On acoustic monitoring of rainfall intensity

Carlo Monti and Stavros Ntalampiras
Department of Computer Science, University of Milan
carlo.monti@studenti.unimi.it, stavros.ntalampiras@unimi.it

***Abstract*—Audio-based environmental monitoring is gaining ever-increasing interest in the last decades facilitating a wide range of applications. An emerging task concerns the automatic estimation of rainfall intensity based on the respective acoustic activity. This work proposes an audio processing and modelling pipeline tailored to the requirements of the specific task. More precisely, we a) preprocessed the audio signals through filtering prior to feature extraction, b) integrated meteorological data as auxiliary features, c) explored different FFT window lengths considering the stationary characteristics of the available data, and d) constructed an ensemble model by stacking multiple transformer-based regressors. Importantly, during this analysis, we employed a publicly available dataset, i.e. SARID, adopting a standardized experimental protocol enabling reliable comparison of different approaches. Finally, the optimised model ensemble achieved a noticeable increase over the state of the art. Last but not least, the implementation of the described experimental pipeline is available at https://www.kaggle.com/code/imemine/ensemble-model-for-rain-intensity-estimation.**

***Index Terms*—Environmental monitoring, rainfall estimation, audio pattern recognition, audio surveillance, transformers, ensemble modeling**

I. Introduction

Audio-based environmental monitoring may offer efficient solutions in contexts with heterogeneous requirements and objectives, such as assessing ecosystems, urban areas, and wildlife habitats, to name but a few [1]–[5]. Audio signal processing and pattern recognition technologies can address a wide gamut of applications ranging from detection of environmental changes to tracking urban noise levels, while allowing for non-invasive data collection. A relatively recent application concerns rainfall monitoring, which is a critical task in environmental sciences, playing a vital role in water resource management, agriculture, flood forecasting and climate studies [6]. Accurate and reliable measurement of rainfall is essential for understanding precipitation patterns and their broader environmental and societal impacts [7], [8].

Traditionally, rainfall intensity and accumulation are measured using rain gauges—devices that collect and quantify precipitation [9]. Unfortunately, due to their construction, which involves a funnel for collecting the raindrops, rain gauges face several technical challenges, particularly in remote or inaccessible locations. Regular maintenance is necessary to ensure their accuracy as debris, insects or sediment can clog the instruments, rendering them ineffective or leading to inaccurate measurements. This maintenance requirement not only increases operational costs but also limits the scalability of rain gauge networks in regions where infrastructure and resources are constrained.

Given these limitations, alternative methods for rainfall monitoring that a) are cost-effective, b) necessitate low-maintenance, and c) can be suitably deployed in remote areas have being explored. Among those, the use of indirect approaches, such as analyzing the generated audio signals, have been investigated. Current research in this area can be categorized into three three main areas of focus.

The first line focuses on underwater rainfall sensing using audio. This approach leverages underwater acoustic signals to estimate rainfall intensity and has been applied in real-world scenarios [10], [11]. By utilizing existing underwater devices, this method provides an effective way to measure rainfall without requiring additional infrastructure in marine environments.

The second line of research involves the development of custom devices designed to infer rainfall intensity by capturing the sound of raindrops impacting a predefined surface. These devices utilize plates of specific materials and dimensions to create a controlled environment for sound detection, ensuring consistency and measurement reliability [12], [13]. This approach could extend the existing strategy of measuring the drop size distribution of atmospheric precipitation, a method traditionally employed using an instrument known as a laser disdrometer [14]. Disdrometers analyze the size, shape, and velocity of raindrops to provide detailed information about rainfall characteristics.

The third line of research, which is the focus of the present article, explores the use of machine learning techniques for rainfall detection based on surveillance audio [15]. This approach aims at exploiting the widespread network of surveillance cameras equipped with audio recording capabilities to enable a cost-effective and dense monitoring network for both urban and remote areas. By utilizing existing infrastructure, this method has the potential to significantly expand rainfall monitoring coverage without the need for additional hardware installations. Existing solutions have explored several feature sets (Chroma, contrast, tonnetz, Mel Frequency Cepstral Coeffients, etc.) along with traditional machine learning and deep learning architectures [6], [16]–[18].

A major challenge in this domain is the lack of a consistent dataset of audio recordings paired with corresponding rainfall measurements. To the best of our knowledge, the only publicly available dataset is the Surveillance-Audio-Rainfall-Intensity-Dataset (SARID) [18]. SARID provides annotated audio recordings taken from six real-world rainfall events occurred at the Nanjing Normal University in China. The audio recordings have been split into chunks of 4 seconds

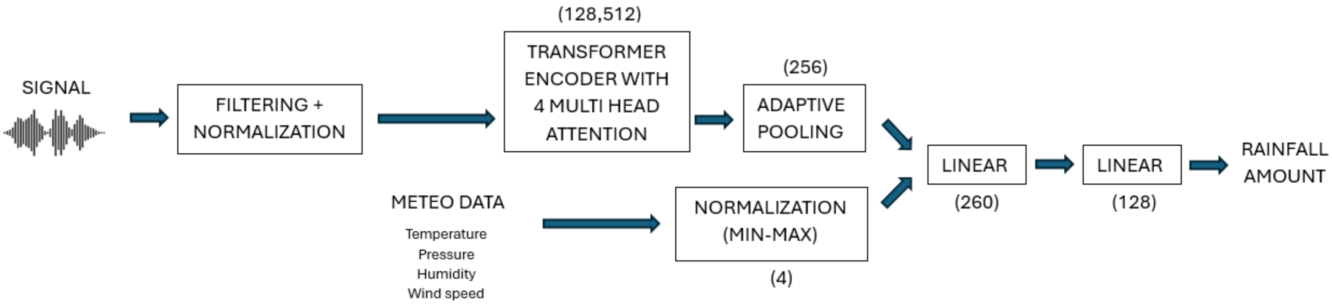

Fig. 1. The proposed experimental pipeline for optimizing the audio signal processing component and incorporating meteorological parameters.

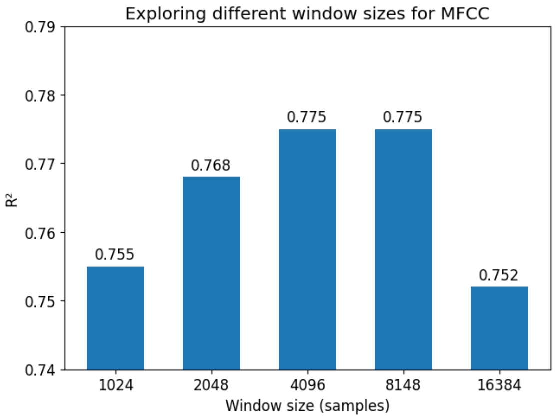

Fig. 2. The obtained results for different window sizes.

and subdivided homogeneously into training and testing sets.

This work advances the existing findings by optimizing the audio processing front-end combined with a suitably-trained ensemble of transformer models. Interestingly, this process provided valuable insights for the problem at hand. At the same time, the incorporation of meteorological parameters is also explored. Importantly, the proposed model ensemble outperforms the state of the art, while adopting a standardized experimental protocol.

This rest of this article is organised as follows: section II formulates the rainfall intensity estimation problem. Section III details and experimental construction of the proposed ensemble model including the optimization of the audio processing stage. Finally, in section IV we draw our conclusions and present fruitful directions for future research.

II. Problem Formulation

Let us consider a rainfall monitoring framework encompassing an acoustic sensor capturing the audiostream $y_t : \mathbb{N} \to \mathbb{R}$ and environmental sensors capturing temperature, pressure, humidity and wind speed $\theta_t, p_t, h_t, w_t : \mathbb{N} \to \mathbb{R}$ over time t. The specific timeseries are associated with rainfall intensity measurements denoted as $r_t : \mathbb{N} \to \mathbb{R}$. The overall aim is to create a model $\mathcal{M}$ accurately predicting r_t using the available measurements θ_t, p_t, h_t, w_t , i.e. $r_t = \mathcal{M}(\theta_t, p_t, h_t, w_t)$.

III. Experimental Construction of the Proposed Ensemble Model

The conducted work can be divided into three main phases, each one corresponding to three different strategies that we adopted to optimize the proposed regression model.

The first part focused on enhancing the audio data before feature extraction by applying low-pass filters at various cut-off frequencies (4000 Hz, 3000 Hz, 2000 Hz, and 1000 Hz) and by exploring different window lengths (and FFT resolutions).

The second part of the work aimed at complementing the audio-based models by incorporating additional meteorological parameters (humidity, pressure, temperature, and wind speed) taken from the SARID dataset. Such parameters were added to the model after the encoding phase and before the linear layer during training (see Fig. 1).

For these two parts, we used the same Transformer architecture proposed in the reference paper, which consists of four stacked encoders (with 4 attention heads and a feed-forward size of 512), followed by a Global Average Pooling layer and two fully connected layers.

During the last phase, we constructed an ensemble stacking model elaborating the outputs of three transformer-based models, each one trained with a suitably-optimized feature set, and a linear regressor.

The above-mentioned phases are explained in the following subsections. It should be mentioned that during all experiments were conducted following the standardized experimental protocol suggested in [18]. Both feature extraction and modelling stages were optimized on a validation set, which is part of the training set, while the presented figures of merit are computed on the test set. Aiming at minimizing the need for domain knowledge, we employed the short-time Fourier transform (STFT) spectrogram, log-Mel spectrogram (MEL), and the Mel-Frequencies Cepstral Coefficients (MFCC) which offer complementary views of the audio structure and different attention levels [19], [20]. The results presented in the following sections are, for simplicity, generally reported using the R^2 score as the sole evaluation metric. This choice aligns with the approach adopted in the reference paper, where performance figures were also based on this metric. Moreover, despite the challenges posed by an asymmetric distribution

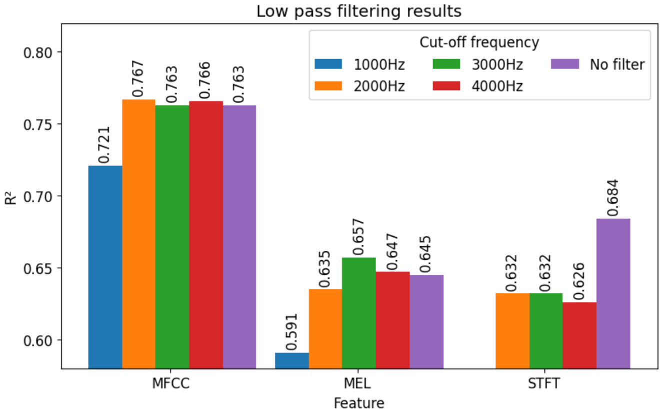

Fig. 3. The obtained results during the low-pass filtering phase.

(see section IV), the R^2 score is used here solely to compare the performance of different models as the input features vary. However, to better highlight the findings in III-D where a direct comparison with the baseline from the reference paper is made—we also report MAE and RMSE values, in order to provide a more comprehensive performance overview. The implementation of the proposed experimental pipeline is publicly available at https://www.kaggle.com/code/imemine/ensemble-model-for-rain-intensity-estimation.

A. Low-pass filtering

The rationale behind the filtering strategy comes from subjective auditory evaluations that suggested that the relevant part of rainfall sounds lies in the lower part of the spectrum. The goal was to reduce potential noise and focus on the frequency range most relevant for rainfall intensity detection. Audio signals were filtered at various cutoff frequencies, specifically 1000Hz, 2000Hz, 3000Hz, and 4000Hz, before undergoing feature extraction.

The results, shown in Fig. 3, reveal that applying a low-pass filter with a cut-off frequency around 2000-3000Hz provided a slight improvement in model performance compared to the unfiltered baseline. It should be noted that the model based on the STFT feature, unlike the rest, showed a consistent decrease in performance when the filter was applied that became worst when the filter cut-off was 1000Hz. However, reducing the cutoff frequency further to 1000Hz or below led to a noticeable decline in performance. This outcome shows that filtering out higher frequencies can be useful up to a certain extent since it may remove noise. At the same time, fine-tuning the filtering phase is needed to avoid losing useful information.

Overall, the results suggest that the information crucial for rainfall estimation is primarily concentrated below 2000 Hz, while the frequency components above this range appear not to introduce significant noise that would negatively impact detection accuracy. Despite the lack of significant improvement on the regression task, these findings could be exploited to effectively reduce the size of the data and the bandwidth required for real-time rainfall detection and estimation in real-world scenarios. By focusing on the critical frequency range, it may be possible to optimize the system for deployment in bandwidth-constrained environments without sacrificing accuracy. This also implies that a full-range microphone is not strictly necessary to collect the audio data.

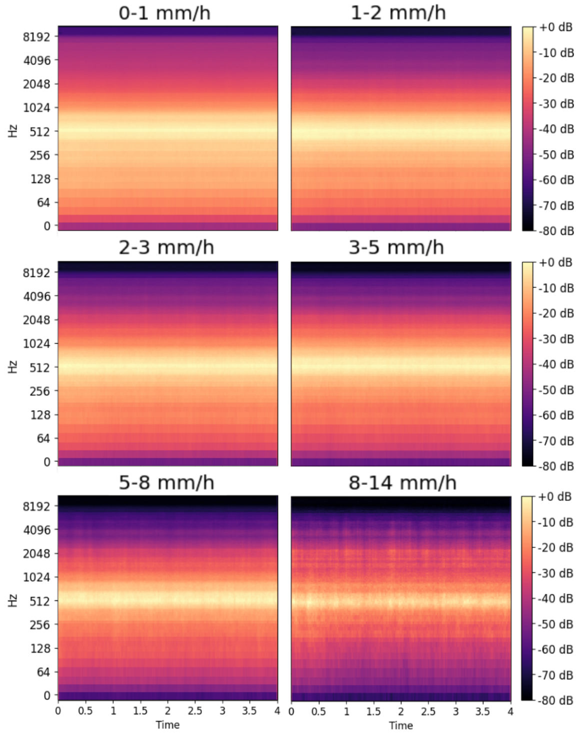

Fig. 4. STFT averages of samples representing various rainfall intensities.

B. Widening the window size

During the specific phase, We explored different window sizes for the FFT, guided by the idea that rainfall sounds are largely stationary and spread over a broader temporal window. Enlarging the window size naturally allows for greater frequency resolution. The respective results are shown in Fig. 2. There, we observe a considerable performance improvement as regards to the model trained with the MFCCs when using a window size of 4096 samples. However, increasing further the window size did not lead to additional improvements.

During this phase, we divided the samples into 6 bins (each containing L^b samples) according to the rainfall intensity range and we calculated the magnitude average $\bar{A}$ of the STFTs for frame m and frequency k as follows: $\bar{A}^b(m,k) = \frac{1}{L^b}\sum_{l=1}^{L^b}|STFT_l(m,k)|$. A visual analysis of the spectrum reveals that it is indeed highly stationary and concentrated around a well-defined central frequency (Fig. 4). Additionally, the spectrum exhibits a pattern that appears to vary in direct proportion to the intensity of the rainfall. In samples with light rain, the spectrum is more distributed across frequencies, whereas in samples with heavy rain, it becomes more concentrated within the main frequency range. This characteristic can therefore be considered the primary cue utilized by the regression algorithm and can be used to explain the functioning of the presented system. It is important to note

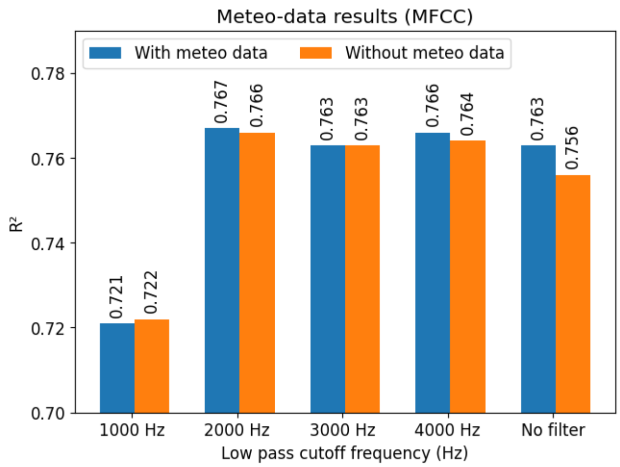

Fig. 5. Results obtained when incorporating meteorogical parameters during model training.

that the STFT is here normalized before plotting, meaning that the observed variations in frequency distribution could also be accompanied by changes in the overall energy of the signal.

C. Incorporating Meteorological Parameters

The third experimental phase focuses on enhancing the constructed models by incorporating additional features already present in the SARID dataset but not utilized during the original training process. These features include generic meteorological data such as *humidity*, *pressure*, *temperature*, and *wind speed*, which could potentially provide complementary information to the audio-based rainfall intensity detection.

The meteorological data were first normalized (min-max) and then added to the linear vector resulting from the pooling phase as shown in Fig. 1. As such, the linear dense layers will integrate both the audio-derived and meteorological features to make their predictions. This strategy follows the state of the art of CNN-based modelling, where the metadata, when needed, are added after the convolution phase [21]. Despite the theoretical advantages of including these additional parameters, the results did not show a significant improvement in performance (see Fig. 5). We argue that further exploration into how auxiliary features can be effectively integrated into such audio-based regression models is needed.

D. Stacking ensemble method

The next phase of this work involved combining the results obtained from the three models into an ensemble model using a *stacking* approach. To this end, we build a synergistic approach using a linear regressor the inputs of which are the best-performing transformer models trained on the three considered feature sets, i.e. MFCC, Mel, and STFT. The choice of a stacking approach was driven by the need to combine models with similar architectures but trained on different features. Additionally, this choice was influenced by the performance of these models, which varied considerably depending on the employed features. More in detail, MFCCs consistently achieved higher performance compared to the others in most rainfall intensities. However, we observed that the remaining feature sets may perform more accurately in cases where MFCCs underperform, thereby improving the ensemble model's prediction.

TABLE I
COMPARISON OF THE BEST-PERFORMING MODELS AND THE STATE OF THE ART. THE BEST FIGURES OR MERIT ARE EMBOLDENED.

Model	R^2	**MAE** (mm/h)	**RMSE** (mm/h)
[18]	0.765	0.563	0.88
Transformer with MEL	0.674	0.645	1.034
Transformer with STFT	0.691	0.633	1.007
Transformer with MFCC	0.777	0.538	0.855
Stacking ensemble	**0.787**	**0.52**	**0.837**

As such, we combined the predictions of the three models via a linear regression model, which surpassed all independent models with respect to all figures of merit (see Table. I). Importantly, the constructed ensemble outperforms the state of the art [18]. This result demonstrates that an ensemble model can enhance prediction bias, particularly in cases where the rainfall intensity is high. Despite this improvement, in Fig. 6 we see that the errors made by the ensemble are considerably higher for high rainfall intensities. This is possibly due to the imbalances existing in the available dataset favouring low intensity ranges as discussed in the following section.

IV. Conclusions and Future Work

Motivated by the characteristics of the present problem, this work investigated multiple strategies to improve rainfall intensity estimation using audio data from surveillance cameras. Importantly, we employed a publicly available dataset along with a standardized experimental protocol. Several of the proposed strategies led to improved performance, while a visual analysis of the audio spectra helped identify a relevant aspect affecting the regression task, which could enable an explicit explanation of the working principle behind the ML system.

We emphasize SARID's potential in driving progress in this area of research as the dataset represents a significant advancement over previous rainfall audio datasets, which were substantially smaller, less comprehensive and not publicly available. However, it also comes with notable limitations:

(a) The first drawback is its highly imbalanced nature, with the vast majority of samples representing light rain events. This poses a considerable challenge for regression tasks, as the model may struggle to generalize effectively across different rainfall intensities, even when the task is transformed in a classification problem (as it is usually done in real-world applications where the rainfall amount is usually classified as *light*, *moderate* or *heavy*). For example, ICAO is a well-known standard for aeronautical meteorological observations that differentiates between the type of precipitation: *drizzle* and *rain* [22].

(b) Another limitation is the absence of samples without rain. While the dataset is suitable for rain accumulation

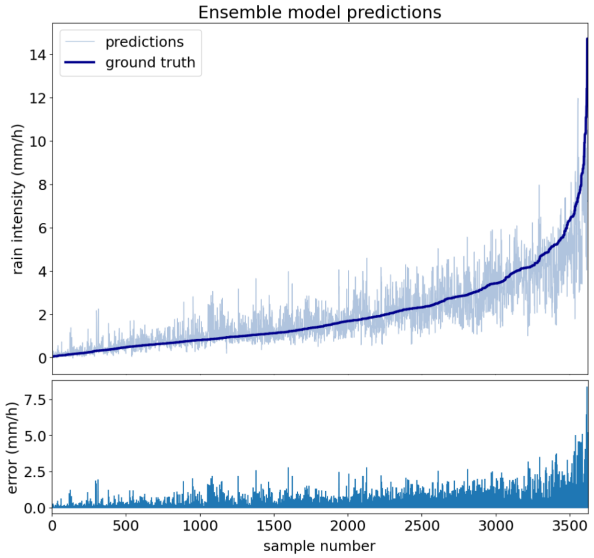

Fig. 6. Ground truth along with predictions and residual errors made by the ensemble model.

measurement tasks, the lack of "no-rain" samples may inhibit the model's ability to function as a "real-time" rain detector in practical applications, such as airport meteorology or urban monitoring for warning.

As such, while SARID provides a critical foundation for developing audio-based rainfall intensity estimation models, addressing these dataset limitations - through balancing, augmentation, and inclusion of no-rain samples — is an essential step towards unlocking its full potential and extend its applicability to real-world scenarios. Including additional data representing high rainfall intensities might be particularly beneficial.

Another potentially fruitful research direction might be the usage of sample-based modelling methods, which may compensate the dataset imbalances up to a certain extent [23]. In this direction, a hierarchical scheme may be applied, where the first step is responsible to classify rainfall intensities (possibly following the ICAO standard [22]), while the estimation is carried out at a second step, thus simplifying the problem space. Last but not least, we argue that audio explainability methods should be considered as they may provide valuable insights to the specific task [24].

References

[1] S. Fan, F. Xiao, S. Qi, Q. Zhu, W. Wang, and J. Guan, "Fine-grained audio feature representation with pretrained model and graph attention for traffic flow monitoring," DCASE2024 Challenge, Tech. Rep., June 2024.

[2] S. Ntalampiras, "Automatic acoustic classification of insect species based on directed acyclic graphs," *The Journal of the Acoustical Society of America*, vol. 145, no. 6, p. EL541–EL546, Jun. 2019.

[3] D. Stowell, "Computational bioacoustics with deep learning: a review and roadmap," *PeerJ*, vol. 10, p. e13152, Mar. 2022.

[4] B. W. Schuller, A. Akman, Y. Chang, H. Coppock, A. Gebhard, A. Kathan, E. Rituerto-González, A. Triantafyllopoulos, and F. B. Pokorny, "Ecology; computer audition: Applications of audio technology to monitor organisms and environment," *Heliyon*, vol. 10, no. 1, p. e23142, Jan. 2024.

[5] S. Ntalampiras and I. Potamitis, "Acoustic detection of unknown bird species and individuals," *CAAI Transactions on Intelligence Technology*, vol. 6, no. 3, p. 291–300, Mar. 2021.

[6] A. Emmanuel, B. Guda, N. D. Hailemariam, M. S. Meshach, and I. T. Barnabas, "Machine learning based rain gauge using acoustic data," in *2023 IEEE AFRICON*, 2023, pp. 1–4.

[7] World Meteorological Organization, "Guide to meteorological instruments and methods of observation: (cimo guide). 2014 edition, updated in 2017.[superseded]," 2017. [Online]. Available: https://www.oceanbestpractices.net/handle/11329/886

[8] D. Liu, Y. Zhang, J. Zhang, L. Xiong, P. Liu, H. Chen, and J. Yin, "Rainfall estimation using measurement report data from time-division long term evolution networks," *Journal of Hydrology*, vol. 600, p. 126530, Sep. 2021.

[9] S. Grimaldi, A. Petroselli, L. Baldini, and E. Gorgucci, "Description and preliminary results of a 100 square meter rain gauge," *Journal of Hydrology*, vol. 556, p. 827–834, Jan. 2018.

[10] S. Pensieri, R. Bozzano, J. A. Nystuen, E. N. Anagnostou, M. N. Anagnostou, and R. Bechini, "Underwater acoustic measurements to estimate wind and rainfall in the mediterranean sea," *Advances in Meteorology*, vol. 2015, p. 1–18, 2015.

[11] A. Trucco, R. Bozzano, E. Fava, S. Pensieri, A. Verri, and A. Barla, "A supervised learning approach for rainfall detection from underwater noise analysis," *IEEE Journal of Oceanic Engineering*, vol. 47, no. 1, pp. 213–225, 2022.

[12] S. Hwang, C. Jun, C. De Michele, H.-J. Kim, and J. Lee, "Rainfall observation leveraging raindrop sounds acquired using waterproof enclosure: Exploring optimal length of sounds for frequency analysis," *Sensors*, vol. 24, no. 13, p. 4281, Jul. 2024.

[13] R. Avanzato and F. Beritelli, "An innovative acoustic rain gauge based on convolutional neural networks," *Information*, vol. 11, no. 4, p. 183, Mar. 2020. [Online]. Available: http://dx.doi.org/10.3390/info11040183

[14] E. Adirosi, F. Porcù, M. Montopoli, L. Baldini, A. Bracci, V. Capozzi, C. Annella, G. Budillon, E. Bucchignani, A. L. Zollo, O. Cazzuli, G. Camisani, R. Bechini, R. Cremonini, A. Antonini, A. Ortolani, S. Melani, P. Valisa, and S. Scapin, "Database of the italian disdrometer network," *Earth System Science Data*, vol. 15, no. 6, p. 2417–2429, Jun. 2023.

[15] M. Wang, M. Chen, Z. Wang, Y. Guo, Y. Wu, W. Zhao, and X. Liu, "Estimating rainfall intensity based on surveillance audio and deep-learning," *Environmental Science and Ecotechnology*, vol. 22, p. 100450, Nov. 2024.

[16] M. I. Alkhatib, A. Talei, T. K. Chang, A. A. Hermawan, and V. R. Pauwels, "Towards the development of a citizens' science-based acoustic rainfall sensing system," *Journal of Hydrology*, vol. 633, p. 130973, Apr. 2024.

[17] R. S. Xavier, M. Gosset, T. F. Maciel, T. Bicudo, L. A. d. Nascimento, E. Ramalho, and A. Fleischmann, "Measuring amazon rainfall intensity with sound recorders," *Geophysical Research Letters*, vol. 51, no. 20, Oct. 2024.

[18] M. Chen, X. Wang, M. Wang, X. Liu, Y. Wu, and X. Wang, "Estimating rainfall from surveillance audio based on parallel network with multi-scale fusion and attention mechanism," *Remote Sensing*, vol. 14, no. 22, p. 5750, Nov. 2022.

[19] H. Purwins, B. Li, T. Virtanen, J. Schlüter, S.-Y. Chang, and T. Sainath, "Deep learning for audio signal processing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 2, pp. 206–219, 2019.

[20] S. Ntalampiras, L. A. Ludovico, G. Presti, M. V. Vena, D. Fantini, T. Ogel, S. Celozzi, M. Battini, and S. Mattiello, "An integrated system for the acoustic monitoring of goat farms," *Ecological Informatics*, vol. 75, p. 102043, Jul. 2023.

[21] E. Vaghefi, S. Hosseini, B. Prorok, and E. Mirkoohi, "Geometrically-informed predictive modeling of melt pool depth in laser powder bed fusion using deep mlp-cnn and metadata integration," *Journal of Manufacturing Processes*, vol. 119, p. 952–963, Jun. 2024.

[22] International Civil Aviation Organization, *Manual on Automatic Meteorological Observing Systems at Aerodromes*, ser. Doc (International Civil Aviation Organization), 2006. [Online]. Available: https://books.google.it/books?id=HR1oUEPsvfQC

[23] S. Ntalampiras and A. Scalambrino, "Automatic prediction of disturbance caused by inter-floor sound events," *IEEE Transactions on Cognitive and Developmental Systems*, pp. 1–8, 2024.

[24] A. Akman and B. W. Schuller, "Audio explainable artificial intelligence: A review," *Intelligent Computing*, vol. 3, Jan. 2024.