Steel and Composite Structures, *Vol. 56, No. 6 (2025) 539-550*
https://doi.org/10.12989/scs.2025.56.6.539

Rainfall intensity estimation via raindrop sounds leveraging convolutional neural networks and low-cost IoT sensors

Seunghyun Hwang[1a], Jinwook Lee[2b], Carlo De Michele[3b], Jongyun Byun[1c], Donghwi Jung[4b] and Changhyun Jun*[4]

[1]*Department of Civil, Environmental and Architectural Engineering, Korea University, 145, Anam-ro, Seongbuk-gu, Seoul, Republic of Korea*
[2]*Department of Civil and Environmental Engineering, University of Hawaii at Manoa, 2500 Campus Road, Honolulu, HI 96822, USA*
[3]*Department of Civil and Environmental Engineering, Politecnico di Milano, 20133 Milano, Italy*
[4]*School of Civil, Environmental and Architectural Engineering, Korea University, 145, Anam-ro, Seongbuk-gu, Seoul, Republic of Korea*

(Received May 31, 2025, Revised August 15, 2025, Accepted August 25, 2025)

Abstract. This study proposes a convolutional neural networks (CNNs)-based framework for estimating rainfall intensity using acoustic signals acquired from raindrops. Raindrop sounds were collected under real-world conditions using an internet of things (IoT) sensor-based acoustic data collection device. The collected signals were then transformed into spectrotemporal representations via short-time Fourier transform (STFT) and mel-spectrogram analysis. A dual-stream CNNs model was constructed to learn from both spectrogram types, leveraging their complementary strengths in capturing high- and low-frequency signal characteristics across various rainfall intensities. The model was trained using a balanced dataset representing no rain, weak, moderate, and heavy rainfall, and validated against ground truth measurements from an optical disdrometer (i.e., OTT PARSIVEL^2). Evaluation results indicate that the proposed method yields promising performance, with a root mean square error of 4.89 mm/h, a mean absolute error of 2.02 mm/h, and a R^2 of 0.75. While the model effectively estimates weak to moderate rainfall, it tends to underestimate extreme rainfall events due to their underrepresentation in the training data. These findings demonstrate the feasibility of rainfall intensity estimation from acoustic signals and highlight the potential of deep learning-based acoustic sensing for hydrometeorological applications in observation-challenged areas.

Keywords: cognitive computing; convolutional neural networks; raindrop sound; rainfall estimation; spectral analysis

1. Introduction

Climate change has been triggering a wide range of extreme weather events across the globe (Hashim and Hashim 2016, Diffenbaugh *et al*. 2017, Ornes 2018, Faranda *et al*. 2021, Robinson 2021). This phenomenon is also evident in precipitation patterns, as the frequency of extreme rainfall events has increased due to climate change, resulting in significant damage in various countries (Guhathakurta *et al*. 2011, Zhu 2013, Yilmaz *et al*. 2014, O'Gorman 2015, Kirchmeier-Young and Zhang 2020). South Korea is no exception to this trend. Over the past six decades, the country has experienced a noticeable rise in both total precipitation and the occurrence of extreme rainfall events (Do *et al*. 2023). In particular, the frequency of localized heavy rainfall, commonly referred to as intense downpours, has shown a consistent upward trend (Park *et al*. 2021, Kim *et al*. 2023, Shin *et al*. 2025).

As these highly localized extreme precipitation events become more prevalent, there is an increasing demand for rapid and accurate monitoring and response strategies. However, conventional rainfall observation systems present several limitations. Ground-based rain gauges, due to their spatial discreteness, provide only partial information on the overall rainfall field (Villarini *et al*. 2008, Paz *et al*. 2020). Weather radar systems, while capable of broader coverage, require substantial investment in both installation and maintenance. Furthermore, since they primarily observe precipitation from several kilometers above the ground surface, discrepancies between radar-derived measurements and actual ground-level rainfall may arise (Chumchean *et al*. 2006, Martens *et al*. 2013). These challenges highlight the need for alternative approaches that can effectively complement existing observation networks.

Meanwhile, rapid advancements in artificial intelligence (AI) and computational power have enabled the analysis of massive datasets, paving the way for cognitive computing technologies that process signals in a manner analogous to human perception (Modha *et al*. 2011, Chen *et al*. 2018, Gupta *et al*. 2018). Cognitive computing has already demonstrated significant utility across various sectors of human life, offering meaningful contributions in numerous applications (Chen *et al*. 2016, Behera *et al*. 2019, Sreedevi *et al*. 2022). In the domain of precipitation monitoring, applications of cognitive computing have begun to emerge, particularly those utilizing visual and auditory signals. Notable examples include the use of surveillance cameras to analyze cloud imagery for rainfall prediction (Byun *et al*.

*Corresponding author, Ph.D. Professor
E-mail: cjun@korea.ac.kr
[a]Integrated Master & Ph.D. Student
[b]Ph.D.
[c]Ph.D. Student

ISSN: 1229-9367 (Print), 1598-6233 (Online)

2023) and to estimate the raindrop size and terminal velocity (Lee *et al*. 2023). Furthermore, several studies have employed sound signals: one categorized rainfall intensity into three levels based on acoustic data recorded via smartphones (Guo *et al*. 2018), while another classified rainfall intensity into five levels using audio collected from surveillance systems (Wang *et al*. 2022). These cases demonstrate the potential to establish a dense rainfall observation network over a wide area with greater cost-effectiveness and ease of installation compared to conventional rainfall monitoring methods. They can offer viable solutions to overcome the inherent limitations of traditional observation approaches. Especially, the use of acoustic signals is promising, as numerous previous studies have demonstrated the versatility and effectiveness of acoustic sensing in environmental monitoring applications (Raimbault and Dubois 2005, Kloser *et al*. 2009, Pijanowski *et al*. 2011, Howe *et al*. 2019). Given its proven potential, acoustic-based observation may offer a highly efficient alternative for rainfall monitoring. However, previous studies on acoustic-based rainfall observation have primarily focused on classifying rainfall intensity into a limited number of categorical levels, while attempts to quantitatively estimate rainfall intensity remain exceedingly rare. This scarcity of research highlights the necessity for further investigations aimed at establishing reliable methodologies for quantitative rainfall estimation using acoustic signals.

With the advancement of AI technologies, AI-based methodologies have been extensively applied in various domains of acoustic signal processing, including speech recognition, emotion classification, and noise reduction (El Ayadi *et al*. 2011, Hinton *et al*. 2012, Xu *et al*. 2013, Wang *et al*. 2018, George *et al*. 2024, Ahlawat *et al*. 2025). Among these, convolutional neural networks (CNNs) have demonstrated significant utility in acoustic signal analysis, particularly when combined with spectrogram extraction techniques such as short-time Fourier transform (STFT) and wavelet transform. The effectiveness of such combinations has been validated through numerous prior studies.

One representative domain in which CNNs-based acoustic analysis has been actively applied is healthcare. Fan *et al*. (2023) employed CNNs to enhance both the accuracy and interpretability of abnormal heart sound detection, aiming for early diagnosis of cardiovascular diseases. Utilizing the 2016 PhysioNet/CinC Challenge dataset and a self-constructed pediatric heart sound dataset, they proposed a learnable lifting wavelet transform (Le-LWT) block by integrating CNNs with the lifting wavelet transform (LWT). As a result, the Le-LWTNet outperformed state-of-the-art models in terms of accuracy, sensitivity, and specificity, while also requiring fewer trainable parameters, thereby demonstrating high applicability to real-time mobile-based implementations.

Another notable example of heart sound-based disease detection that integrates CNNs and STFT is presented in Guo *et al*. (2023). They proposed a method for efficiently screening cardiovascular diseases by analyzing heart sounds collected via wearable devices. Their study developed a dual-stream CNN (DS-CNN) model that utilized two types of inputs derived respectively from STFT and higher-order spectral (HOS) analysis. Using the same 2016 PhysioNet/CinC Challenge dataset as in Fan *et al*. (2023), the proposed DS-CNN model showed superior performance compared to existing cardiovascular disease detection models.

Beyond the healthcare domain, CNNs-based acoustic signal analysis has also found wide application in environmental monitoring. Mu *et al*. (2021) proposed a temporal-frequency attention-based CNN (TFCNN) to address various error factors in environmental sound classification (ESC). This model incorporated frequency attention and temporal attention mechanisms, assigning higher weights to critical frequency bands and important time frames in the log-mel spectrogram, respectively, and used the resulting features as inputs to the CNN. Validation with the UrbanSound8K and ESC-50 datasets demonstrated the model's superior performance, achieving classification accuracies of 93.1% and 84.4%, respectively.

Another example in ESC is the study by Garcia-Ordas *et al*. (2023), in which the authors developed a fully convolutional neural network (FCN) to recognize multiple bird species, aiming to identify climate-sensitive avian populations. The study utilized 2,699 audio samples collected from 17 bird species, with preprocessing focused on segments containing bird calls. The input features consisted of mel-spectrograms and mel frequency cepstral coefficients (MFCCs). The proposed FCN achieved an accuracy of 85.71%, representing more than a 67% improvement over conventional methods. Notably, mel-spectrograms yielded superior classification performance compared to MFCCs.

In this study, we propose a novel approach for acoustic-based rainfall observation method that leverages a CNNs for spectral analysis. The proposed method carries out quantitative estimation of rainfall intensity by utilizing spectrograms extracted from acoustic signals of rainfall, which are subsequently analyzed through CNNs. By exploiting the high applicability of CNNs in acoustic signal processing, the approach not only achieves enhanced accuracy in acoustic-based rainfall observation but also addresses the absence of prior studies attempting quantitative rainfall intensity estimation from acoustic signals. With this aim, a custom acoustic data collection device was developed to record rainfall acoustics under real rainfall conditions. From the collected data, frequency-domain spectrograms were extracted using STFT and mel-spectrogram transform. These extracted spectrograms were used as input to the CNNs model for rainfall intensity estimation. The resulting rainfall intensities were then validated by comparison with measurements obtained from an optical disdrometer (i.e., OTT PARSIVEL^2) to assess the model's reliability.

2. Materials

The overview of the proposed rainfall intensity estimation method using CNNs is illustrated in Fig. 1.

Rainfall acoustics are continuously recorded in 60 s

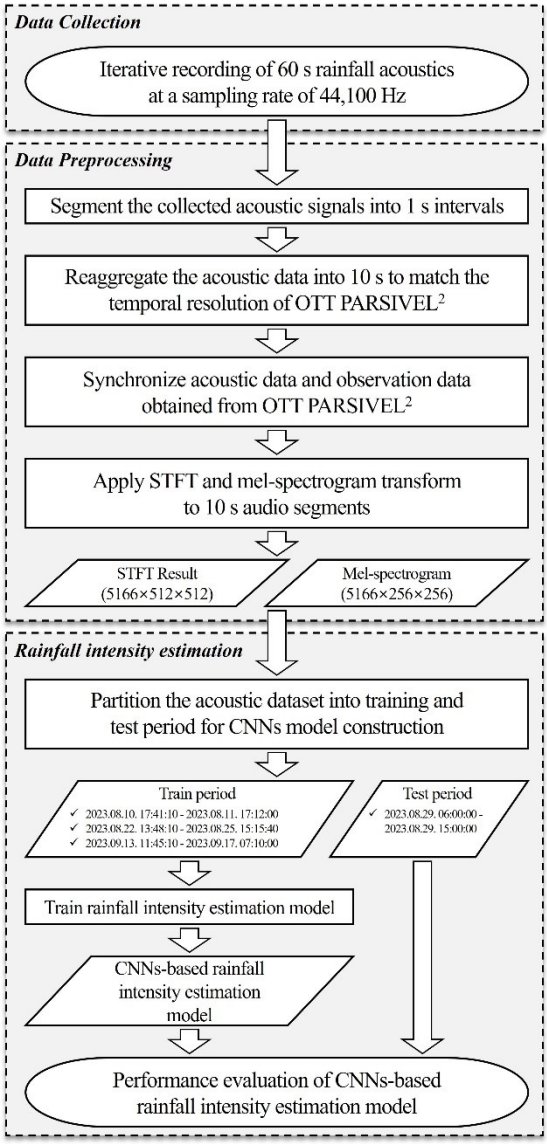


Fig. 1 Flowchart of the study for rainfall intensity estimation based on rainfall acoustics

segments through a custom-built acoustic data collection device consisting of a Raspberry Pi and a condenser microphone. During the data storage process, slight temporal delays were introduced, resulting in progressive shifts in the recording start times. To address this issue and to ensure precise temporal alignment between the collected acoustic signals and the PARSIVEL^2 observations used as validation data, the recorded signals were first segmented into 1 s intervals and subsequently reaggregated into 10 s segments, consistent with the temporal resolution of PARSIVEL^2. For each 10 s audio segment, both STFT and mel-spectrogram transform were applied to extract spectrogram representations. Finally, a CNNs-based model for rainfall intensity estimation was developed using the two spectrogram types as input features. Detailed descriptions of each step will be provided in subsequent sections.

The acoustic data collection device employed for collecting rainfall acoustics in this study was identical to the equipment utilized in our previous study by Hwang *et al*. (2024), which was another case of rainfall intensity estimation using rainfall acoustics. Hwang *et al*. (2024) presented a study that attempted to quantitatively estimate rainfall intensity based on rainfall acoustics. In their approach, the collected acoustic signals were segmented into fixed-length intervals, and four characteristic features were extracted from the STFT results of each segment. These features were then utilized in conjunction with an extreme gradient boosting (XGBoost) model to perform binary classification of rainfall occurrence and to estimate rainfall intensity. Identical to the system employed in Hwang *et al*. (2024), the acoustic data collection device used in this study was constructed by placing a Raspberry Pi and a condenser microphone inside a waterproof plastic enclosure. This configuration enables stable recording of acoustic signals under rainfall conditions, while the plastic housing not only prevents water penetration but also effectively suppresses environmental noise. Consequently, the device is highly effective in capturing the acoustic signals generated by raindrops impacting the enclosure surface. The developed acoustic data collection device was installed at Chung-Ang University in Seoul, Republic of Korea (37°20′13″ N, 126°57′27″ E).

Based on this device, rainfall acoustic data were collected across four events between August 10 and September 17, 2023. An overview of each event is provided in Table 1. Event 1 corresponds to a rainfall event induced by Typhoon Khanun, while Event 2 was associated with precipitation resulting from a high-pressure system situated over the East Sea. Event 3 represents an autumnal monsoonal rainfall event, formed by a stationary front caused by the interaction between the northward-moving East Sea high-pressure system and the cold Siberian air mass. Compared to the other rainfall events observed in August, Event 3 was characterized by shorter duration and lower accumulated precipitation. Exhibiting characteristics analogous to Event 2, Event 4 was likewise induced by a high-pressure system situated over the East Sea. Events 1 through 3 were utilized during the training process of the rainfall intensity estimation model, whereas Event 4 was employed to evaluate the accuracy of the final developed model. All events were collected at the same location using the identical acoustic collection device. For each event, the rainfall intensity observed at 10 s intervals by an optical disdrometer installed within a 10 m radius of the acoustic data collection site is presented in Fig. 2.

As illustrated in Fig. 2, the majority of rainfall acoustic data collected during the study period corresponded to no-rain conditions or low rainfall intensities below 15 mm/h. Notably, 98.47% of the acoustic recordings were associated with rainfall intensities less than 15 mm/h. Such a highly imbalanced distribution poses the risk of biased learning, wherein the model may overfit to the dominant class and fail to generalize across different rainfall intensities (Buda *et al*. 2018). To mitigate this issue, this study classified rainfall intensity into three distinct categories and applied a random sampling strategy to extract an equal number of acoustic samples from each category for model training. Constructing a balanced dataset through the modification of

Table 1 Description of the data used in the study

Event No.	Period	Maximum rainfall intensity (mm/h)	Precipitation (mm)	Usage
Event 1	2023.08.10. 17:41:10 - 2023.08.11. 17:12:00	48.750	25.495	Train set
Event 2	2023.08.22. 13:48:10 - 2023.08.25. 15:15:40	141.414	61.358	Train set
Event 3	2023.09.13. 11:45:10 - 2023.09.17. 07:10:00	107.846	61.728	Train set
Event 4	2023.08.29. 06:00:00 - 2023.08.29. 15:00:00	101.885	37.169	Test set

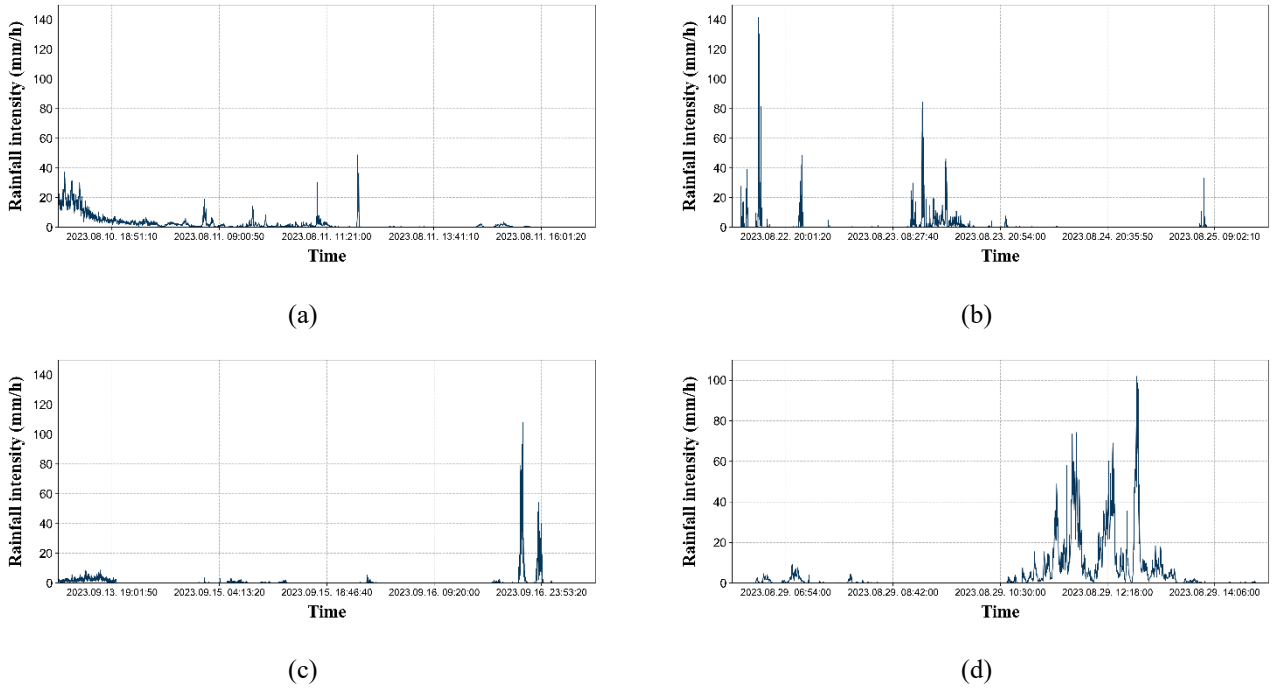


Fig. 2 Rainfall events observed during the research period: (a) Event 1; (b) Event 2; (c) Event 3; (d) Event 4

an imbalanced dataset using sampling methods has been shown to improve classification accuracy across a wide range of models (He and Garcia, 2009). A total of 1,998 acoustic samples were utilized for model training. Specifically, 333 acoustic samples were extracted for each rainfall intensity intervals: weak rainfall (0.5–15 mm/h), moderate rainfall (15–30 mm/h), and heavy rainfall (≥ 30 mm/h). To maintain class balance, an additional 999 acoustic samples corresponding to no rainfall conditions (< 0.5 mm/h) were also extracted, resulting in an equal number of rainy and non-rainy acoustic samples. Each acoustic sample corresponds to a 10 s segment of recorded sound.

To validate the estimated rainfall intensity in this study, observational data from an optical disdrometer collected at 10 s intervals were utilized, in accordance with the approach established in the prior work of Hwang *et al*. (2024). The optical disdrometer operates based on the principle of light extinction. It uses a horizontal laser light sheet (780 nm, 3 mW) and a single photodiode to measure the size and fall velocity of individual hydrometeors. As particles traverse the laser beam, they cause a transient reduction in the detected signal. The amplitude of this attenuation correlates with the cross-sectional area of the particle, allowing estimation of its diameter, while the duration of signal reduction corresponds to the fall velocity.

The system is capable of detecting particles within a diameter range of 0.3 to 30 mm and fall speeds up to 20 m s^{-1}. Calibration procedures involve empirical relationships derived from controlled free-fall experiments using particles of known size and terminal velocity, which ensure accuracy in both size and velocity measurements. For raindrops, size estimation errors are generally within ±100 μm and ±5%, and velocity estimation errors are within 25% for small drops (0.3 mm) and 10% for largest drops (5 mm). The optical disdrometer provides a robust and accurate solution for high-resolution precipitation measurements, making it especially suitable for validation in hydrometeorological studies (Löffler-Mang and Joss, 2000). Based on its high precision and accuracy, the optical disdrometer has been utilized as reference data for validation purposes in numerous previous studies (Das *et al*. 2017, Barcaroli *et al*. 2018, Bracci *et al*. 2023, Pu *et al*. 2023).

3. Methodology

As illustrated in the flowchart in Fig. 1, the methodology of this study is divided into three major components following the collection of rainfall acoustic data. First, two types of spectrograms are extracted from the

recorded acoustic signals. Second, a CNNs model is developed for rainfall intensity estimation using the extracted spectrograms as input features. Finally, the performance of the constructed CNNs model is evaluated. Detailed descriptions of these three components are provided in Sections 3.1 through 3.3.

3.1 Spectrogram extraction based on STFT and mel-spectrogram transform

The STFT is a representative time-frequency analysis technique designed for interpreting non-stationary signals whose statistical properties vary over time. STFT operates by segmenting the entire signal into fixed-duration temporal windows, within which the signal is assumed to be quasi-stationary. A Fourier transform is then applied to each windowed segment, enabling simultaneous resolution in both time and frequency domains. Mathematically, the STFT can be expressed as Eq. (1) (Allen and Rabiner, 1977). $x(m)$ denotes the input signal, $w(n)$ represents the analysis window function, and ω_k indicates the sampled frequency components. This formulation can be interpreted as applying a window function to weight a localized portion of the signal, followed by the application of the Fourier transform to extract its frequency content.

$$X_n(e^{j\omega_k}) = \sum_{m=-\infty}^{\infty} w\,(n-m)x(m)e^{-j\omega_k m} \tag{1}$$

The "mel" is a unit on a numerically constructed subjective scale that quantifies the perceived magnitude of pitch, developed to represent human auditory perception independently of physical frequency or musical scales (Stevens *et al*. 1937). The mel scale assigns higher resolution to lower frequency bands compared to higher frequencies, thereby enhancing sensitivity to spectral variations in the lower frequency range.

By applying the mel scale to the result of the STFT, it is possible to obtain a mel-spectrogram (Ustubioglu *et al*. 2023). To convert the STFT result into a mel-spectrogram, a mel-frequency filter bank in the form of triangular filters must be employed. The transformation function that determines the weight for each frequency component k in the STFT result, when computing the power spectral value of the m-th bin in the mel-spectrogram, is defined as Eq. (2) (Abdul and Al-Talabani 2022). $f(m)$ denotes the function that converts the mel-scale frequency of the m-th bin in the mel-spectrogram back to linear frequency in hertz (Hz).

$$H_m(k) = \begin{cases} 0, & k < f(m-1) \\ \dfrac{k - f(m-1)}{f(m) - f(m-1)}, & f(m-1) \le k < f(m) \\ 1, & k = f(m) \\ \dfrac{f(m+1) - k}{f(m+1) - f(m)}, & f(m) < k \le f(m+1) \\ 0, & k > f(m+1) \end{cases} \tag{2}$$

The mel frequency and the linear frequency can be interconverted using the following equation (Abdul and Al-Talabani 2022):

$$Mel\ frequency = 2595 \times \log_{10}\left(1 + \frac{linear\ frequency\ (Hz)}{700}\right) \tag{3}$$

The power spectrum value of the m-th bin in the mel-spectrogram can then be obtained by multiplying each power spectrum value of the STFT result by the corresponding weight ($H_m(k)$) and summing the results.

Rainfall acoustics corresponding to higher rainfall intensities tend to exhibit acoustic signals with higher frequency components, which are distinctly separated from background sounds typically found in lower frequency ranges (Hwang *et al*. 2024). In other words, the primary characteristics of rainfall acoustics under weaker rainfall conditions tend to appear in lower frequency bands. The mel-spectrogram, which applies a logarithmic scale and smoothing the high-frequency region, maintains the overall spectrogram size while achieving higher resolution in the lower frequency domain (Malayath and Hermansky 2003, Aziz and Shahnawazuddin 2023). By securing enhanced resolution in the lower frequency range, the mel-spectrogram has demonstrated improved analytical performance compared to uniform spectrogram extraction methods such as STFT, with reported cases showing superior outcomes (Han *et al*. 2024). Considering that weak rainfall events are characterized by acoustic features in the low-frequency range, the incorporation of mel-spectrograms in this study was expected to enhance the analysis of weak rainfall events relative to approaches that rely solely on STFT.

3.2 Rainfall intensity estimation model via CNNs

CNNs, inspired by the perceptual mechanisms of biological visual systems, represent a class of deep neural network models primarily utilized for pattern recognition in images (Li *et al*. 2021). CNNs are generally structured as a hierarchical architecture comprising convolutional layers, pooling layers, and fully connected layers. Convolutional layers extract progressively abstract features by computing the scalar product between learned weights and localized regions of the input, thereby capturing spatial patterns. Subsequently, pooling layers downsample the spatial dimensions of the feature maps, effectively reducing computational complexity and mitigating overfitting. Finally, fully-connected layers integrate the high-level features and generate class scores for final prediction, functioning analogously to conventional artificial neural networks (O'shea and Nash 2015).

Compared to fully connected neural networks, CNNs offer several advantages. First, through local connectivity, each neuron is connected to only a subset of neurons in the preceding layer rather than to all neurons, which effectively reduces the number of parameters and accelerates the convergence of learning. Second, weight sharing allows multiple connections to utilize the same set of weights, thereby further decreasing the model complexity. Third, pooling layers leverage the local correlation of image data to perform downsampling, which not only reduces the amount of data but also preserves informative features while discarding irrelevant ones, leading to additional parameter reduction. These architectural characteristics have been pivotal in establishing CNNs as one of the most prominent algorithms in the field of deep learning (Li *et al*.

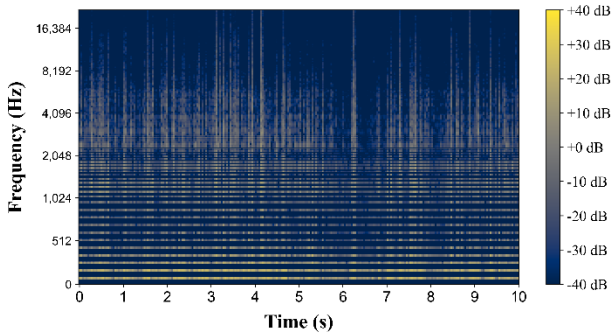


Fig. 3 Example of noise artifacts (horizontal stripes) in the low-frequency range of the mel-spectrogram due to resolution enhancement

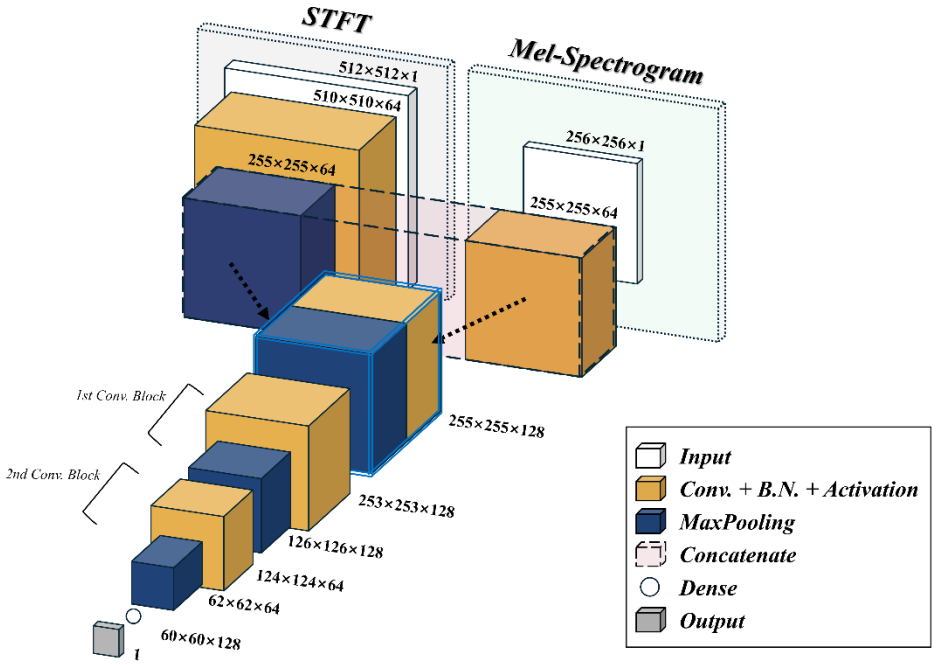


Fig. 4 Architecture of CNNs-based rainfall intensity estimation model leveraging STFT and mel-spectrogram

2021).

In this study, a CNNs model was developed to estimate rainfall intensity using two types of spectrograms (i.e., STFT, mel-spectrogram) as input data. The collected rainfall sounds were divided into 10 s segments, and spectrograms for each 10 s segment were extracted. As previously mentioned, the mel-spectrogram, which provides high resolution in the low-frequency range, was used alongside the STFT in an effort to improve the detection performance for weak rainfall.

When utilizing the two spectrograms as input data for the CNNs model, it is necessary for both spectrograms to be extracted at the same size in order to merge them before inputting them into the CNNs. However, since the mel-spectrogram requires high resolution in the low-frequency range, extracting it at the same size as the STFT may result in the introduction of horizontal line noise in the low-frequency range during the process of increasing the resolution with the mel-scale (Fig. 3). To address this issue, frequency domain analysis should be conducted with finer segmentation. However, this approach may lead to excessive processing time and delays in mel-spectrogram extraction.

Therefore, this study adopted a structure where the STFT and mel-spectrogram are extracted at different sizes, and the two input data are merged through a concatenation layer at the feature level of the CNNs. This method allows the two input data to have different sizes, setting the input size of the mel-spectrogram to be smaller than that of the STFT. Fig. 4 illustrates the CNNs structure used in this study. The STFT extracted at a size of (512 × 512) undergoes convolution and max pooling layers, reducing its size, and is designed to be merged with the mel-spectrogram extracted at a size of (256 × 256) after one convolution layer has been applied.

3.3 Performance evaluation

To evaluate the performance of the proposed CNNs-based rainfall intensity estimation method using rainfall

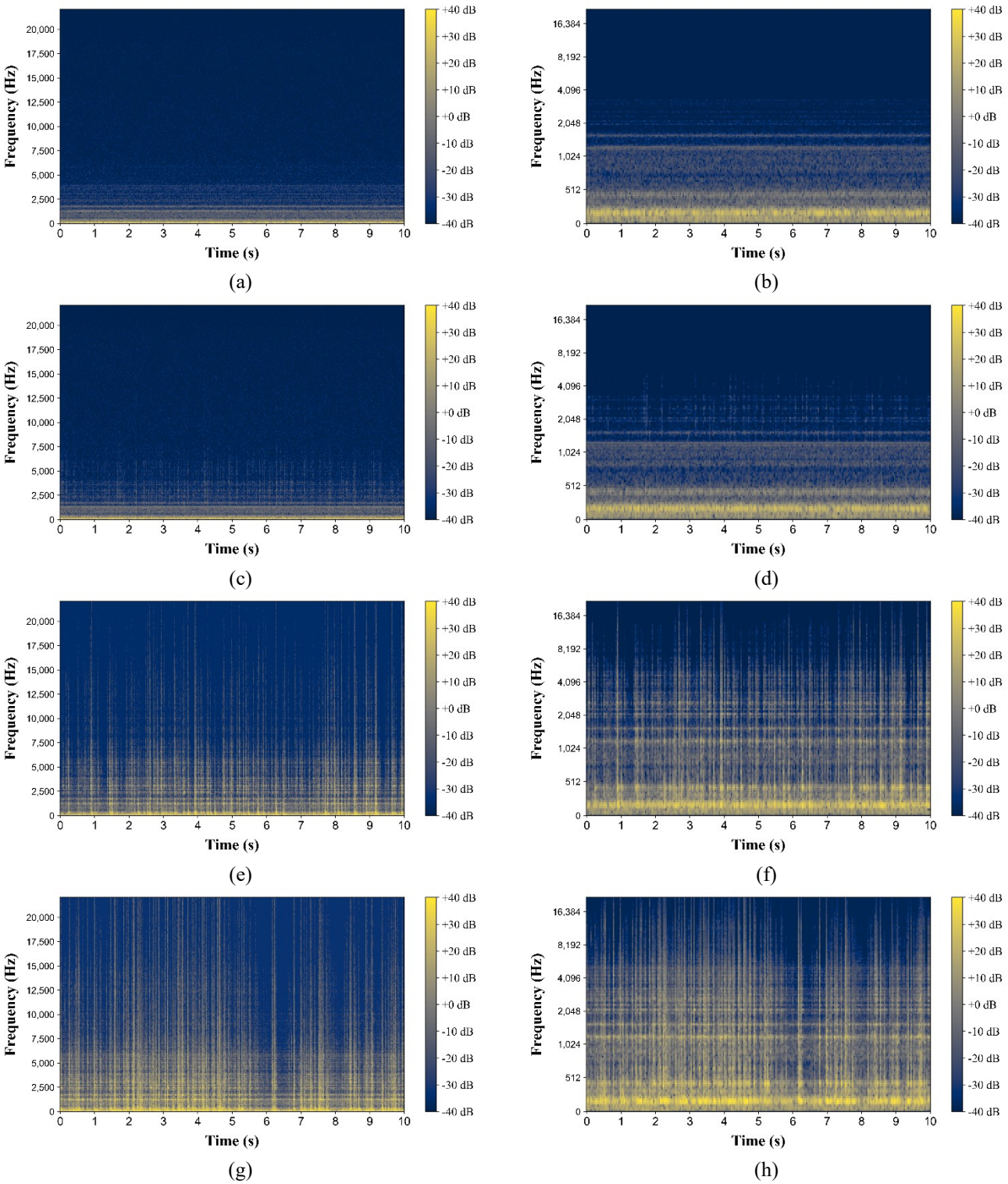
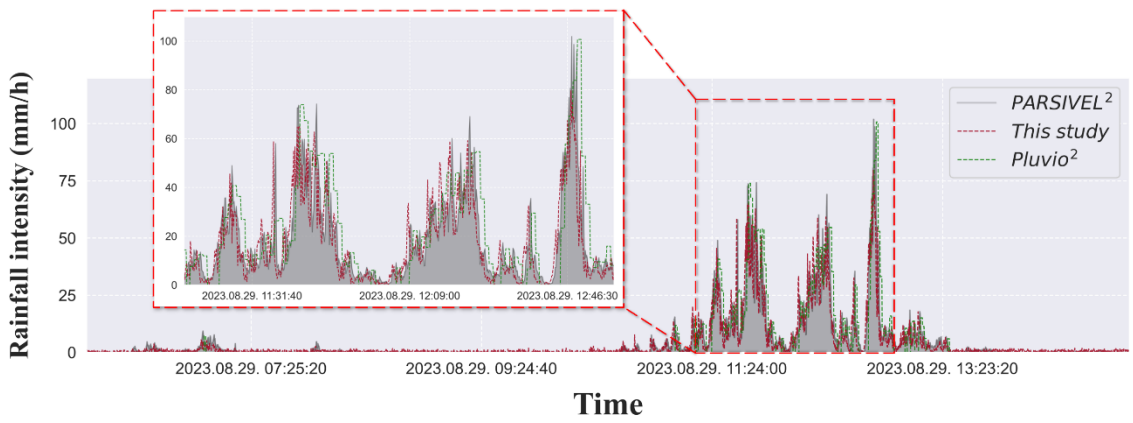


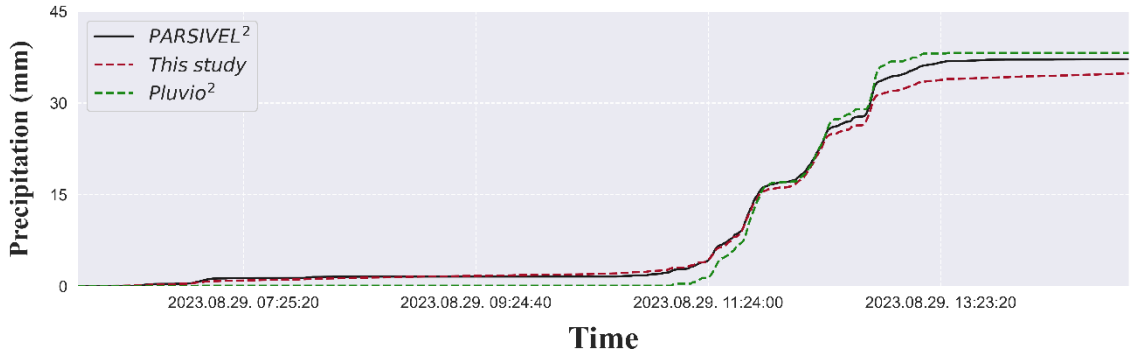
Fig. 5 Examples of spectrogram extraction using STFT and mel-spectrogram for different rainfall intensity levels (left: STFT, right: mel-spectrogram): (a), (b) no rain (0 mm/h); (c), (d) weak rain (1.483 mm/h); (e), (f) moderate rain (18.261 mm/h); (g), (h) heavy rain (141.414 mm/h)

acoustics, rainfall intensity was estimated based on acoustic signals recorded during a rainfall event that occurred between 06:00 and 15:00 on August 29, 2023 (Event 4). The estimated intensities were then compared with those measured by an optical disdrometer installed at the same location. Fig. 2(d) presents the rainfall intensities observed by the optical disdrometer during the validation period. The maximum rainfall intensity recorded during this event was 101.885 mm/h, with a total accumulated rainfall of 37.169 mm.

To quantitatively evaluate the performance of rainfall intensity estimation, three validation metrics were



(a)



(b)

Fig. 6 Rainfall intensity estimation results using rainfall acoustics based on CNNs: (a) time series rainfall intensities; (b) cumulative precipitation

computed: root mean square error (RMSE), mean absolute error (MAE), and R^2. The calculation methods for each metric are expressed in Eqs. (4)-(6). y_i is observed rainfall intensity, $\hat{y}_i$ is estimated rainfall intensity, $\bar{y}$ is average of observed rainfall intensity, and n is the number of samples.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{4}$$

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \tag{5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{6}$$

4. Results

4.1 Spectrogram extraction results according to rainfall intensity

Fig. 5 presents the STFT results and mel-spectrograms derived from acoustic signals collected under four different rainfall conditions: no rain, weak rain, moderate rain, and heavy rain.

Based on the extracted spectrograms, it is evident that signals with higher frequency components emerge as rainfall intensity increases. A comparative analysis between the spectrograms for no rain (Figs. 5(a), 5(b)) and weak rain (Figs. 5(c), 5(d)) reveals that, in the case of weak rain, irregular acoustic patterns appear below 6,000 Hz, which are absent under no-rain conditions.

Given that the mel-spectrogram provides higher resolution in the lower frequency range due to its perceptual scale, the spectral differences between no rain and weak rain are more distinctly captured in the mel-spectrogram than in the STFT. Conversely, a comparison between moderate rain (Figs. 5(e), 5(f)) and heavy rain (Figs. 5(g), 5(h)) indicates a noticeable difference in spectrograms within the high-frequency region above 15,000 Hz. While this difference is clearly identifiable in the STFT, it is less discernible in the mel-spectrogram due to its lower resolution in the high-frequency domain.

These observations highlight that the mel-spectrogram is particularly advantageous in distinguishing the acoustic characteristics between no rain and weak rain due to its enhanced resolution in low frequencies. However, it exhibits limitations in capturing differences associated with higher rainfall intensities compared to the STFT. Accordingly, it is anticipated that by jointly considering both STFT and mel-spectrogram representations, the model can achieve consistent rainfall intensity estimation performance across a broad range of rainfall conditions without being biased toward either weak or heavy rainfall

scenarios.

4.2 Performance of CNNs-based rainfall intensity estimation model

Fig. 6 illustrates the time series of rainfall intensity and the cumulative precipitation derived from rainfall acoustics collected during the evaluation period (06:00 to 15:00 on August 29, 2023).

The recorded acoustic signals were divided into 10 s segments, from which both STFT results and mel-spectrograms were extracted. These spectrograms were then input into the CNNs-based rainfall intensity estimation model trained using Events 1 to 3, as summarized in Table 1. In addition to the reference observations obtained from the optical disdrometer used during the training and validation phases, data from a co-located weighing rain gauge (i.e., OTT Pluvio^2 L) were also incorporated in Fig. 6. The acoustic data collection device, optical disdrometer, and weighing rain gauge were all installed within a 10-meter radius to minimize spatial discrepancies. Overall, the model demonstrated satisfactory performance, yielding a RMSE of 4.89 mm/h, a MAE of 2.02 mm/h, and an R^2 value of 0.75. To clearly assess the performance improvement achieved through the application of CNNs and the incorporation of mel-spectrograms, the approach proposed in our previous study (Hwang *et al*. 2024), which involved extracting four features from STFT results and applying them to an XGBoost model, was implemented using the same dataset. The results yielded an RMSE of 5.20 mm/h, an MAE of 1.97 mm/h, and an R^2 of 0.73. While a slight increase in MAE was observed, both RMSE and R^2 demonstrated improved performance, thereby confirming the effectiveness of CNNs combined with mel-spectrograms in enhancing rainfall intensity estimation.

To evaluate the accuracy of the rainfall intensity estimates derived in this study, we benchmarked our results against reported rainfall intensity estimation accuracies in prior work. As previously noted, quantitative estimation of rainfall intensity from rainfall acoustics has been attempted only rarely; accordingly, we compared against closely related, low-cost IoT sensor contexts in the video-based rainfall intensity estimation literature. Zheng *et al*. (2024) estimated rainfall intensity from surveillance camera imagery using Bayesian deep learning, utilizing data collected on the Zijingang Campus of Zhejiang University in Hangzhou, China. As the data were collected within a university setting, Zheng *et al*. (2024) deemed the research environment comparable to the present study. In that study, accuracy was assessed against 1 min rainfall observations, and the test set yielded RMSE of 2.5 to 3.6 mm/h. For our method, we averaged the 10 s rainfall intensity estimates to 1 min resolution and obtained an RMSE of approximately 3.05 mm/h, indicating performance comparable to Zheng *et al*. (2024). As another image-based reference, Yin *et al*. (2023) estimated 1 min rainfall intensity from videos captured by smartphones and surveillance cameras, reporting RMSE of 3.36 to 4.16 mm/h for smartphone imagery and 2.87 to 3.28 mm/h for surveillance cameras; our results are therefore comparable to, or slightly better than, these ranges.

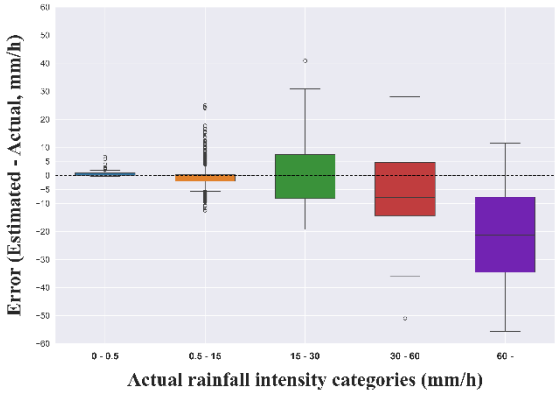


Fig. 7 Distribution of estimation errors across five rainfall intensity intervals based on optical disdrometer observations during the validation period

Beyond vision-based approaches, Pu and Liu (2024) estimated 1 min rainfall intensity using microwave links and reported RMSE of 2.07 to 2.93 mm/h depending on link type, which is similar in magnitude to the error observed in our study. Overall, across multiple representative baselines, the final performance achieved here is on par with, and in some cases superior to, prior work. Notably, the instruments used in the referenced studies, such as microwave links and surveillance cameras, are relatively high cost compared with the acoustic data collection device employed in this research, which underscores the economic advantage of the proposed approach without sacrificing accuracy.

Fig. 6(a) also presents an enlarged graph focusing on the segment where rainfall intensity exceeded 20 mm/h. As shown in the enlarged view, a tendency toward underestimation of certain peak values can be observed. In addition, slight shifts in the timing of peak rainfall intensity are also evident. This discrepancy appears to result from imperfect synchronization between the timestamps recorded during acoustic data collection and those recorded by the optical disdrometer used as reference data. This issue may be partially addressed in future studies by intentionally introducing a distinctive acoustic noise during the data acquisition process that can serve as a temporal marker, thereby enabling post-processing correction based on precise alignment with the disdrometer observations. As shown in Fig. 6(b), the cumulative precipitation estimated from acoustic data tends to be lower than those recorded by the optical disdrometer and the weighing rain gauge, indicating a consistent underestimation. This trend is considered to be caused by the underestimation of peak values during extreme rainfall events exceeding approximately 60 mm/h.

5. Discussions

This study proposed a rainfall intensity estimation method based on rainfall acoustics using CNNs. By leveraging the powerful capability of CNNs in acoustic signal processing, the study aimed to improve the accuracy

of rainfall intensity estimation and to address the paucity of quantitative approaches in prior rainfall acoustics research.

The proposed model demonstrated satisfactory performance in estimating rainfall intensity and, in particular, showed improved accuracy compared with the preceding work by Hwang *et al*. (2024). The final model achieved an RMSE of 4.89 mm/h, an MAE of 2.02 mm/h, and an R^2 of 0.75. When compared with the discrepancies observed between the two rainfall observation instruments considered in this study (i.e., optical disdrometer and weighing rain gauge), which yielded an RMSE of 6.23 mm/h, an MAE of 2.45 mm/h, and an R^2 of 0.76, the proposed rainfall acoustics-based methodology appears to exhibit high adequacy for quantitative rainfall intensity estimation.

Nevertheless, the systematic underestimation observed for extreme rainfall intensities (> 60 mm/h) highlights a remaining limitation. To further investigate the cause of the underestimation tendency observed in the preceding results, the accuracy of the acoustics-based estimated rainfall intensities was compared across different rainfall intensity levels. Fig. 7 presents a boxplot illustrating the distribution of estimation errors for different rainfall intensities, categorized into five intervals based on the observations from the optical disdrometer during the validation period. For rainfall events with intensities less than 30 mm/h, the average estimation error remained within 5 mm/h. However, for events with intensities between 30 mm/h and 60 mm/h, the average error increased to approximately 7 mm/h, and for events exceeding 60 mm/h, the average error surpassed 20 mm/h. The underestimation tendency for rainfall intensities exceeding 60 mm/h observed in Fig. 6 and Fig. 7 is presumed to result from the disparity between the high intensities characteristic of extreme rainfall events and the overall distribution of rainfall intensities represented in the training dataset.

6. Conclusions

In this study, we proposed a CNNs-based rainfall intensity estimation technique that utilizes acoustic signals recorded under natural rainfall conditions. By employing time-frequency analysis techniques, namely STFT and mel-spectrogram transform, we extracted spectrograms that serve as inputs to a dual-stream CNNs architecture. This design enabled the model to capture discriminative acoustic patterns across a wide range of rainfall intensities, taking advantage of the complementary characteristics of STFT, which offers high-frequency resolution, and mel-spectrogram, which provides enhanced sensitivity to low-frequency components.

To address the severe class imbalance commonly observed in natural rainfall datasets, particularly the predominance of low-intensity and no-rain conditions, we adopted a stratified random sampling strategy that ensured equal representation across four rainfall intensity classes, including no rain. This approach contributed to the stability of model training and improved the model's ability to generalize across varied hydrometeorological conditions.

Quantitative evaluation of the model performance against observation from an optical disdrometer during an independent rainfall event confirmed the robustness and reliability of the proposed method. Specifically, the model achieved a RMSE of 4.89 mm/h, a MAE of 2.02 mm/h, and a R^2 of 0.75. These results demonstrate that rainfall acoustics can serve as a viable and effective data source for estimating rainfall intensity with reasonable accuracy, particularly for light and moderate precipitation events. Furthermore, in comparison with representative baselines, including our previous XGBoost-based approach utilizing STFT-derived features (Hwang *et al*. 2024) and other research cases, the performance achieved in this study can be regarded as being at an appropriate level.

Overall, this research confirms the technical feasibility and practical potential of using rainfall-induced acoustic signals, in conjunction with CNNs-based spectrogram analysis, as an alternative or complementary methodology for rainfall monitoring. Given its cost-effectiveness, mobility, and ease of deployment, the proposed framework is particularly suitable for use in resource-constrained or remote environments where conventional meteorological sensors are unavailable or impractical. Furthermore, the integration of acoustic-based sensing with data-driven learning models represents a promising direction for future development in the domain of smart environmental monitoring and real-time hydrometeorological assessment.

Acknowledgments

The research described in this paper was financially supported by the National Research Foundation of Korea (NRF) (RS-2021-NR060085) funded by the Korea government (MSIT).

References

Abdul, Z.K. and Al-Talabani, A.K. (2022), "Mel frequency cepstral coefficient and its applications: A review", *IEEE Access*, **10**, 122136-122158. https://doi.org/10.1109/ACCESS.2022.3223444.

Ahlawat, H., Aggarwal, N. and Gupta, D. (2025), "Automatic Speech Recognition: A survey of deep learning techniques and approaches", *Int. J. Cognitive Comput. Eng.*, **6**, 201-237. https://doi.org/10.1016/j.ijcce.2024.12.007.

Allen, J.B. and Rabiner, L.R. (1977), "A unified approach to short-time Fourier analysis and synthesis", *Proceedings of the IEEE*, **65**(11), 1558-1564. https://doi.org/10.1109/PROC.1977.10770.

Aziz, S. and Shahnawazuddin, S. (2023), "Effective preservation of higher-frequency contents in the context of short utterance based children's speaker verification system", *Appl. Acoustics*, **209**, 109420. https://doi.org/10.1016/j.apacoust.2023.109420.

Barcaroli, E., Lupidi, A., Facheris, L., Cuccoli, F., Chen, H. and Chandra, C.V. (2018), "A validation procedure for a polarimetric weather radar signal simulator", *IEEE Transact. Geosci. Remote Sensing*, **57**(1), 609-622. https://doi.org/10.1109/TGRS.2018.2868144.

Behera, R.K., Bala, P.K. and Dhir, A. (2019), "The emerging role of cognitive computing in healthcare: a systematic literature review", *Int. J. Medical Inform.*, **129**, 154-166. https://doi.org/10.1016/j.ijmedinf.2019.04.024.

Bracci, A., Sato, K., Baldini, L., Porcù, F. and Okamoto, H. (2023), "Development of a methodology for evaluating spaceborne W-band Doppler radar by combined use of Micro Rain Radar and a disdrometer in Antarctica", *Remote Sensing of Environ.*, **294**, 113630. https://doi.org/10.1016/j.rse.2023.113630.

Buda, M., Maki, A. and Mazurowski, M.A. (2018), "A systematic study of the class imbalance problem in convolutional neural networks", *Neural Networks*, **106**, 249-259. https://doi.org/10.1016/j.neunet.2018.07.011

Byun, J., Jun, C., Kim, J., Cha, J. and Narimani, R. (2023), "Deep learning-based rainfall prediction using cloud image analysis", *IEEE Transact. Geosci. Remote Sensing*, **61**, 1-11. https://doi.org/10.1109/TGRS.2023.3263872.

Chen, M., Herrera, F. and Hwang, K. (2018), "Cognitive computing: architecture, technologies and intelligent applications", *Ieee Access*, **6**, 19774-19783. https://doi.org/10.1109/ACCESS.2018.2791469.

Chen, Y., Argentinis, J.E. and Weber, G. (2016), "IBM Watson: how cognitive computing can be applied to big data challenges in life sciences research", *Clinical Therapeutics*, **38**(4), 688-701. https://doi.org/10.1016/j.clinthera.2015.12.001.

Chumchean, S., Sharma, A. and Seed, A. (2006), "An integrated approach to error correction for real-time radar-rainfall estimation", *J. Atmos. Oceanic Technol.*, **23**(1), 67-79. https://doi.org/10.1175/JTECH1832.1.

Das, S.K., Konwar, M., Chakravarty, K. and Deshpande, S., (2017), "Raindrop size distribution of different cloud types over the Western Ghats using simultaneous measurements from Micro-Rain Radar and disdrometer", *Atmos. Res.* **186**, 72-82. https://doi.org/10.1016/j.atmosres.2016.11.003.

Diffenbaugh, N.S., Singh, D., Mankin, J.S., Horton, D.E., Swain, D.L., Touma, D., Charland, A., Liu, Y., Haugen, M., Tsiang, M. and Rajaratnam, B. (2017), "Quantifying the influence of global warming on unprecedented extreme climate events", *Proceedings of the National Academy of Sciences*, **114**(19), 4881-4886. https://doi.org/10.1073/pnas.1618082114.

Do, H.S., Kim, J., Cha, E.J., Chang, E.C., Son, S.W. and Lee, G. (2023), "Long-term change of summer mean and extreme precipitations in Korea and East Asia", *Int. J. Climatology*, **43**(7), 3476-3492. https://doi.org/10.1002/joc.8039.

El Ayadi, M., Kamel, M.S. and Karray, F. (2011), "Survey on speech emotion recognition: Features, classification schemes, and databases", *Pattern Recognition*, **44**(3), 572-587. https://doi.org/10.1016/j.patcog.2010.09.020.

Fan, J., Tang, S., Duan, H., Bi, X., Xiao, B., Li, W. and Gao, X. (2023), "Le-lwtnet: A learnable lifting wavelet convolutional neural network for heart sound abnormality detection", *IEEE Transact. Instrumentation Measure.*, **72**, 1-14. https://doi.org/10.1109/TIM.2023.3246513.

Faranda, D., Bourdin, S., Ginesta, M., Krouma, M., Messori, G., Noyelle, R., Pons, F., Yiou, P. and Messori, G. (2022), "A climate-change attribution retrospective of some impactful weather extremes of 2021", *Weather Climate Dyn. Discussions*, **2022**, 1-37. https://doi.org/10.5194/wcd-3-1311-2022.

Garcia-Ordas, M.T., Rubio-Martin, S., Benitez-Andrades, J.A., Alaiz-Moreton, H. and Garcia-Rodriguez, I. (2023), "Multispecies bird sound recognition using a fully convolutional neural network", *Appl. Intell.*, **53**(20), 23287-23300. https://doi.org/10.1007/s10489-023-04704-3.

George, S.M. and Ilyas, P.M. (2024), "A review on speech emotion recognition: A survey, recent advances, challenges, and the influence of noise", *Neurocomputing*, **568**, 127015. https://doi.org/10.1016/j.neucom.2023.127015.

Guhathakurta, P., Sreejith, O.P. and Menon, P.A. (2011), "Impact of climate change on extreme rainfall events and flood risk in India", *J. Earth Syst. Sci.*, **120**, 359-373. https://doi.org/10.1007/s12040-011-0082-5.

Guo, H., Huang, H., Sun, Y.E., Zhang, Y., Chen, S. and Huang, L. (2018), "Chaac: Real-time and fine-grained rain detection and measurement using smartphones", *IEEE Internet Things J.*, **6**(1), 997-1009. https://doi.org/10.1109/JIOT.2018.2866690.

Guo, Z., Chen, J., He, T., Wang, W., Abbas, H. and Lv, Z. (2023), "DS-CNN: Dual-stream convolutional neural networks-based heart sound classification for wearable devices", *IEEE Transact. Consumer Electronics*, **69**(4), 1186-1194. https://doi.org/10.1109/TCE.2023.3247901.

Gupta, S., Kar, A.K., Baabdullah, A. and Al-Khowaiter, W.A. (2018), "Big data with cognitive computing: A review for the future", *Int. J. Inform. Manage.*, **42**, 78-89. https://doi.org/10.1016/j.ijinfomgt.2018.06.005.

Han, S., Kang, T., Lee, J., Kim, N., Won, H., Kim, Y.H., Gong, W. and Kwak, I.Y. (2024), "A deep neural network approach to heart murmur detection using spectrogram and peak interval features", *Eng. Appl. Artificial Intell.*, **137**, 109156. https://doi.org/10.1016/j.engappai.2024.109156.

Hashim, J.H. and Hashim, Z. (2016), "Climate change, extreme weather events, and human health implications in the Asia Pacific region", *Asia Pacific J. Public Health*, **28**(2_suppl), 8S-14S. https://doi.org/10.1177/1010539515599030.

He, H. and Garcia, E.A. (2009), "Learning from imbalanced data", *IEEE Transact. Knowledge Data Eng.*, **21**(9), 1263-1284. https://doi.org/10.1109/TKDE.2008.239.

Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N. and Kingsbury, B. (2012), "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups", *IEEE Signal Processing Mag.*, **29**(6), 82-97. https://doi.org/10.1109/MSP.2012.2205597.

Howe, B.M., Miksis-Olds, J., Rehm, E., Sagen, H., Worcester, P.F. and Haralabus, G. (2019), "Observing the oceans acoustically", *Front. Marine Sci.*, **6**, 426. https://doi.org/10.3389/fmars.2019.00426.

Hwang, S., Jun, C., De Michele, C., Kim, H.J. and Lee, J. (2024), "Rainfall observation leveraging raindrop sounds acquired using waterproof enclosure: exploring optimal length of sounds for frequency analysis", *Sensors*, **24**(13), 4281. https://doi.org/10.3390/s24134281.

Kim, H.R., Moon, M., Yun, J. and Ha, K.J. (2023), "Trends and spatio-temporal variability of summer mean and extreme precipitation across South Korea for 1973–2022", *Asia-Pacific J. Atmos. Sci.*, **59**(3), 385-398. https://doi.org/10.1007/s13143-023-00323-7.

Kirchmeier-Young, M.C. and Zhang, X. (2020), "Human influence has intensified extreme precipitation in North America", *Proceedings of the National Academy of Sciences*, **117**(24), 13308-13313. https://doi.org/10.1073/pnas.1921628117.

Kloser, R.J., Ryan, T.E., Young, J.W. and Lewis, M.E. (2009), "Acoustic observations of micronekton fish on the scale of an ocean basin: potential and challenges", *ICES J. Marine Sci.*, **66**(6), 998-1006. https://doi.org/10.1093/icesjms/fsp077.

Lee, J., Byun, J., Baik, J., Jun, C. and Kim, H.J. (2023), "Estimation of raindrop size distribution and rain rate with infrared surveillance camera in dark conditions", *Atmos. Measure. Techniques*, **16**(3), 707-725. https://doi.org/10.5194/amt-16-707-2023.

Li, Z., Liu, F., Yang, W., Peng, S. and Zhou, J. (2021), "A survey of convolutional neural networks: analysis, applications, and prospects", *IEEE Transact. Neural Networks Learning Systems*, **33**(12), 6999-7019. https://doi.org/10.1109/TNNLS.2021.3084827.

Löffler-Mang, M. and Joss, J. (2000), "An optical disdrometer for measuring size and velocity of hydrometeors", *J. Atmos. Oceanic Technol.*, **17**(2), 130-139. https://doi.org/10.1175/1520-

0426(2000)017%3C0130:AODFMS%3E2.0.CO;2.

Malayath, N. and Hermansky, H. (2003), "Data-driven spectral basis functions for automatic speech recognition", *Speech Commun.*, **40**(4), 449-466. https://doi.org/10.1016/S0167-6393(02)00127-9.

Martens, B., Cabus, P., De Jongh, I. and Verhoest, N.E.C. (2013), "Merging weather radar observations with ground-based measurements of rainfall using an adaptive multiquadric surface fitting algorithm", *J. Hydrology*, **500**, 84-96. https://doi.org/10.1016/j.jhydrol.2013.07.011.

Modha, D.S., Ananthanarayanan, R., Esser, S.K., Ndirango, A., Sherbondy, A.J. and Singh, R. (2011), "Cognitive computing", *Commun. ACM*, **54**(8), 62-71. https://doi.org/10.1145/1978542.1978559.

Mu, W., Yin, B., Huang, X., Xu, J. and Du, Z. (2021), "Environmental sound classification using temporal-frequency attention based convolutional neural network", *Sci. Reports*, **11**(1), 21552. https://doi.org/10.1038/s41598-021-01045-4.

O'shea, K. and Nash, R. (2015), "An introduction to convolutional neural networks", arXiv preprint arXiv:1511.08458. https://doi.org/10.48550/arXiv.1511.08458.

O'Gorman, P.A. (2015), "Precipitation extremes under climate change", *Current Climate Change Reports*, **1**, 49-59. https://doi.org/10.1007/s40641-015-0009-3.

Ornes, S. (2018), "How does climate change influence extreme weather? Impact attribution research seeks answers", *Proceedings of the National Academy of Sciences*, **115**(33), 8232-8235. https://doi.org/10.1073/pnas.1811393115.

Park, C., Son, S.W., Kim, H., Ham, Y.G., Kim, J., Cha, D.H., Chang, E.C., Lee, G., Kug, J.S., Lee W.S., Lee Y.Y., Lee H.C. and Lim, B. (2021), "Record-breaking summer rainfall in South Korea in 2020: Synoptic characteristics and the role of large-scale circulations", *Month. Weather Rev.*, **149**(9), 3085-3100. https://doi.org/10.1175/MWR-D-21-0051.1.

Paz, I., Tchiguirinskaia, I. and Schertzer, D. (2020), "Rain gauge networks' limitations and the implications to hydrological modelling highlighted with a X-band radar", *J. Hydrology*, **583**, 124615. https://doi.org/10.1016/j.jhydrol.2020.124615.

Pijanowski, B.C., Villanueva-Rivera, L.J., Dumyahn, S.L., Farina, A., Krause, B.L., Napoletano, B.M., Gage, S.H. and Pieretti, N. (2011), "Soundscape ecology: the science of sound in the landscape", *BioScience*, **61**(3), 203-216. https://doi.org/10.1525/bio.2011.61.3.6.

Pu, K. and Liu, X. (2024), "Precipitation measurements experiment using microwave links in Eastern China from October 2020 to March 2022", *J. Hydrology*, **643**, 131922. https://doi.org/10.1016/j.jhydrol.2024.131922.

Pu, K., Liu, X., Liu, L. and Zhao, Y. (2023), "A novel adaptive rain-induced attenuation model based on clustering algorithm for commercial microwave link-based rainfall inversion", *IEEE Geosci. Remote Sensing Lett.*, **21**, 1-5. https://doi.org/10.1109/LGRS.2023.3342976.

Raimbault, M. and Dubois, D. (2005), "Urban soundscapes: Experiences and knowledge", *Cities*, **22**(5), 339-350. https://doi.org/10.1016/j.cities.2005.05.003.

Robinson, W.A. (2021), "Climate change and extreme weather: A review focusing on the continental United States", *J. Air Waste Manage. Association*, **71**(10), 1186-1209. https://doi.org/10.1080/10962247.2021.1942319.

Shin, J.Y., Park, J., Sung, K. and Kim, Y. (2025), "Recent spatiotemporal changes of subhourly extreme rainfall events in Seoul", *Appl. Sci.*, **15**(3), 1672. https://doi.org/10.3390/app15031672.

Sreedevi, A.G., Harshitha, T.N., Sugumaran, V. and Shankar, P. (2022), "Application of cognitive computing in healthcare, cybersecurity, big data and IoT: A literature review", *Inform. Processing Manage.*, **59**(2), 102888. https://doi.org/10.1016/j.ipm.2022.102888.

Stevens, S.S., Volkmann, J. and Newman, E.B. (1937), "A scale for the measurement of the psychological magnitude pitch", *J. Acoustic. Soc. America*, **8**(3), 185-190. https://doi.org/10.1121/1.1915893.

Ustubioglu, A., Ustubioglu, B. and Ulutas, G. (2023), "Mel spectrogram-based audio forgery detection using CNN", *Signal, Image Video Processing*, **17**(5), 2211-2219. https://doi.org/10.1007/s11760-022-02436-4.

Villarini, G., Mandapaka, P.V., Krajewski, W.F. and Moore, R.J. (2008), "Rainfall and sampling uncertainties: A rain gauge perspective", *J. Geophy. Res. Atmos.*, **113**(D11). https://doi.org/10.1029/2007JD009214.

Wang, D. and Chen, J. (2018), "Supervised speech separation based on deep learning: An overview", *IEEE/ACM Transact. Audio, Speech language Processing*, **26**(10), 1702-1726. https://doi.org/10.1109/TASLP.2018.2842159.

Wang, X., Wang, M., Liu, X., Glade, T., Chen, M., Xie, Y., Yuan, H. and Chen, Y. (2022), "Rainfall observation using surveillance audio", *Appl. Acoust.*, **186**, 108478. https://doi.org/10.1016/j.apacoust.2021.108478.

Xu, Y., Du, J., Dai, L.R. and Lee, C.H. (2013), "An experimental study on speech enhancement based on deep neural networks", *IEEE Signal Processing Lett.*, **21**(1), 65-68. https://doi.org/10.1109/LSP.2013.2291240.

Yilmaz, A.G., Hossain, I. and Perera, B.J.C. (2014), "Effect of climate change and variability on extreme rainfall intensity–frequency–duration relationships: a case study of Melbourne", *Hydrology Earth Syst. Sci.*, **18**(10), 4065-4076. https://doi.org/10.5194/hess-18-4065-2014.

Yin, H., Zheng, F., Duan, H.F., Savic, D. and Kapelan, Z. (2023), "Estimating rainfall intensity using an image-based deep learning model", *Engineering*, **21**, 162-174. https://doi.org/10.1016/j.eng.2021.11.021.

Zheng, F., Yin, H., Zhang, J., Duan, H.F. and Gupta, H.V. (2024), "A Bayesian deep learning approach for video-based estimation and uncertainty quantification of urban rainfall intensity", *J. Hydrology*, **640**, 131706. https://doi.org/10.1016/j.jhydrol.2024.131706.

Zhu, J. (2013), "Impact of climate change on extreme rainfall across the United States", *J. Hydrol. Eng.*, **18**(10), 1301-1309. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000725.