Safety*, vol. 7, no. 24, p. e4, 2021.

[17] D. Nagothu, Y. Chen, E. Blasch, A. Aved, and S. Zhu, "Detecting malicious false frame injection attacks on surveillance systems at the edge using electrical network frequency signals," *Sensors*, vol. 19, no. 11, p. 2424, 2019.

[18] D. Nagothu, R. Xu, Y. Chen, E. Blasch, and A. Aved, "Defake: Decentralized enf-consensus based deepfake detection in video conferencing," in *Proceedings of the IEEE 23rd International Workshop on Multimedia Signal Processing, Tampere, Finland*, 2021, pp. 6–8.

[19] ——, "Detecting compromised edge smart cameras using lightweight environmental fingerprint consensus," in *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, 2021, pp. 505–510.

[20] ——, "Deterring deepfake attacks with an electrical network frequency fingerprints approach," *Future Internet*, vol. 14, no. 5, p. 125, 2022.

[21] P. Newswire, "Video conferencing market to be worth $19.73 billion by 2030," *https://www.bloomberg.com/press-releases/2022-06-20/video-conferencing-market-to-be-worth-19-73-billion-by-2030-grand-/view-research-inc*, June 20 2022.

[22] S. Y. Nikouei, Y. Chen, S. Song, B.-Y. Choi, and T. R. Faughnan, "Toward intelligent surveillance as an edge network service (isense) using lightweight detection and tracking algorithms," *IEEE Transactions on Services Computing*, 2019.

[23] K. A. Pantserev, "The malicious use of ai-based deepfake technology as the new threat to psychological security and political stability," in *Cyber Defence in the Age of AI, Smart Societies and Augmented Humanity*. Springer, 2020, pp. 37–55.

[24] I. Perov, D. Gao, N. Chervoniy, K. Liu, S. Marangonda, C. Umé, M. Dpfks, C. S. Facenheim, L. RP, J. Jiang *et al.*, "Deepfacelab: A simple, flexible and extensible face swapping framework," *arXiv preprint arXiv:2005.05535*, 2020.

[25] H. Su, A. Hajj-Ahmad, R. Garg, and M. Wu, "Exploiting rolling shutter for enf signal extraction from video," in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 5367–5371.

[26] H. Su, A. Hajj-Ahmad, C.-W. Wong, R. Garg, and M. Wu, "Enf signal induced by power grid: A new modality for video synchronization," in *Proceedings of the 2Nd ACM International Workshop on Immersive Media Experiences*, 2014, pp. 13–18.

[27] C. Vaccari and A. Chadwick, "Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news," *Social Media+ Society*, vol. 6, no. 1, p. 2056305120903408, 2020.

[28] R. Wen, K. Torkkola, B. Narayanaswamy, and D. Madeka, "A multi-horizon quantile recurrent forecaster," *arXiv preprint arXiv:1711.11053*, 2017.

[29] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 12, 2021, pp. 11 106–11 115.