

InterGridNet: An Electric Network Frequency Approach for Audio Source Location Classification Using Convolutional Neural Networks

Christos Korgialas*, Ioannis Tsingalis*, Georgios Tzolopoulos*, and Constantine Kotropoulos

Department of Informatics
Aristotle University of Thessaloniki
 Thessaloniki 54124, Greece

email: {ckorgial, tsingalis, gtzolopo, costas}@csd.auth.gr

* Equal contribution

Abstract—A novel framework, called InterGridNet, is introduced, leveraging a shallow RawNet model for geolocation classification of Electric Network Frequency (ENF) signatures in the SP Cup 2016 dataset. During data preparation, recordings are sorted into audio and power groups based on inherent characteristics, further divided into 50 Hz and 60 Hz groups via spectrogram analysis. Residual blocks within the classification model extract frame-level embeddings, aiding decision-making through softmax activation. The topology and the hyperparameters of the shallow RawNet are optimized using a Neural Architecture Search. The overall accuracy of InterGridNet in the test recordings is 92%, indicating its effectiveness against the state-of-the-art methods tested in the SP Cup 2016. These findings underscore InterGridNet’s effectiveness in accurately classifying audio recordings from diverse power grids, advancing state-of-the-art geolocation estimation methods.

Keywords—*electric network frequency (ENF); grid location estimation; audio processing; multimedia forensics*

I. INTRODUCTION

Due to power grid disturbances, the Electric Network Frequency (ENF) is a dynamic time series that exhibits fluctuations around its nominal frequency of 50 Hz in Europe and 60 Hz in the United States/Canada. These oscillations result from instantaneous load variations within the power grid, displaying a consistent pattern within interconnected grids. ENF signals become embedded in multimedia recordings captured in proximity to power sources. This distinctive signal can subsequently be extracted [1]–[4] from digital recordings for various applications, such as verification of recording timestamps [5]–[8].

Another application where ENF is also utilized is grid localization. Grid localization can be treated as inter-grid [9]–[11] or intra-grid [10] [12] [13] localization. Inter-grid localization focuses on identifying the grid in which a media recording was captured, while intra-grid localization aims to determine the recording’s location within the specific grid precisely. Inter-grid localization is briefly surveyed in Section II.

The intra-grid localization is considered more challenging due to the highly subtle distinctions in the ENF signatures encoded within the recorded signals. However, this assumption is challenged by [14], who detail noticeable differences due to varying city power consumptions and the time for load changes to impact different grid segments, a concept further

explored in [15]. Additionally, ENF fluctuations can stem from system disruptions like short circuits, power line switching, and generator failures, as noted in [16]. Minor local load changes affect ENF differently than major events like generator disconnections, which impact the entire grid at about 500 miles per second [17]. Given the aforementioned intra-grid characteristics, various methods have been proposed to tackle the problem of intra-grid localization [18] [19] [20].

A novel framework, termed InterGridNet, is introduced for geolocation classification exploiting the ENF. The framework offers a comprehensive approach that includes data preparation and preprocessing techniques using a shallow RawNet [21] for classification. The topology and the hyperparameters of InterGridNet are optimized through Neural Architecture Search (NAS), enhancing its capability to tackle inter-grid localization in audio recordings. It incorporates innovative techniques, including filtering to isolate the relevant ENF signal, using residual layers to extract frame-level embeddings, and employing a softmax activation function in the decision-making process. To our knowledge, this represents a pioneering advancement spanning from preprocessing techniques to the classification stage, establishing a novel framework in geolocation classification using deep learning methodologies. The Signal Processing (SP) Cup 2016 dataset [22], the only benchmark dataset publicly available, is employed for assessing geolocation classification.

The key contributions of the paper are as follows:

- A novel framework, coined InterGridNet, is proposed to treat geolocation estimation as a classification problem among nine power grids, employing a shallow RawNet optimized with NAS and leveraging ENF signatures from the benchmark SP Cup 2016 dataset. It should be noted that a shallow RawNet is utilized to reduce the number of parameters and achieve comparable performance with that using a deeper neural network.
- Experimental evaluation, including extensive testing of the SP Cup 2016 dataset, showcases the effectiveness of InterGridNet in geolocation classification across nine distinct power grids, where it is compared with state-of-the-art methods.

The remainder of the paper is organized as follows. Sec-

tion II provides an overview of related work. The proposed framework is detailed in Section III. Experimental evaluation is conducted in Section IV. Section V concludes the paper by providing information for future work.

II. RELATED WORK

ENF variations due to load fluctuations and grid frequency control help to localize audio recordings. Grigoras’s research demonstrated this by correlating ENF from audio recordings with reference ENF signals from different power grids to estimate the location of the recording [23]. Extensive research was conducted in grid localization using ENF by employing diverse datasets [12]. Additionally, location estimation at various scales was addressed in [24] and [13]. In [10], a machine learning system was developed to ascertain where an ENF-containing media file was recorded, even when no simultaneous ENF reference was available. Five machine learning algorithms were explored to identify the recording location of power and audio recordings obtained from ten distinct power grids in [25]. The hypothesis that variations in load conditions could generate unique location-specific patterns within the ENF signal was assessed in [14]. In [26], an ENF region classification model, UniTS-SinSpec, was introduced within the UniTS framework, integrating a sinusoidal activation function and a spectral attention mechanism and trained on a public dataset. Addressing the complexities of inter-grid classification, field specialists have formulated methodologies to distinguish audio recordings across global power grids, exemplified by the 2016 SP Cup. This work substantially improved the forensic analysis based on ENF, fortifying the verification of the authenticity of multimedia recordings. These distinctive patterns could pinpoint the precise location within a grid where the recording was made.

III. THE INTERGRIDNET FRAMEWORK

A. Dataset Preparation

The SP Cup 2016 competition [22] benchmark dataset [27] is employed, with data split into three sets: a training set for the model’s development and training, a practice set for validation, and a testing set for evaluating performance on unseen data (see Section IV-A). The dataset encompasses audio recordings capturing ENF signals from power grids across different countries. Specifically, it consists of recordings from nine distinct power grids, denoted as **A** through **I**. Grids **A**, **C**, and **I** are characterized by nominal ENF at 60Hz, while the remaining grids exhibit ENF around 50Hz.

The dataset consists of audio and power recordings for each grid. The power recordings were generated from a specialized circuit designed to capture the ENF time series directly from the power mains and have a temporal span of 30 to 60 minutes. The audio recordings were acquired using a microphone near substantial electrical devices, capturing the ENF hum for 30 minutes. In particular, power recordings are distinguished by stronger ENF traces than audio recordings.

The testing set has been augmented with 100 samples (40 Audio and 60 Power), each spanning 10 minutes. This

subset comprises 8-11 samples from each of the nine grids (**A** - **I**) and 10 additional samples from networks other than these, categorized as “None” (**N**). This diverse sample set is a benchmark for assessing the proposed InterGridNet’s efficiency and generalization.

Figure 1 illustrates the InterGridNet framework, highlighting the two critical stages of data preparation and classification. The data preparation process is depicted within the yellow dashed box in Figure 1. Initially, the recordings’ inherent characteristics, encompassing ENF signals at either 50Hz or 60Hz, are utilized to classify the recordings as audio or power recordings. Four distinct and independent data groups were created: `audio50`, `audio60`, `power50`, and `power60`.

During the training phase, this categorization is direct since the differences between audio and power recordings are perceptible, mainly due to the higher Signal-to-Noise Ratio (SNR) in power recordings. In contrast, during the testing phase, an automated grouping method is required to classify recordings based on their spectral characteristics, mainly to distinguish between the ENF frequencies of 50Hz and 60Hz. This method is described as follows:

- 1) For each recording, the average spectrogram magnitude is calculated for the first three harmonics associated with the nominal frequencies of 50 Hz and 60 Hz.
- 2) For each nominal ENF, the harmonic with the smallest value from step 1 is ignored. Since the ENF may not be present in every harmonic, the two harmonics with the stronger traces are enough for the categorization.
- 3) The average of the magnitude values at the retained frequencies in step 2 is calculated.
- 4) The largest value reveals the nominal frequency of the network.

After classifying each recording into its data group, a filtering process is applied to isolate the corresponding ENF within a range of 2 Hz. For instance, samples from the `audio60` group undergo filtering using a bandpass filter set to frequencies between 59 Hz and 61 Hz. Subsequently, the waveforms are segmented into 16-second frames with a 50% overlap and normalized to the interval $[-1, 1]$. These processed frames are subsequently fed into the classification model for power grid classification, shown as the blue dashed box in Figure 1. The aggregated count of frames for each grid is depicted in Figure 2, providing an overview of the distribution of frames across the dataset.

B. Classification Architecture

The spectral content of the frames exhibits variation based on the grid of origin, providing valuable information for the location estimation of the recording. Figure 3 displays a spectrogram concentrated around the nominal ENF from four distinct grids. Notably, the ENF behavior differs depending on the grid, wherein Figures 3(a), 3(b), 3(c), and 3(d) the frequency content is around 60Hz, 50Hz, 50Hz, and 60Hz, respectively. Consequently, the technique elucidated following harnesses these ENF characteristics to classify the samples according to the grid where the recording was made.

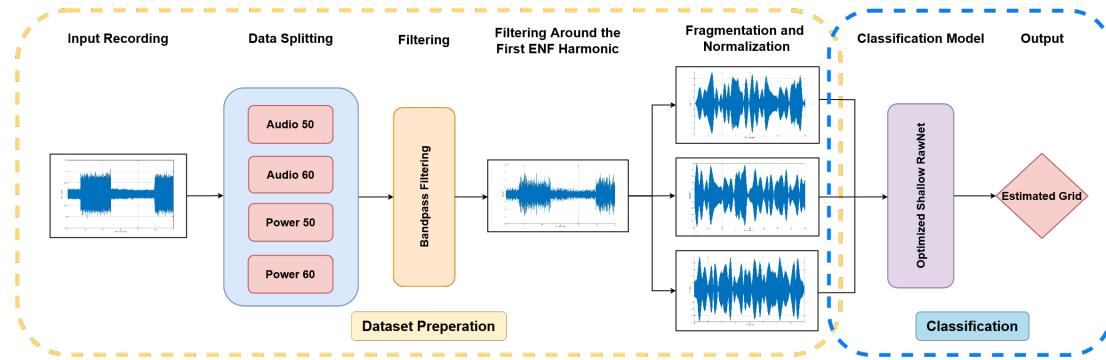


Figure 1. Flowchart of the proposed InterGridNet framework.

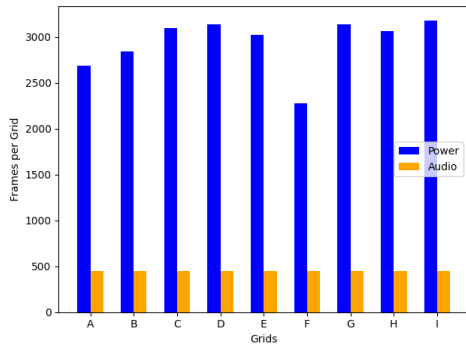


Figure 2. Number of audio and power recording frames in each grid.

TABLE I. OPTIMIZED HYPERPARAMETERS OF THE SHALLOW RAWNET.

| | G_{Audio}^{50} | G_{Audio}^{60} | G_{Power}^{50} | G_{Power}^{60} |
|---------------|-------------------------|-------------------------|-------------------------|-------------------------|
| Learning Rate | 6.5×10^{-4} | 7×10^{-4} | 1.1×10^{-3} | 9.7×10^{-4} |
| β_1 | 0.96 | 0.97 | 0.98 | 0.98 |
| β_2 | 0.998 | 0.998 | 0.992 | 0.993 |

To address the classification problem, individual classes are defined for each grid, comprising 16-second frames derived in Section III-A. These 16-second frames are called samples hereafter. The classification problem for each data group is denoted as $G_{\text{REC}}^{\text{ENF}}$, where REC represents the recording type (Audio or Power), and ENF signifies the nominal frequency of the grid. Consequently, the classification problems are denoted as G_{Audio}^{50} , G_{Power}^{50} , G_{Audio}^{60} , and G_{Power}^{60} . Each $G_{\text{REC}}^{\text{ENF}}$ is expressed as $G_{\text{REC}}^{\text{ENF}} = \{C_1, C_2, \dots, C_n\}$, where $n = 3$ for G_{Audio}^{60} and G_{Power}^{60} , and $n = 6$ for the others. Each C_i class in the classification problem contains all samples from the corresponding grid in the respective data group.

As an illustrative example, the classification problem for the data group $\text{audio}60$ is denoted as $G_{\text{Audio}}^{60} = \{C_1, C_2, C_3\}$, where C_1 encompasses the audio frames from grid **A**, C_2 from grid **C**, and C_3 from grid **I**. Each sample has a label $l \in \{1, 2, \dots, n\}$.

For the last part of the flowchart in Figure 1, a shallow RawNet architecture has been implemented to perform the classification. The topology of the shallow RawNet was opti-

mized through NAS using the `Keras-Tuner` library. During this search, several parameters were fine-tuned, including the number of convolutional layers (ranging between 3 and 5), the filter sizes (128 to 256), Gated Recurrent Unit (GRU) units (512 to 1024), and dense layer units (64 to 512). After extensive experimentation, the optimal configuration for this architecture was determined to include two residual blocks.

Specifically, as depicted in Figure 4, the network begins with an input layer that processes frames of size 15,999. These frames are passed through a Strided Convolution block consisting of a one-dimensional convolution layer, batch normalization (BN), and LeakyReLU activation (with a slope of 0.01 for negative inputs). This initial block outputs a feature map of size 5333×128 . The first residual block follows, comprising two convolutional layers, batch normalization, LeakyReLU activation, and a max-pooling layer, resulting in an output of 593×128 . Following a similar structure, another residual block with 256 filters is applied next, reducing the output to 66×256 . These residual blocks are crucial for extracting frame-level embeddings from the input data. Next, the network incorporates a GRU to aggregate these frame-level embeddings into a single ENF-level representation. The output from the GRU is then passed through a dense layer, reducing the dimensionality to a 128-dimensional vector. This layer condenses the extracted features into a more abstract, higher-level representation. Finally, the 128-dimensional vector is processed by the output layer, which uses a softmax activation function to map the vector to a probability distribution over the 9 classes, completing the classification task.

In addition to optimizing the topology of the shallow RawNet, NAS is also employed to fine-tune the hyperparameters. The optimization process explicitly targets the learning rate and parameters associated with the Adaptive Moment Estimation (Adam) optimizer [28]. Initially, the learning rate is set within a range from 10^{-4} to 10^{-2} , and the β values for the Adam optimizer vary between 0.9 to 0.999 and 0.99 to 0.999, respectively. Following the optimization with the `Keras-Tuner` library, the optimal hyperparameter settings for each data group are summarized in Table I. These configurations effectively balance the influence of past and current gradients, contributing to efficient optimization.

To perform grid localization using InterGridNet, we adhere

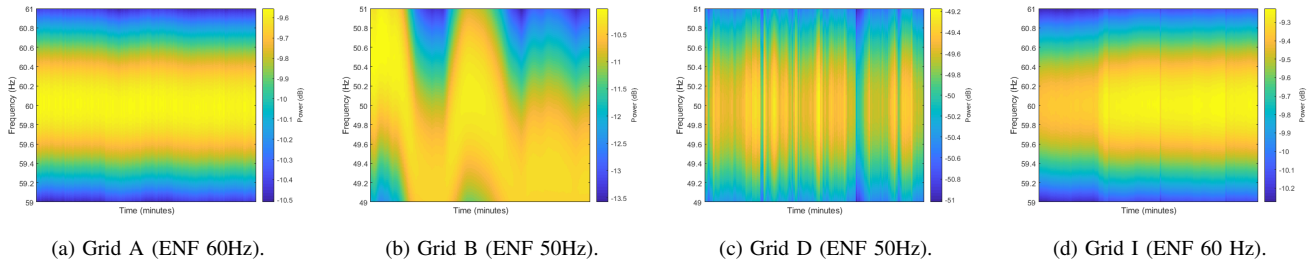


Figure 3. Spectrograms focused on the nominal ENF value for different grids.

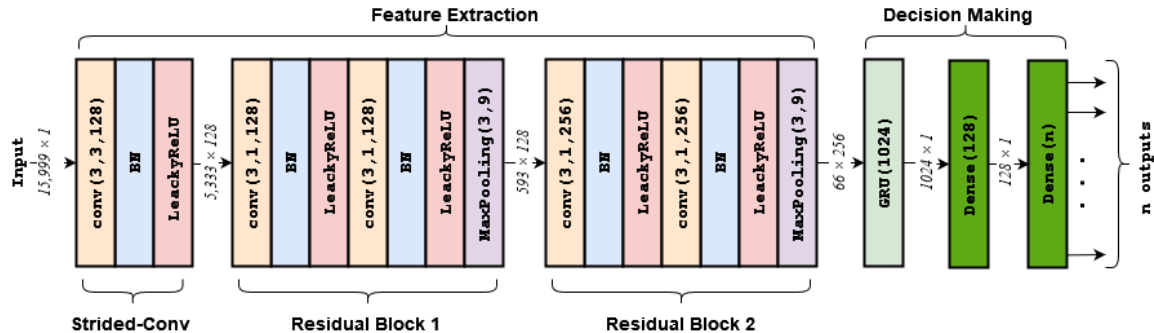


Figure 4. Architecture of the proposed optimized shallow RawNet model. The operators utilized in the network include Conv1D(kernel size, strides, filters), MaxPool1D(pool size, strides), GRU(units), and Dense(nodes).

to the outlined steps depicted in Figure 1. After data preparation, each recording frame undergoes classification by the neural network, resulting in a probability distribution across classes as determined by the softmax activation function of the last layer. For the classification of a frame into one of the known classes, the predictions should satisfy the rule:

$$-\sum_{i=1}^n p_i(x) \log_2 p_i(x) < \alpha_1 \cdot \log_2(n), \quad (1)$$

where p_i is the probability for each class prediction from the softmax and n is the number of classes in the frame's data group. In cases where the frame fails to meet (1), it is classified as **N**.

Subsequently, a majority voting mechanism is employed to ascertain the final estimate. The final estimate is deemed valid only if it appears in at least α_2 of the frames' predictions; otherwise, it is designated as **N**. Through the validation process, thresholds α_1 and α_2 have been set to 0.8 and 0.75, respectively. This approach ensures robustness in the grid localization process by requiring a consistent majority agreement across frames for a conclusive final estimation.

IV. EXPERIMENTAL RESULTS

In this section, the validation and testing of the InterGridNet are disclosed [29]. Additionally, limitations are discussed, providing valuable insights into the model's performance and areas for potential improvement.

TABLE II. INTERGRIDNET VALIDATION ACCURACY.

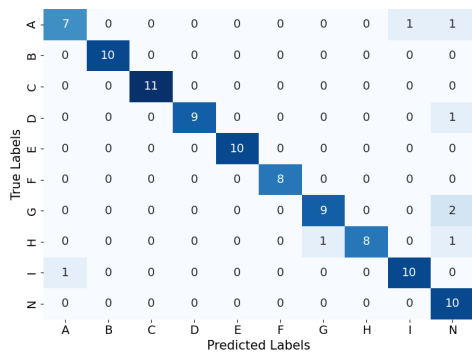
| Type | A | B | C | D | E | F | G | H | I | N | Overall |
|-------|------|------|------|------|------|------|-----|------|------|------|---------|
| Audio | 80% | 100% | 100% | 100% | 80% | 100% | 80% | 80% | 100% | 100% | 80% |
| Power | 100% | 100% | 100% | 100% | 100% | 100% | 80% | 100% | 100% | 100% | 96.67% |
| All | 80% | 100% | 100% | 100% | 80% | 100% | 60% | 80% | 100% | 100% | 90% |

A. Model Validation and Testing

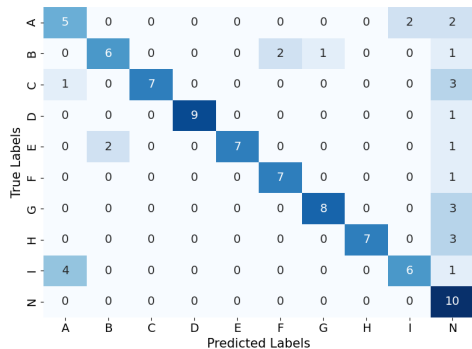
At the training phase of each model, all available training data depicted in Figure 2, corresponding to each data group, were utilized. For validation purposes and to experimentally determine the coefficients α_1 and α_2 , the practice set from the SP Cup 2016 dataset was employed. This shares identical characteristics with the testing set described in Section III-A and consists of 50 samples (5 samples for each class).

Table II summarizes the accuracy achieved for each class in the practice set of applying InterGridNet after completing model training and coefficient tuning. The classifier exhibits superior performance in the Power recordings compared to the Audio recordings as the Power recordings contain stronger ENF traces, and the corresponding classifiers benefit from a larger volume of training data, contributing to enhanced performance. In addition, class "None" has 100% accuracy, as shown in column **N**, underscoring the effectiveness of the "None" sample identification method outlined in Section III-B. The aggregate accuracy of the framework culminates at 90%.

The final assessment of InterGridNet's performance was conducted utilizing the dataset testing set. In Figure 5(a), the confusion matrix derived from the predictions of the proposed framework is illustrated, yielding an overall accuracy



(a) Bandpass filtering is applied.



(b) No filtering is applied.

Figure 5. Confusion matrices predictions on the testing set employing the proposed InterGridNet.

of 92%. Notably, misclassifications between classes **A-I** are minimal, owing to the inherent constraints of the data splitting technique, which refrains from classifying a sample with ENF at 50Hz into classes **A**, **C**, or **I** with ENF at 60Hz, and vice versa. Consistent with expectations, the testing accuracy closely aligns with the validation accuracy.

B. Discussion

The achieved testing accuracy of 92% underscores the unique characteristics embedded in the ENF signal per grid. Unlike analytical feature extraction methods [30]–[35], these distinctive features, crucial for solving the classification problem, are effectively extracted by the residual blocks and the GRU layer of the neural network described in Section III-B. This observation suggests that the chosen architecture demonstrates exceptional suitability for processing the ENF signal within raw audio data.

Figures 5(a) and 5(b) present the impact of frequency filtering around the nominal ENF on the classification. When this filtering is not applied, the overall accuracy is 72%, significantly lower compared to the scenario with bandpass filtering. This underscores the significant contribution of the ENF signal to accurately determining the grid corresponding to the recording location. In Figure 5(a), the misclassifications by InterGridNet predominantly categorize samples as “None” (class **N**). This exposes a vulnerability of (1) in the framework

TABLE III. TESTING ACCURACIES (%) IN SP CUP 2016 DATASET.

| Method | Characteristic | Accuracy |
|--|----------------------------|------------|
| SVM [30] | One-vs-one classification | 86% |
| SVM [31] | Multi-class classification | 77% |
| SVM [35] | Multi-class classification | 88% |
| Random Forest, SVM, AdaBoost [32] | Ensemble method | 88% |
| Binary SVM [33] | Binary classification | 87% |
| Multi-Harmonic Histogram Comparison [34] | Frequency domain analysis | 88% |
| InterGridNet (Ours) | Shallow RawNet | 92% |

but also underscores its confidence when handling samples from grids on which it has been trained. This dual observation provides insights into the framework’s strengths and areas for potential improvement.

Table III summarizes the testing accuracy of other methods using the same testing set. The data highlights the superiority of the proposed InterGridNet framework over previous works, reaffirming its effectiveness in geolocating sound recordings. Hence, InterGridNet is a powerful tool in the field, showcasing its potential for advancing state-of-the-art audio source grid location classification.

In [11], authored by our team, a fusion model comprising five machine learning classifiers was developed, trained, and tested using audio spectrograms from the nine ENF grids. This model achieved a testing accuracy of 96%, compared to the 92% accuracy of the proposed InterGridNet. While the higher accuracy of the fusion model can be attributed to its combination of multiple classifiers, it’s important to note that it required a significantly larger parameter count, with 11 million parameters for the CNN alone, which further increased when including the parameters of the fusion framework’s classifiers. In contrast, InterGridNet, with a streamlined architecture of 7 million parameters, adopts a novel unified single-classifier approach based on raw audio input via a DNN, highlighting its innovation and efficiency in power grid classification without the need for classifier fusion.

V. CONCLUSIONS

This paper presents InterGrid, a novel framework for geolocating audio recordings across different power grids, incorporating optimization through NAS. Inspired by RawNet’s architecture, InterGridNet has employed a shallow version of RawNet, offering a dynamic framework that includes preprocessing techniques to tackle the complex challenge of inter-grid localization within audio recordings. Key techniques have been crucial, such as bandpass filtering of ENF data, integration of residual layers for extracting frame-level embeddings, and softmax activation for decision-making. This research has marked the first implementation of DNN methodology for classification with preprocessing methods, achieving a 92% accuracy rate on the SP Cup 2016 dataset. Future research will employ a transformer architecture for grid location classification. To enhance transparency and understand the model’s decision-making process, explainable AI (xAI) techniques will also be integrated to extract specific patterns associated with each grid.

ACKNOWLEDGMENT

This research was supported by the Hellenic Foundation for Research and Innovation (HFRI) under the “2nd Call for HFRI Research Projects to support Faculty Members & Researchers” (Project Number: 3888).

REFERENCES

[1] S. Vatansever, A. E. Dirik, and N. Memon, “Analysis of rolling shutter effect on ENF-based video forensics,” *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 9, pp. 2262–2275, 2019.

[2] C. Korgialas, C. Kotropoulos, and K. N. Plataniotis, “Leveraging electric network frequency estimation for audio authentication,” *IEEE Access*, 2024.

[3] C. Korgialas and C. Kotropoulos, “A robust RELAX-based algorithm for enhanced electric network frequency estimation,” in *Proc. of the 13th Hellenic Conference on Artificial Intelligence (SETN)*. ACM, 2024, pp. 1–6.

[4] C. Moysiadis, G. Karantaidis, and C. Kotropoulos, “Electric network frequency optical sensing devices,” in *Proc. of the 2023 IARIA Annual Congress on Frontiers in Science, Technology, Services, and Applications*. IARIA, 2023, pp. 1–6.

[5] R. Garg, A. L. Varna, and M. Wu, “Modeling and analysis of electric network frequency signal for timestamp verification,” in *Proc. of the IEEE International Workshop on Information Forensics and Security (WIFS)*. IEEE, 2012, pp. 67–72.

[6] G. Hua, “Error analysis of forensic ENF matching,” in *Proc. of the IEEE International Workshop on Information Forensics and Security (WIFS)*, 2018, pp. 1–7.

[7] L. Zheng, Y. Zhang, C. E. Lee, and V. L. L. Thing, “Time-of-recording estimation for audio recordings,” *Digital Investigation*, vol. 22, pp. S115–S126, 2017.

[8] G. Hua, J. Goh, and V. L. Thing, “A dynamic matching algorithm for audio timestamp identification using the ENF criterion,” *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 7, pp. 1045–1055, 2014.

[9] A. Hajj-Ahmad, R. Garg, and M. Wu, “ENF based location classification of sensor recordings,” in *Proc. of the IEEE International Workshop on Information Forensics and Security (WIFS)*, 2013, pp. 138–143.

[10] —, “ENF-based region-of-recording identification for media signals,” *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 6, pp. 1125–1136, 2015.

[11] G. Tzolopoulos, C. Korgialas, and C. Kotropoulos, “On spectrogram analysis in a multiple classifier fusion framework for power grid classification using electric network frequency,” in *Proc. of the 13th International Conference on Pattern Recognition Applications and Methods*. SCITEPRESS, 2024, pp. 91–99.

[12] W. Yao *et al.*, “Source location identification of distribution-level electric network frequency signals at multiple geographic scales,” *IEEE Access*, vol. 5, pp. 11 166–11 175, 2017.

[13] R. Garg, A. Hajj-Ahmad, and M. Wu, “Feasibility study on intra-grid location estimation using power ENF signals,” 2021, arXiv:2105.00668.

[14] —, “Geo-location estimation from electrical network frequency signals,” in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 2862–2866.

[15] M. M. Elmesalawy and M. M. Eissa, “New forensic ENF reference database for media recording authentication based on harmony search technique using GIS and wide area frequency measurements,” *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 4, pp. 633–644, 2014.

[16] M. H. J. Bollen and I. Y. H. Gu, *Signal Processing of Power Quality Disturbances*. John Wiley & Sons, 2006.

[17] S.-J. Tsai *et al.*, “Frequency sensitivity and electromechanical propagation simulation study in large power systems,” *IEEE Transactions on Circuits and Systems*, vol. 54, no. 8, pp. 1819–1828, 2007.

[18] A. Hajj-Ahmad, R. Garg, and M. Wu, “Instantaneous frequency estimation and localization for ENF signals,” in *Proc. of the Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, 2012, pp. 1–10.

[19] Y. Jeon, M. Kim, H. Kim, H. Kim, J. H. Huh, and J. Yoon, “I’m listening to your location! Inferring user location with acoustic side channels,” in *Proc. of the World Wide Web Conference*, 2018, pp. 339–348.

[20] Y. Cui, Y. Liu, P. Fuhr, and M. Morales-Rodriguez, “Exploiting spatial signatures of power ENF signal for measurement source authentication,” in *Proc. of the IEEE International Symposium on Technologies for Homeland Security*, 2018, pp. 1–6.

[21] J.-W. Jung, H.-S. Heo, J.-H. Kim, H.-J. Shim, and H.-J. Yu, “RawNet: Advanced end-to-end deep neural network using raw waveforms for text-independent speaker verification,” *arXiv preprint arXiv:1904.08104*, 2019.

[22] M. Wu, A. Hajj-Ahmad, M. Kirchner, Y. Ren, C. Zhang, and P. Campisi, “Location signatures that you don’t see: Highlights from the IEEE Signal Processing Cup student competition,” *IEEE Signal Processing Magazine*, vol. 33, no. 5, pp. 149–156, 2016.

[23] C. Grigoras, “Digital audio recording analysis—the electric network frequency criterion,” *International Journal of Speech Language and the Law*, vol. 12, no. 1, pp. 63–76, 2005.

[24] M. Sarkar, D. Chowdhury, C. Shahnaz, and S. A. Fattah, “Application of electrical network frequency of digital recordings for location-stamp verification,” *Applied Sciences*, vol. 9, no. 15, p. 3135, 2019.

[25] Ž. Šarić, A. Žunić, T. Zrnić, M. Knežević, D. Despotović, and T. Delić, “Improving location of recording classification using electric network frequency (ENF) analysis,” in *Proc. of the IEEE International Symposium on Intelligent Systems and Informatics*, 2016, pp. 51–56.

[26] Y. Li, T. Lu, G. Zeng, K. Zhao, and S. Peng, “Advanced enf region classification using units-sinspec: A novel approach integrating sinusoidal activation function and spectral attention,” *Applied Sciences*, vol. 14, no. 19, p. 9081, 2024.

[27] A. Hajj-Ahmad, “ENF power frequency data for location forensics,” <https://dx.doi.org/10.21227/H2159S>, August 2016, Signal Processing Cup.

[28] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

[29] “InterGridNet: An electric network frequency approach for audio source location classification using convolutional neural networks,” [Accessed: October 10, 2024]. [Online]. Available: <https://github.com/ckorgial/InterGridNet>

[30] A. Triantafyllopoulos *et al.*, “Exploring power signatures for location forensics of media recordings,” University of Patras, Greece, Tech. Rep., 2016, Signal Processing Cup.

[31] R. Ohib, S. Y. Arnob, R. Arefin, M. Amin, and T. Reza, “ENF based grid classification system: Identifying the region of origin of digital recordings,” *Criterion*, vol. 3, no. 4, p. 5, 2017.

[32] M. El Helou, A. W. Turkmani, R. Chanouha, and S. Charbaji, “A novel ENF extraction approach for region-of-recording identification of media recordings,” *Forensic Science International*, vol. 155, no. 2-3, p. 165, 2005.

[33] D. Despotović *et al.*, “Exploring power signatures for location forensics of media recordings,” University of Novi Sad, Serbia, Tech. Rep., 2016, Signal Processing Cup.

[34] C. Chow *et al.*, “Multi-harmonic histogram comparison,” Purdue University, Tech. Rep., 2016, Signal Processing Cup.

[35] H. Zhou, *et al.*, “Geographic location estimation from ENF signals with high accuracy,” University of Science and Technology of China, Tech. Rep., 2016, Signal Processing Cup.