Examining Recognition of Occupants' Cooking Activity Based on Sound Data Using Deep Learning Models

Yuhwan Kim [1] , Chang-Ho Choi [2], Chang-Young Park [3] and Seonghyun Park [3,*]

1   Department of Architectural Engineering, Graduate School, Kwangwoon University, Seoul 01897, Republic of Korea; soc05125@kw.ac.kr
2   Department of Architectural Engineering, Kwangwoon University, Seoul 01897, Republic of Korea; choi1967@kw.ac.kr
3   Institute of Green Building and New Technology, Mirae Environment Plan, Seoul 01905, Republic of Korea; cypark@mrplan.co.kr
*   Correspondence: parksh@mrplan.co.kr; Tel.: +82-2-6459-6036

Abstract: In today's society, where people spend over 90% of their time indoors, indoor air quality (IAQ) is crucial for sustaining human life. However, as various indoor activities such as cooking generate diverse types of pollutants in indoor spaces, IAQ has emerged as a serious issue. Previous studies have employed methods such as CO_2 sensors, smart floor systems, and video-based pattern recognition to distinguish occupants' activities; however, each method has its limitations. This study delves into the classification of occupants' cooking activities using sound recognition technology. Four deep learning-based sound recognition models capable of recognizing and classifying sounds generated during cooking were presented and analyzed. Experiments were carried out using sound data collected from real kitchen environments and online data-sharing websites. Additionally, changes in performance according to the amount of collected data were observed. Among the developed models, the most efficient is found to be the convolutional neural network, which is relatively unaffected by fluctuations in the amount of sound data and consistently delivers excellent performance. In contrast, other models exhibited a tendency for reduced performance as the amount of sound data decreased. Consequently, the results of this study offer insights into the classification of cooking activities based on sound data and underscore the research potential for sound-based occupant behavior recognition classification models.

Keywords: occupant behaviors; indoor air quality; deep learning; sound recognition; cooking activities

1. Introduction

The building sector has implemented both passive and active technologies, including high-quality insulation, high-efficiency equipment, and renewable energy facilities to conserve energy. Enhancing the insulation of building facades is crucial for minimizing energy losses attributed to external heat transfer. However, the resulting increase in airtightness can have intricate implications for indoor air quality (IAQ). While improved insulation aids in preventing the infiltration of outdoor pollutants, including traffic emissions, it also has the potential to restrict natural indoor air circulation, thereby impacting IAQ. IAQ is influenced by various factors, encompassing a building's airtightness, ventilation quality, the utilization of household appliances, and occupant activities, each of which independently contributes to the overall IAQ. This, in turn, can result in health issues, including sick building syndrome (SBS) [1–3], a concern which has been widely researched in numerous studies, highlighting SBS-related risk factors [4–7]. Given that indoor air is an essential factor for sustaining human life, several studies have focused on IAQ [8,9]. In contemporary society, where people spend more than 90% of their time indoors [10], the relationship between residential spaces and IAQ has garnered significant attention. Pollution originating from substances emitted directly from indoor spaces is considered

a more significant problem than outdoor air pollution because it can pose a direct and significant threat to individuals. The primary contributors to elevated indoor particulate matter (PM) levels are cooking, combustion, and cleaning activities [11]. Previous studies have indicated that cooking activities have a considerable impact on IAQ [12–19]. Common indoor activities, such as cooking, release PM (e.g., $PM_{2.5}$ and PM_{10}) [20,21], along with various types of indoor air pollutants, including volatile organic compounds (VOCs) and semi-VOCs. The increase in indoor air pollutants can have adverse effects on occupants' cardiorespiratory systems [22,23], lung function [24–27], cardiovascular health [28,29], human brain activity [30], and cognitive performance [31].

Therefore, in this study, our primary objective is to facilitate the enhancement of IAQ by precisely identifying occupants' actions through the creation of an automated model. To achieve this, our focus is on developing a cooking activity classification model. The concept is that devices capable of autonomously monitoring and improving IAQ can potentially reduce exposure to health-hazardous indoor environments without occupants having to actively intervene. The introduction of these systems can have positive impacts on health and convenience aspects, making it an important consideration in modern architectural environments.

Various studies have been conducted to detect occupants' behavior inside buildings. Although several methods such as CO_2 sensors [32,33], smart floor systems [34,35], and video-based pattern recognition [36–39] have been applied, each method still has certain limitations. For instance, the CO_2 sensor method involves the slow diffusion of CO_2 within an indoor space where air is introduced, which can lead to ambiguous interpretations due to ventilation and the consequent alteration of CO_2 concentrations inside the building [40,41]. Smart flooring systems require highly instrumented floors, demanding thousands of sensors, which is a notable disadvantage [34,35]. Moreover, the use of video recording devices may be challenging in real environments, as it raises privacy concerns. Therefore, this study aims to overcome these limitations and accurately recognize occupants' behavior by utilizing sound recognition technology, which has a relatively low risk of privacy infringement and can distinguish cooking activities.

The concept of artificial intelligence (AI) was introduced to automate the classification of cooking activities. AI, which emerged in the 1950s in the field of computer science, is dedicated to developing machines capable of mimicking human thinking and processing data and systematizing them. This technology proves especially valuable when handling large datasets [42]. AI-based machines have extensive applications across various industries and have significantly contributed to diverse research fields, including image processing [43,44], natural language processing [45,46], disease diagnosis, medicine, and engineering sectors [47].

In conclusion, the remarkable progress in machine learning (ML) has enabled computers to excel in discerning intricate patterns, making precise predictions, and anticipating new data attributes through the analysis of training data [48]. Going further into this domain, deep learning (DL), a specific subset of ML, mimics natural human learning, enabling computers to autonomously learn from examples and directly extract valuable functions from data [44]. The diminished reliance on manually crafted functions and the decrease in human intervention stand out as distinctive features that differentiate DL from traditional ML methods.

The recent surge of interest in this domain [49] has prompted exploration into various neural network architectures, such as deep feed-forward, Siamese neural, and graph neural networks [50]. As AI, ML, and DL rapidly evolve, they are poised to play an increasingly pivotal role in shaping the future [51]. Recognizing the potential of these technologies, this study employs a DL methodology to investigate the applicability and performance of recognizing occupants' cooking activities based on sound data. Furthermore, the insights derived from this research not only enrich the existing body of knowledge but also provide valuable considerations for advancing smart environments and human–machine interactions.

2. Methods

This study collects and preprocesses sound data generated during cooking activities and then classifies and evaluates the performance of each model. Additionally, the dataset is divided into subsets of 883, 400, 200, 100, and 50 samples; finally, changes in model efficiency are determined based on the quantity of data. Through this, we aim to propose the most efficient model for automatically classifying cooking activities that significantly impact IAQ. Figure 1 illustrates the flowchart of the research process.
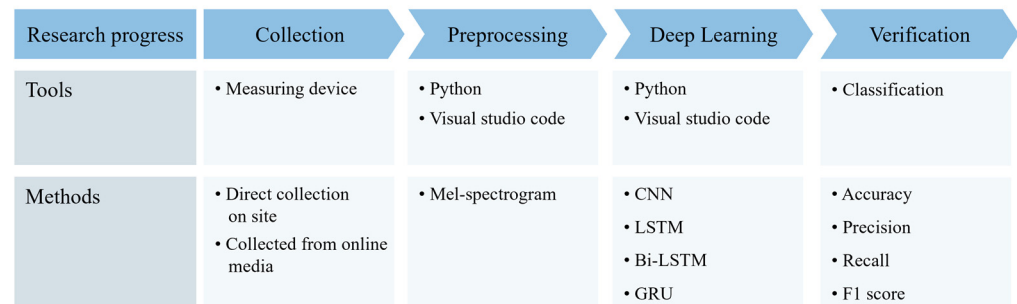
Research progress	Collection	Preprocessing	Deep Learning	Verification
Tools	• Measuring device	• Python • Visual studio code	• Python • Visual studio code	• Classification
Methods	• Direct collection on site • Collected from online media	• Mel-spectrogram	• CNN • LSTM • Bi-LSTM • GRU	• Accuracy • Precision • Recall • F1 score

Figure 1. Research process flow chart.

2.1. Sound Data Collection

To classify cooking activities, the collected sound datasets were primarily categorized as "steaming" or "boiling" and "frying or grilling". Classifying cooking activities based on sound data is inherently challenging due to their limited availability. Therefore, this study strategically prioritized the use of voice data recorded in real kitchens. In cases with insufficient data, an additional layer of diversity was introduced by incorporating data downloaded from online media. The measurements were conducted in a single-person living space with an area of 19.7 m^2 and a height of 2.4 m, as shown in Figure 2. This space is fully equipped for cooking tasks and was considered to enhance the real-world performance of the model by mimicking a real residential environment. This approach allowed the model to learn and generalize kitchen sounds occurring in specific residential spaces. Each data measurement was conducted for less than 1 min, and the measurements were taken during lunch and dinner over a period of 3 months. The room had minimal other sound sources, and background sounds were not removed. Table 1 provides an overview of the sound dataset composition. In total, 883 sound data samples were collected, of which 460 were for "boiling (steaming)" and 423 were for "frying (grilling)".
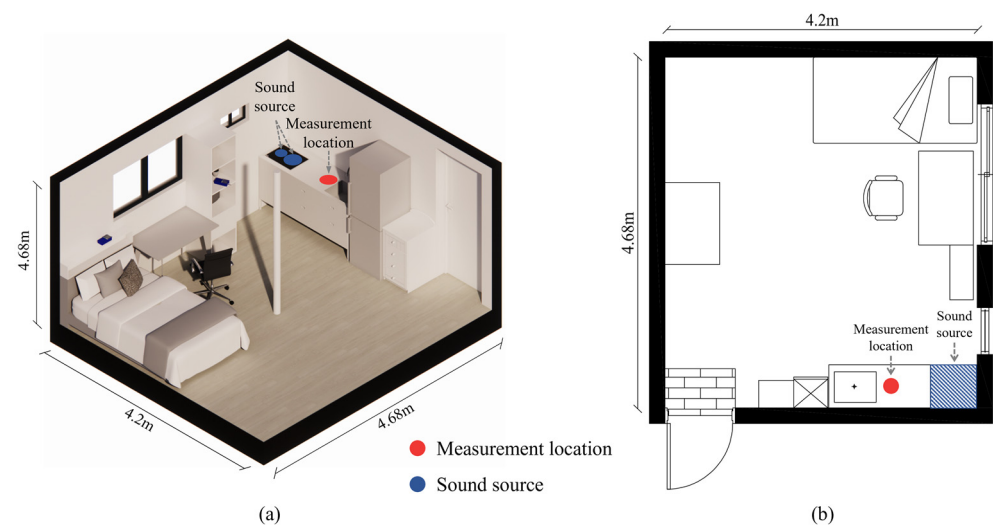

(a)   (b)

Figure 2. (**a**) Interior of a single-person household and (**b**) floor plan, depicting the measurement location.

Table 1. Organization of the collected sound datasets.

Category		Boiling (Steaming)	Frying (Grilling)
Number of data (N)		460	423
Range	File size (MB)	0.01–361.39 [Avg. 3.10]	0.10–39.57 [Avg. 3.39]
	File length (s)	2.71–221.73 [Avg. 35.89]	1.25–377.88 [Avg. 27.98]
	Sample rate (Hz)	44,100–96,000 [Avg. 48152]	44,100–96,000 [Avg. 48089]
	Amplitude (-)	0.000–0.267 [Avg. 0.018]	0.001–0.173 [Avg. 0.032]

2.2. Preprocessing and Acoustic Feature Extraction

Preprocessing sound data refers to the process of converting sound data into a format suitable for use as input in a model. This step is crucial for effectively managing sound data, which are large, complex, and high-dimensional, making them unsuitable for direct input into a model.

Initially, the sound data were loaded using the Librosa library, and the sampling rate, representing the number of data points sampled per second, was set to 16,000 Hz. This value was selected to manage the computational load while adequately capturing the high-frequency components of the sound signal. The loaded sound data were initially transformed into a Mel-Spectrogram using the Librosa library, maintaining the 16,000 Hz sampling rate. The window size for fast Fourier transform (FFT), referred to as "n_fft", was set to 2048, and the "hop_length", representing the degree of window overlap, was adjusted to 512. Furthermore, "n_mels", indicating the number of Mel-Filters, was set to 64. These settings collectively converted the sound data into a format suitable for classification models. Additionally, when the spectrogram's length exceeded the "max_length", certain data were trimmed, while padding was added when the length was insufficient to maintain a consistent data length. Through this process, the sound data were converted into a format suitable for use as input to the model, as shown in Figure 3, making acoustic feature extraction straightforward.
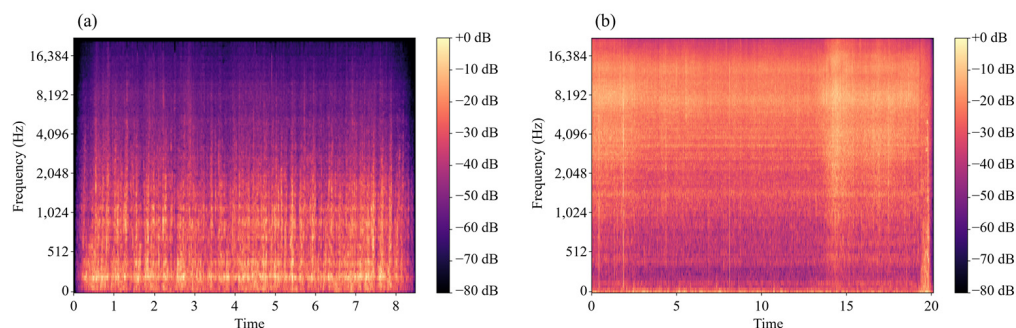
Figure 3. Mel-Spectrogram of the audio signal for (**a**) "Boiling" and (**b**) "Frying".

3. Classification Model Structure and Training

To identify occupant behavior, this study established four DL-based sound recognition models: convolutional neural network (CNN), long short-term memory (LSTM), bidirectional long short-term memory (Bi-LSTM), and gated recurrent unit (GRU). Each model was employed to learn and classify sound data, which were visualized as a Mel-Spectrogram.

3.1. Convolutional Neural Network (CNN)

The model comprises a total of three convolutional layers, each utilizing a 3×3 filter for extracting data features. The first convolutional layer incorporates 64 filters with the rectified linear unit (ReLU) activation function, enhancing the model's capacity to learn complex data relationships and features while introducing non-linearities.

Next, the MaxPooling layer downsamples data through a 2×2 pooling window. In this process, input images or data are subdivided into small grid-patterned areas, with the largest value within each area selected as a representative value. This procedure reduces

the size of images or data, improves computational efficiency, preserves crucial information, and enhances data abstraction.

The second convolutional layer utilizes 128 filters with a 3×3 dimension, taking the output from the first layer as input. Once again, the ReLU activation function is applied to introduce non-linearities and extract data features. The subsequent MaxPooling layer further reduces data dimensions using a 2×2 pooling window, effectively minimizing data size, retaining essential information, and simplifying the model.

The third convolutional layer employs 256 filters with a 3×3 dimension to process input data. The ReLU activation function is then applied to further abstract data features. The third MaxPooling layer again reduces data size through a 2×2 pooling window.

Following this, the data are flattened into a one-dimensional form via the flatten layer, preparing them for transmission to the fully connected layer. The first fully connected layer comprises 512 neurons, introducing non-linearity through the ReLU activation function. The second fully connected layer encompasses 256 neurons and also utilizes the ReLU activation function.

In the middle layer, 128 filters are applied using the same kernel size and activation function. The final convolution layer employs 256 filters with a 4×3 kernel size. Following the convolution layer, the flatten layer converts the data from the convolution to the fully connected layer, which includes 512 and 256 neurons. The ReLU activation function and dropout regularization are applied to prevent overfitting.

Finally, the output layer utilizes the SoftMax activation function to classify data into the "boiling" or "frying" categories. The SoftMax function transforms the model's output into a probability distribution for each class, effectively conducting predictions and classification tasks by converting input values into probability values between 0 and 1.

3.2. Long Short-Term Memory (LSTM)

LSTM is a technology engineered to address the limitations of the existing recurrent neural network (RNN), which involves a directed cycle in which past data can influence future results. RNN uses gradient descent to adjust parameters for minimizing the cost function, but it encounters the challenge of vanishing gradients when processing long sequences, making it difficult to capture long-term dependencies in the data [52]. LSTM overcomes these drawbacks of RNN and provides a structure that adeptly learns long-term dependencies from sequential data [53]. The fundamental structure of LSTM is depicted in Figure 4.
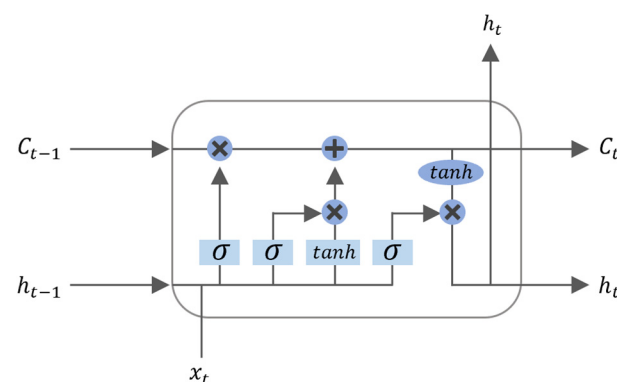

Figure 4. Structure of long short-term memory (LSTM) [54].

LSTM has proven to be highly effective in handling diverse, sequence data types, including time series, text, and sound data. Its significance becomes apparent in the context of the sound-based, cooking activities classification model examined in this study.

The initial LSTM layer comprises 128 neurons, actively learning diverse features from the input sequence data and extracting patterns within the sequence. Notably, as this

layer is configured with "return_sequence=True", it plays a crucial role in transmitting information up to the current point to the subsequent layer.

3.3. Bidirectional Long Short-Term Memory (Bi-LSTM)

Bi-LSTM, an extended iteration of LSTM, proficiently manages sequence data bidirectionally. This capability empowers the model to leverage both past and future information when making decisions in the present [55]. In contrast to LSTM, Bi-LSTM incorporates both forward and backward directions, enabling it to discern meaningful patterns by simultaneously integrating information from the past and the future into the present. The structure of Bi-LSTM is delineated in Figure 5.
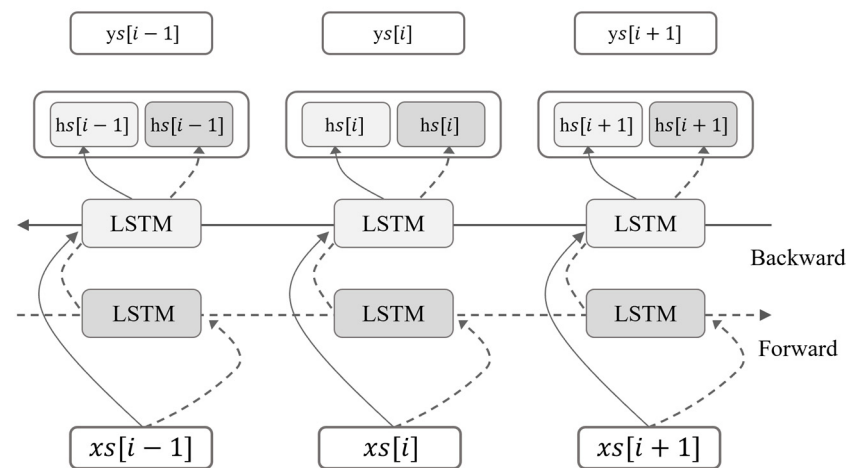

Figure 5. Structure of bidirectional long short-term memory (Bi-LSTM).

This model processes an input sequence bidirectionally for several reasons. First, it enables the model to extract meaningful patterns bidirectionally by incorporating information from the future into the present, while also facilitating the transfer of information from the past to the present. Second, it allows for a more effective consideration of the long-term dependency of data through two-way processing.

Concerning the hierarchical structure of Bi-LSTM, as detailed in Section 3.2, the first LSTM layer comprises 128 neurons. By setting "return_sequence=True", the information up to the current point can be transmitted to the next layer. The second LSTM layer takes the output of the first layer as input and features 64 neurons. Through this two-way hierarchical structure, meaningful sequences can be learned by considering the future information of the data.

3.4. Gated Recurrent Unit (GRU)

GRU represents a type of RNN with a structure simpler than that of LSTM. Prior studies have demonstrated that, despite requiring fewer learning parameters, GRU performs similarly to LSTM [56]. Similar to LSTM, GRU is effective in learning long-term dependencies. Figure 6 illustrates the structure of GRU.

The initial layer of the GRU comprises 256 neurons, with "return_sequence=True" facilitating the transfer of information to the subsequent layer. The second layer features 128 neurons, and, similarly, "return_sequence=True" is set to retain sequence information. The third layer encompasses 64 neurons, and these layers are designed to process sequence data. Introducing batch normalization or dropout layers in-between can enhance the model's stability.
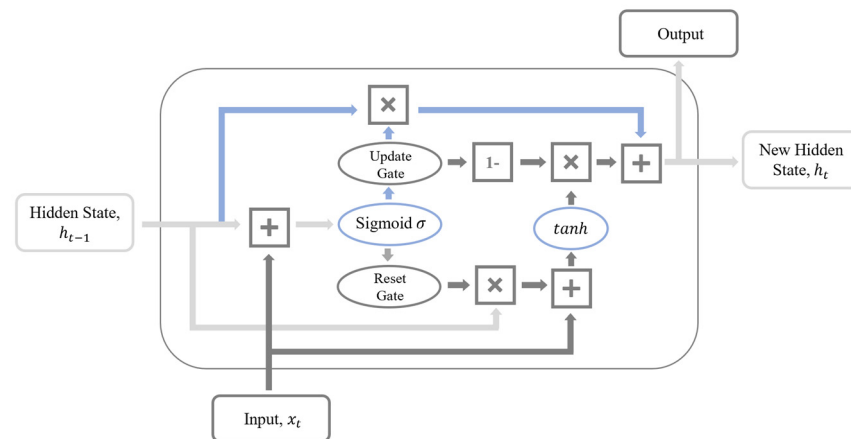
Figure 6. Structure of gated recurrent unit (GRU).

4. Experimental Results and Performance Comparison

This section offers an in-depth discussion of the experimental results and performance comparisons among various models. This study meticulously examined the learning and performance evaluation of each model, and the detailed results are presented below.

4.1. Performance Derivation Method for Each Model

To assess the performance of each model based on experimental results, we scrutinized their performance using four evaluation indicators: accuracy, precision, recall, and F1 score. Figure 7 illustrates the performance evaluation indicators employed in the model evaluation stage. Calculating these four performance evaluation indicators necessitated the use of a confusion matrix. This matrix represented the distribution of observations based on the actual value of the dependent variable and the value predicted by the model for the dependent variable. In this study, a model was constructed to classify the occupants' cooking behavior using sound data. Therefore, we examined the use of classification methods for performance evaluation.
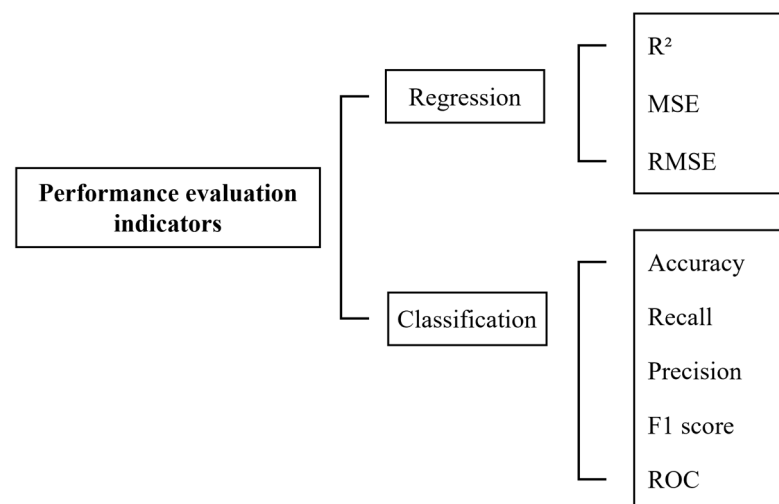

Figure 7. Performance evaluation metrics for deep learning classification models.

The confusion matrix is represented as a 2 × 2 matrix in the context of classification, as depicted in Figure 8. In this scenario, the potential values for y are 0 and 1, reflecting the binary nature of the classification problem. The main components of the confusion matrix are as follows: TP represents instances when the model correctly predicts the positive class, with both the actual and predicted values being positive. TN indicates instances when the model correctly predicts the negative class, with both the actual and predicted values being

negative. FP denotes instances when the model incorrectly predicts the positive class when the actual value is negative, but the model predicts it to be positive. FN signifies instances when the model incorrectly predicts the negative class when the actual value is positive, but the model predicts it to be negative.

Predicted Class

Figure 8. Confusion matrix.

Accuracy serves as the evaluation index that most intuitively indicates model performance. It is calculated by dividing the number of correctly predicted data by the total number of data, as shown in Equation (1).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \tag{1}$$

Precision is the proportion of items that the model classifies as true that are actually true. It can be expressed as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

Recall is an indicator that represents the proportion of actual positive observations among those predicted by the model.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

The F1 score serves as an indicator representing the harmonic mean of precision and recall. Among various metrics, the harmonic mean is generally preferred as the most balanced way to calculate the mean while considering both precision and recall. The F1 score can be assessed on a scale from 0 to 1. A low F1 Score implies that both precision and recall are low, indicating a scenario in which both metrics have decreased. A low precision score suggests that the machine learning model is making numerous incorrect predictions, while a low recall indicates that the model is failing to capture some events which should have been detected.
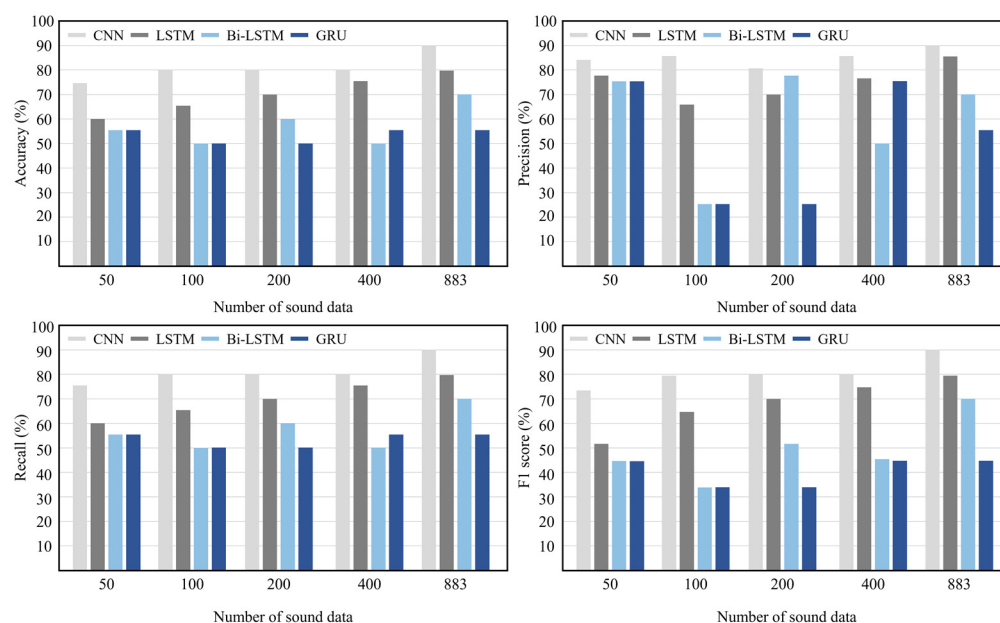
$$\text{F1 score} = 2 \times \frac{\text{Precision} + \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

4.2. Model Performance Comparison Based on the Number of Sound Data Samples

The volume of sound data significantly influences model performance. In this study, a total of 20 models were created using 883, 400, 200, 100, and 50 sound data samples, as outlined in Table 2. Each model received input from 20 data samples to evaluate the results. The changes in model performance based on the number of sound data samples in Table 2 and Figure 9 are summarized as follows. The results indicated that CNN outperforms the other models in classifying cooking activities based on 883 sound data samples. Additionally, LSTM showed the second-highest performance, followed by Bi-LSTM and GRU.

Table 2. Comparison of model performance based on the number of collected data samples.

Number of Data	Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
883	CNN	90	90	90	90
	LSTM	80	86	80	79
	Bi-LSTM	70	70	70	70
	GRU	55	76	55	44
400	CNN	80	86	80	80
	LSTM	75	77	75	74
	Bi-LSTM	50	50	50	45
	GRU	55	76	55	44
200	CNN	80	81	80	80
	LSTM	70	70	70	70
	Bi-LSTM	60	78	60	52
	GRU	50	25	50	33
100	CNN	80	86	80	79
	LSTM	65	66	65	64
	Bi-LSTM	50	25	50	33
	GRU	50	25	50	33
50	CNN	75	83	75	73
	LSTM	60	78	60	52
	Bi-LSTM	55	76	55	44
	GRU	55	76	55	44

Figure 9. Performance of four deep learning models based on the evaluation indicators.

With a reduced number (400) of sound data samples, the CNN model maintained a high performance, with over 90% accuracy, precision, recall, and F1 scores. Thus, the CNN model demonstrated stable performance without being significantly impacted by the amount of sound data. The performance of the LSTM model slightly decreased with fewer data samples but was still high. In contrast, the Bi-LSTM and GRU models showed relatively lower performances.

When the number of sound data samples was 200, the CNN model maintained an excellent performance. The LSTM and Bi-LSTM models demonstrated a moderate performance, while the GRU model still exhibited a relatively low performance. Even when the number of sound data samples was as low as 100 or 50, the CNN model sustained an excellent performance. However, the performance of the other models decreased significantly. Thus, models other than CNN exhibited lower performances with fewer sound data samples.

4.3. Strategies for Improving IAQ and Research Limitations

This study commenced with the development of a model that leverages sound data from residential spaces to automatically identify cooking activities. Our objective was to utilize sound recognition technology for developing a deep learning model capable of distinguishing a broad range of sounds, arising not only from cooking activities but also from the occupants' daily routines. Future research aims to employ this model to devise strategies for the automatic management of IAQ and expand the model for accurately identifying a variety of resident activities. This model is meant to be integrated with measurements of specific, particle-sized concentrations of indoor particulate matter; this will enable the improvement of IAQ—by following action guidelines provided by an automated system—with no direct awareness of the residents. However, the cooking behavior classification model is limited by the lack of a direct correlation with IAQ; it might inadvertently recognize sounds from peripheral devices, such as televisions, while classifying activities based on voice data.

5. Conclusions

This study aims to classify cooking behaviors, which are a main source of indoor particle generation, as a method for improving IAQ based on sound recognition technology. This technology not only presents a low risk to privacy but also effectively distinguishes different cooking activities. Therefore, we compared and evaluated the performances of CNN, LSTM, Bi-LSTM, and GRU, which were the DL models used for classifying cooking activities based on sound data.

The CNN model, based on sound data, exhibited excellent performance. With 883 sound data samples, the model achieved a high performance, exceeding 90% on all the evaluation indicators. The LSTM model demonstrated the second-best performance, while the Bi-LSTM and GRU models showed relatively lower performances.

Even with a reduced number of sound data samples, the CNN model maintained a stable performance. With as few as 400 sound data samples, it sustained performance levels exceeding 80% across all the evaluation criteria and demonstrated a relatively high performance even when the number of sound data samples was further decreased. In contrast, the performances of the other models exhibited a tendency to decline as the number of sound data samples decreased.

As a result, this study confirms that the CNN model consistently demonstrates exceptional performance and is relatively resilient to variations in the size of sound data. This discovery implies that the CNN model should be the preferred choice in developing a sound-based cooking activity classification model. Thoughtful consideration is imperative when fine-tuning the number of sound data samples and selecting a model.

To conclude, this study verifies the potential for further exploration in classifying cooking activities based on sound data. Future research will aim to improve accuracy by focusing on the development of automated models not only for cooking activities but also by collecting sound data from a wide range of activities occurring in the occupants' daily lives.

Abbreviations

AI	Artificial intelligence
Bi-LSTM	Bidirectional long short-term memory
CNN	Convolutional neural network
DL	Deep learning
FFT	Fast Fourier transform
FLAC	Free lossless audio codec
FN	False negative
FP	False positive
GRU	Gated recurrent unit
IAQ	Indoor air quality
LSTM	Long short-term memory
ML	Machine learning
PM	Particulate matter
ReLU	Rectified linear unit
RNN	Recurrent neural network
SBS	Sick building syndrome
TN	True negative
TP	True positive
VOC	Volatile organic compound

References

1. Azuma, K.; Uchiyama, I.; Katoh, T.; Ogata, H.; Arashidani, K.; Kunugita, N. Prevalence and characteristics of chemical intolerance: A Japanese population-based study. *Arch. Environ. Occup. Health* **2015**, *70*, 341–353. [CrossRef] [PubMed]
2. Hojo, S.; Mizukoshi, A.; Azuma, K.; Okumura, J.; Ishikawa, S.; Miyata, M.; Mizuki, M.; Ogura, H.; Sakabe, K. Survey on changes in subjective symptoms, onset/trigger factors, allergic diseases, and chemical exposures in the past decade of Japanese patients with multiple chemical sensitivity. *Int. J. Hyg. Environ. Health* **2018**, *221*, 1085–1096. [CrossRef] [PubMed]
3. Belachew, H.; Assefa, Y.; Guyasa, G.; Azanaw, J.; Adane, T.; Dagne, H.; Gizaw, Z. Sick building syndrome and associated risk factors among the population of Gondar town, northwest Ethiopia. *Environ. Health Prev. Med.* **2018**, *23*, 54. [CrossRef] [PubMed]
4. Indoor Air Facts No. 4 Sick Building Syndrome. 2020. Available online: https://www.epa.gov/sites/default/files/2014-08/documents/sick_building_factsheet.pdf (accessed on 3 August 2020).
5. Redlich, C.A.; Sparer, J.; Cullen, M.R. Sick-building syndrome. *Lancet* **1997**, *349*, 1013–1016. [CrossRef]
6. Burge, P.S. Sick building syndrome. *Occup. Environ. Med.* **2004**, *61*, 185–190. [CrossRef]
7. Hodgson, M.J. Sick Building Syndrome. In *Encyclopedia of Occupational Health and Safety*; International Labor Organization: Geneva, Switzerland, 2011.
8. Saidin, H.; Razak, A.A.; Mohamad, M.F.; UI-Saufie, A.Z.; Zaki, S.A.; Othman, N. Hazard evaluation of indoor air quality in bank offices. *Buildings* **2023**, *13*, 798. [CrossRef]
9. Szczepanik-Scislo, N.; Scislo, L. Dynamic real-time measurements and a comparison of gas and wood furnaces in a dual-fuel heating system in order to evaluate the occupants' safety and indoor air quality. *Buildings* **2023**, *13*, 2125. [CrossRef]
10. Klepeis, N.E.; Nelson, W.C.; Ott, W.R.; Robinson, J.P.; Tsang, A.M.; Switzer, P.; Behar, J.V.; Hern, S.C.; Engelmann, W.H. The National Human Activity Pattern Survey (NHAPS): A resource for assessing exposure to environmental pollutants. *J. Expo. Anal. Environ. Epidemiol.* **2001**, *11*, 231–252. [CrossRef]
11. Tran, V.V.; Park, D.; Lee, Y.-C. Indoor air pollution, related human diseases, and recent trends in the control and improvement of indoor air quality. *Int. J. Environ. Res. Publ. Health* **2020**, *17*, 2927. [CrossRef]
12. Cheng, S.; Wang, G.; Lang, J.; Wen, W.; Wang, X.; Yao, S. Characterization of volatile organic compounds from different cooking emissions. *Atmos. Environ.* **2016**, *145*, 299–307. [CrossRef]
13. Lee, Y.Y.; Park, H.; Seo, Y.; Yun, J.; Kwon, J.; Park, K.W.; Han, S.B.; Oh, K.C.; Jeon, J.M.; Cho, K.S. Emission characteristics of particulate matter, odors, and volatile organic compounds from the grilling of pork. *Environ. Res.* **2020**, *183*, 109162. [CrossRef]
14. Chao, C.Y.; Cheng, E.C. Source apportionment of indoor $PM_{2.5}$ and PM_{10} in homes. *Indoor Built Environ.* **2002**, *11*, 27–37. [CrossRef]
15. Liu, Q.; Son, Y.J.; Li, L.; Wood, N.; Senerat, A.M.; Pantelic, J. Healthy home interventions: Distribution of $PM_{2.5}$ emitted during cooking in residential settings. *Build. Environ.* **2002**, *207*, 108448. [CrossRef]
16. See, S.W.; Balasubramanian, R. Risk assessment of exposure to indoor aerosols associated with Chinese cooking. *Environ. Res.* **2006**, *102*, 197–204. [CrossRef]
17. See, S.W.; Karthikeyan, S.; Balasubramanian, R. Health risk assessment of occupational exposure to particulate-phase polycyclic aromatic hydrocarbons associated with Chinese, Malay and Indian cooking. *J. Environ. Monit.* **2006**, *8*, 369–376. [CrossRef]
18. See, S.W.; Balasubramanian, R. Chemical characteristics of fine particles emitted from different gas cooking methods. *Atmos. Environ.* **2008**, *42*, 8852–8862. [CrossRef]

19. Li, Y.C.; Shu, M.; Ho, S.S.H.; Wang, C.; Cao, J.J.; Wang, G.H.; Wang, X.X.; Wang, K.; Zhao, X.Q. Characteristics of $PM_{2.5}$ emitted from different cooking activities in China. *Atmos. Res.* **2015**, *166*, 83–91. [CrossRef]

20. Xiang, J.; Hao, J.; Austin, E.; Shirai, J.; Seto, E. Residential cooking-related $PM_{2.5}$: Spatial-temporal variations under various intervention scenarios. *Build. Environ.* **2021**, *201*, 108002. [CrossRef]

21. Koistinen, K.J.; Edwards, R.D.; Mathys, P.; Ruuskanen, J.; Künzli, N.; Jantunen, M.J. Sources of fine particulate matter in personal exposures and residential indoor, residential outdoor and workplace microenvironments in the Helsinki phase of the EXPOLIS study. *Scand. J. Work Environ. Health* **2004**, *30* (Suppl. 2), 36–46.

22. Park, S.K.; O'Neill, M.S.; Vokonas, P.S.; Sparrow, D.; Schwartz, J. Effects of air pollution on heart rate variability: The VA normative aging study. *Environ. Health Perspect.* **2005**, *113*, 304–309. [CrossRef]

23. Rajagopalan, S.; Brauer, M.; Bhatnagar, A.; Bhatt, D.L.; Brook, J.R.; Huang, W.; Münzel, T.; Newby, D.; Siegel, J.; Brook, R.D.; et al. Personal-level protective actions against particulate matter air pollution exposure: A scientific statement from the American Heart Association. *Circulation* **2020**, *142*, e411–e431. [CrossRef]

24. Du, B.; Gao, J.; Chen, J.; Stevanovic, S.; Ristovski, Z.; Wang, L.; Wang, L. Particle exposure level and potential health risks of domestic Chinese cooking. *Build. Environ.* **2017**, *123*, 564–574. [CrossRef]

25. Agrawal, S. Effect of indoor air pollution from biomass and solid fuel combustion on prevalence of self-reported asthma among adult men and women in India: Findings from a nationwide large-scale cross-sectional survey. *J. Asthma* **2012**, *49*, 355–365. [CrossRef]

26. Stabile, L.; Fuoco, F.C.; Marini, S.; Buonanno, G. Effects of the exposure to indoor cooking-generated particles on nitric oxide exhaled by women. *Atmos. Environ.* **2015**, *103*, 238–246. [CrossRef]

27. Wong, G.W.K.; Brunekreef, B.; Ellwood, P.; Anderson, H.R.; Asher, M.I.; Crane, J.; Lai, C.K.W. Cooking fuels and prevalence of asthma: A global analysis of phase three of the international study of asthma and allergies in childhood (ISAAC). *Lancet Respir. Med.* **2013**, *1*, 386–394. [CrossRef]

28. Lewtas, J. Air pollution combustion emissions: Characterization of causative agents and mechanisms associated with cancer, reproductive, and cardiovascular effects. *Mutat. Res. Rev. Mutat. Res.* **2007**, *636*, 95–133. [CrossRef]

29. Tan, Y.Q.; Rashid, S.K.A.; Pan, W.C.; Chen, Y.C.; Yu, L.E.; Seow, W.J. Association between microenvironment air quality and cardiovascular health outcomes. *Sci. Total Environ.* **2020**, *716*, 137027. [CrossRef]

30. Naseri, M.; Jouzizadeh, M.; Tabesh, M.; Malekipirbazari, M.; Gabdrashova, R.; Nurzhan, S.; Farrokhi, H.; Khanbabaie, R.; Mehri-Dehnavi, H.; Bekezhankyzy, Z.; et al. The impact of frying aerosol on human brain activity. *Neurotoxicology* **2019**, *74*, 149–161. [CrossRef]

31. Zhang, X.; Chen, X.; Zhang, X. The impact of exposure to air pollution on cognitive performance. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 9193–9197. [CrossRef]

32. Candanedo, L.M.; Feldheim, V. Accurate occupancy detection of an office room from light, temperature, humidity and CO_2 measurements using statistical learning models. *Energy Build.* **2016**, *112*, 28–39. [CrossRef]

33. Jiang, C.; Masood, M.K.; Soh, Y.C.; Li, H. Indoor occupancy estimation from carbon dioxide concentration. *Energy Build.* **2016**, *131*, 132–141. [CrossRef]

34. Serra, R.; Di Croce, P.; Peres, R.; Knittel, D. Human step detection from a piezoelectric polymer floor sensor using normalization algorithms. In Proceedings of the Sensors, 2014 IEEE, Valencia, Spain, 2–5 November 2014; pp. 1169–1172. [CrossRef]

35. Serra, R.; Knittel, D.; Di Croce, P.; Peres, R. Activity recognition with smart polymer floor sensor: Application to human footstep recognition. *IEEE Sens. J.* **2016**, *16*, 5757–5775. [CrossRef]

36. Wang, L.; Tan, T.; Ning, H.; Hu, W. Silhouette analysis-based gait recognition for human identification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 1505–1518. [CrossRef]

37. Gautam, K.S.; Thangavel, S.K. Video analytics-based intelligent surveillance system for smart buildings. *Soft Comput.* **2019**, *23*, 2813–2837. [CrossRef]

38. Sugandi, B.; Kim, H.; Tan, J.K.; Ishikawa, S. Real time tracking and identification of moving persons by using a camera in outdoor environment. *Int. J. Innov. Comput. Inf. Control* **2009**, *5*, 1179–1188. Available online: https://kyutech.repo.nii.ac.jp/record/5063/files/ijicic1179_1188.pdf (accessed on 2 November 2023).

39. Zheng, W.-S.; Gong, S.; Xiang, T. Person re-identification by probabilistic relative distance comparison. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; pp. 649–656. [CrossRef]

40. Tekler, Z.D.; Low, R.; Gunay, B.; Andersen, R.K.; Blessing, L. A scalable Bluetooth low energy approach to identify occupancy patterns and profiles in office spaces. *Build. Environ.* **2020**, *171*, 106681. [CrossRef]

41. Weekly, K.; Bekiaris-Liberis, N.; Jin, M.; Bayen, A.M. Modeling and estimation of the humans' effect on the CO_2 dynamics inside a conference room. *IEEE Trans. Contr. Syst. Technol.* **2015**, *23*, 1770–1781. [CrossRef]

42. Saranya, A.; Subhashini, R. A systematic review of explainable artificial intelligence models and applications: Recent developments and future trends. *Decis. Anal. J.* **2023**, *7*, 100230. [CrossRef]

43. Izonin, I.; Tkachenko, R.; Peleshko, D.; Rak, T.; Batyuk, D. Learning-based image super-resolution using weight coefficients of synaptic connections. In Proceedings of the 2015 Xth International Scientific and Technical Conference "Computer Sciences and Information Technologies" (CSIT), Lviv, Ukraine, 14–17 September 2015; pp. 25–29. [CrossRef]

44. Shen, D.; Wu, G.; Suk, H.-I. Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* **2017**, *19*, 221–248. [CrossRef]

45. Barakhnin, V.; Duisenbayeva, A.; Kozhemyakina, O.Y.; Yergaliyev, Y.; Muhamedyev, R. The automatic processing of the texts in natural language. Some bibliometric indicators of the current state of this research area. *J. Phys. Conf. Ser.* **2018**, *1117*, 012001. [CrossRef]
46. Hirschberg, J.; Manning, C.D. Advances in natural language processing. *Science* **2015**, *349*, 261–266. [CrossRef]
47. Exarchos, K.P.; Aggelopoulou, A.; Oikonomou, A.; Biniskou, T.; Beli, V.; Antoniadou, E.; Kostikas, K. Review of artificial intelligence techniques in chronic obstructive lung disease. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 2331–2338. [CrossRef]
48. Shi, F.; Wang, J.; Shi, J.; Wu, Z.; Wang, Q.; Tang, Z.; He, K.; Shi, Y.; Shen, D. Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19. *IEEE Rev. Biomed. Eng.* **2021**, *14*, 4–15. [CrossRef]
49. Goutam, B.; Hashmi, M.F.; Geem, Z.W.; Bokde, N.D. A comprehensive review of deep learning strategies in retinal disease diagnosis using fundus images. *IEEE Access* **2022**, *10*, 57796–57823. [CrossRef]
50. Mukhamediev, R.I.; Popova, Y.; Kuchin, Y.; Zaitseva, E.; Kalimoldayev, A.; Symagulov, A.; Levashenko, V.; Abdoldina, F.; Gopejenko, V.; Yakunin, K.; et al. Review of artificial intelligence and machine learning technologies: Classification, restrictions, opportunities and challenges. *Mathematics* **2022**, *10*, 2552. [CrossRef]
51. Andras, I.; Mazzone, E.; Van Leeuwen, F.W.B.; De Naeyer, G.; Van Oosterom, M.N.; Beato, S.; Buckle, T.; O'Sullivan, S.; Van Leeuwen, P.J.; Beulens, A.; et al. Artificial intelligence and robotics: A combination that is changing the operating room. *World J. Urol.* **2020**, *38*, 2359–2366. [CrossRef]
52. Hong, J.-K.; Lee, Y.K. LSTM-based anomal motor vibration detection. In Proceedings of the 2021 21st ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter), Ho Chi Minh City, Vietnam, 28–30 January 2021; pp. 98–99. [CrossRef]
53. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
54. Liu, S.; Zhang, Y.; Xu, T.; Wu, J. Short-term power prediction of wind turbine applying machine learning and digital filter. *Appl. Sci.* **2023**, *13*, 1751. [CrossRef]
55. Schuster, M.; Paliwal, K.K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. [CrossRef]
56. Gao, S.; Huang, Y.; Zhang, S.; Han, J.; Wang, G.; Zhang, M.; Lin, Q. Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. *J. Hydrol.* **2020**, *589*, 125188. [CrossRef]