

1D-CNN-LSTM Hybrid-Model-Based Pet Behavior Recognition through Wearable Sensor Data Augmentation

Hyungju Kim and Namme Moon*

Abstract

The number of healthcare products available for pets has increased in recent times, which has prompted active research into wearable devices for pets. However, the data collected through such devices are limited by outliers and missing values owing to the anomalous and irregular characteristics of pets. Hence, we propose pet behavior recognition based on a hybrid one-dimensional convolutional neural network (CNN) and long short-term memory (LSTM) model using pet wearable devices. An Arduino-based pet wearable device was first fabricated to collect data for behavior recognition, where gyroscope and accelerometer values were collected using the device. Then, data augmentation was performed after replacing any missing values and outliers via preprocessing. At this time, the behaviors were classified into five types. To prevent bias from specific actions in the data augmentation, the number of datasets was compared and balanced, and CNN-LSTM-based deep learning was performed. The five subdivided behaviors and overall performance were then evaluated, and the overall accuracy of behavior recognition was found to be about 88.76%.

Keywords

Behavior Recognition, CNN-LSTM, Data Augmentation, Deep Learning, Sensor data, Wearable Device

1. Introduction

The size of the pet care market has rapidly grown in recent times as more people tend to think of their pets as family members. Moreover, the number of consumers adopting pets has increased owing to prolonged quarantine at home and telecommuting because of the COVID-19 pandemic [1,2]. Accordingly, the number of users interested in the healthcare concerns of pets as well as related industries have increased. Various devices have been launched in the pet healthcare service segment, such as monitoring systems with cameras and wearable devices for pet activity measurement. Among these, wearable devices constitute a part of the healthcare segment that uses the Internet of Things (IoT), on which many studies have been reported [3,4]. Since these devices should not be uncomfortable or heavy for the pets to wear, miniaturized and lightweight products are being developed to enable continuous detection during daily activities without encumbrance from temporal and spatial limitations [5,6]. These devices can also be used for the daily health management of pets or observation of pet behaviors.

Data collected from wearable devices may often have missing values or outliers because of poor

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuscript received January 21, 2022; first revision May 10, 2022; accepted August 20, 2022.

* Corresponding Author: Namme Moon (namme.moon@gmail.com)

Dept. of Computer Science and Engineering, Hoseo University, Asan, Korea (kimhyungju01@gmail.com, namme.moon@gmail.com)

communication. Moreover, since the wearing direction or angle of a wearable device is not always guaranteed to be constant, various types of noise may be generated. The data obtained from such sensors are often augmented to achieve high accuracy. Data augmentation is typically used to process the missing values or outliers, where noise is artificially added to the sensor data. At this time, statistical methods of augmentation are preferred over deep-learning-based methods for behavior recognition because of the irregular and anomalous characteristics of pets.

Accordingly, this work proposes pet behavior recognition based on a hybrid one-dimensional (1D) convolutional neural network (CNN) and long short-term memory (LSTM) model through wearable sensor data augmentation. Most devices meant for pet healthcare purposes only show simple activities of the pets, and the range of data available to users is extremely limited or sometimes missing. Thus, an Arduino-based wearable device was manufactured for sensor data collection, and three-axis gyroscope as well as three-axis accelerometer data were collected. After preprocessing and augmentation, the acquired data were applied as inputs to a deep-learning-based model, from which we expect to recognize five behaviors, namely walking, standing, sitting, running, and lying.

The remainder of this manuscript is structured as follows: Section 2 describes research related to existing wearable devices and behavior recognition. Section 3 describes the proposed pet behavior recognition. Section 4 describes the experimental results from the proposed method. Section 5 discusses the conclusions of this study and future research directions.

2. Related Works

2.1 Sensor Data Augmentation

Activity recognition using sensor data requires a large dataset for deep learning. Thus, the performance of a model is often degraded when the training data are sparse [7]. Data augmentation is the most common method of solving this problem [8], which involves adding new data or increasing the amount of available data. The method used to augment the data in this study is the generative adversarial network (GAN), which uses a generative model to approximate the data distribution and a differential model to discriminate true from false [9]. This approach is widely used in image data augmentation studies [10,11]. Another method of augmentation involves artificially generating data using mathematical calculations, and the representative techniques for such calculations include jittering, rotation, permutation, combination, scaling, and time warping, among others [12]. Examples of data augmentation results by these methods are shown in Fig. 1. Considering the anomalous and irregular characteristics of pets, methods involving device rotation and noise generation are considered suitable for this study, so a mathematical approach was used herein.

2.2 Behavior Recognition using Wearable Devices

Sensor data may entail various types and communication methods based on the modules configured into a wearable device. In general, data collected through wearable devices are communicated via Bluetooth so that users can interface using a smartphone application or receive service through a website. For continuous data collection, low power requirement is also essential. However, currently available pet wearable devices use three-axis accelerometers to measure the number of steps and show only the activity

of the pet, and some devices do not have sensors. Considering the characteristics of pets, this is an essential configuration module for behavior recognition. Moreover, general users cannot access such collected data directly [13]. Table 1 compares some of the available pet wearable devices.

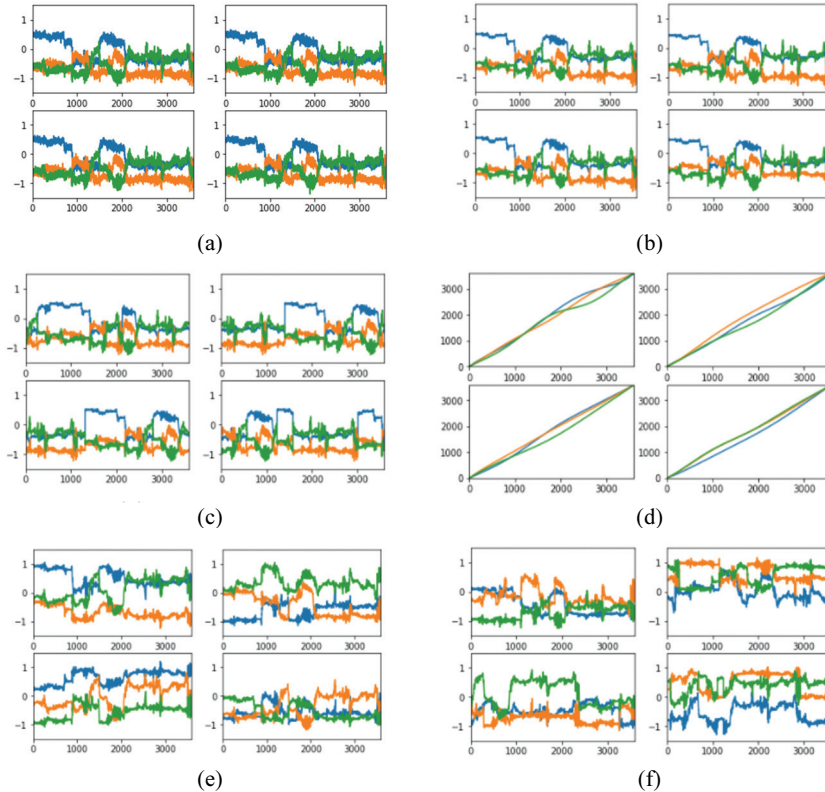


Fig. 1. Examples of sensor data augmentation results using mathematical calculations: (a) jittering, (b) scaling, (c) permutation, (d) time warping, (e) rotation, and (f) combination.

Table 1. Comparison of available pet wearable devices

Device	3-axis Acc	3-axis Gyro	Bluetooth	Low power	Behavior recognition
FitBark	O		O	O	
PetPace				O	
Whistle	O		O	O	
PitPat	O		O	O	
Proposed device	O	O	O	O	O

In this study, we develop a device comprising a three-axis accelerometer and three-axis gyroscope sensor module for behavior recognition in an effort to go beyond simple activity measurement, as in the case of existing pet wearable devices. The proposed device also provides accessibility to transmit the collected data using Bluetooth, thereby ensuring continuous data collection with low power consumption.

In general, behavior recognition using sensor data is based on information gathered via a three-axis accelerometer or three-axis gyroscope. To improve the efficiency and reliability of behavior recognition

performance, it is necessary to fuse information from multiple sensors instead of data from a single sensor [14,15]. To this end, data collected from wearable devices are generally subjected to preprocessing operations, such as filtering and resolving data omission problems due to sensor errors.

Existing behavior recognition research is mainly human centered and is often conducted using the sensors in a smartphone or smartwatch rather than through body sensors [16,17]. In previous studies, high accuracies were obtained for behavior classification and recognition by constructing training models using the processed data and by deep learning. In general, these models used the CNN-LSTM hybrid approach, where the CNN extracted features and LSTM reflected the time-series data characteristics [18,19].

3. 1D-CNN-LSTM Hybrid-Model-based Pet Behavior Recognition

The process for the 1D-CNN-LSTM hybrid-model-based pet behavior recognition using a wearable device is shown in Fig. 2. The data collected through the developed pet wearable device are first transmitted to a server. After replacing the missing values and outliers in this data through preprocessing, a sliding-window sequence is constructed for use as input values to a learning model. Data augmentation is then performed on the preprocessed data for each labeled behavior. Finally, behavior recognition is implemented to derive five types of behaviors, namely lying, running, sitting, standing, and walking, using the 1D-CNN-LSTM hybrid learning mode.

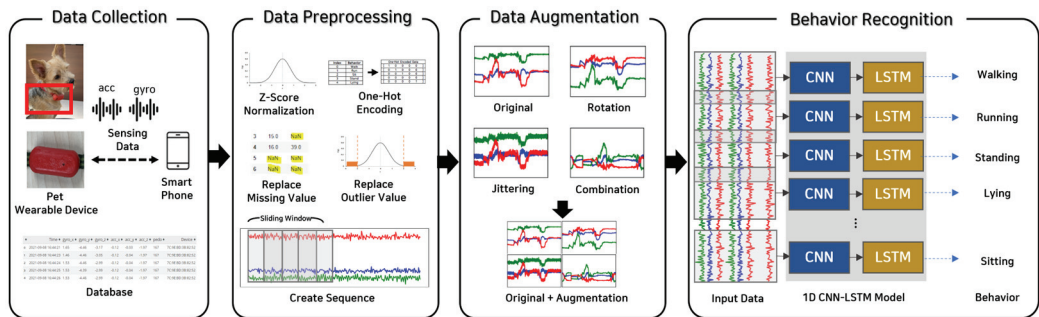


Fig. 2. Process of 1D-CNN-LSTM hybrid-model-based pet behavior recognition using a wearable device.

3.1 Data Collection

To collect the necessary data for this study, a pet wearable device was manufactured using the Arduino Nano 33 IoT board. A lithium-polymer battery and TP4056, which is a type-C charging module, were soldered to the board to provide power supply and rechargeability. Thereafter, a custom case was manufactured for the device using a 3D printer, and the entire device was configured so as to be worn as a pet collar, as shown in Fig. 3.

The board collects three-axis accelerometer and three-axis gyroscope data using the built-in inertial measurement unit module (LSM6DS3). Then, the pet pedometer information is calculated from the collected data. A Bluetooth low energy module (NINA-W102) is used to connect the device to a smartphone for data transmission. The application checks the mac address of the connected wearable

device, adds a mac column to the collected data, and transmits the data to the database. The data table structure in the database is shown in Table 2.

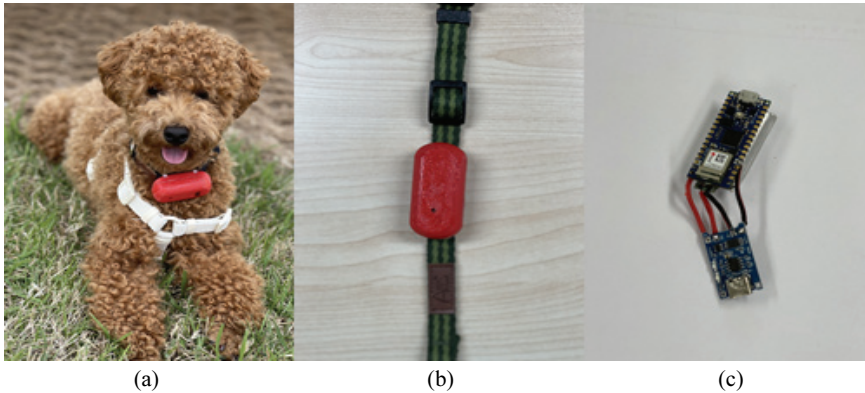


Fig. 3. Illustration showing (a) the wearing of the proposed device, (b) manufactured pet wearable device, and (c) device components.

Table 2. Data table structure in the database

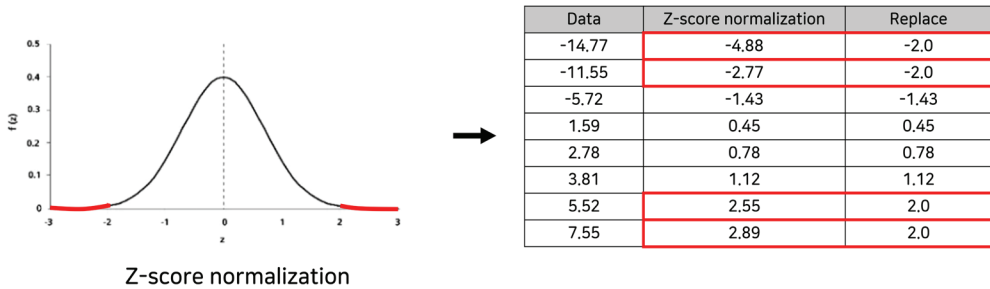
Time	Gyro_x	Gyro_y	Gyro_z	Acc_x	Acc_y	Acc_z	Pedo	Mac address
2022-01-09 14:08:27	-1.1	20.45	-137.76	-0.1	0.09	-0.34	1810	7C:9E:BD:3B:82:52
2021-01-09 14:08:28	74.71	52.49	2.01	-0.55	-1.07	-0.43	1813	7C:9E:BD:3B:82:52
2021-01-09 14:08:29	-5.07	-155.88	-21.55	-1.57	-1.67	-1.12	1815	7C:9E:BD:3B:82:52
2021-01-09 14:08:30	-74.22	-17.46	-33.94	-0.88	-0.98	-1.5	1817	7C:9E:BD:3B:82:52
2021-01-09 14:08:31	23.62	3.05	117.98	-2.15	-2.73	-0.51	1820	7C:9E:BD:3B:82:52

3.2 Data Preprocessing

Using the collected data, missing value processing, Z-score normalization, and sequence generation are performed consecutively. The missing values are removed by filtering when their durations exceed 2 seconds in the collected data. Thereafter, the average values are calculated for the remaining missing values, and a substitution process is implemented. Z-score normalization is then performed to handle outliers, where the mean and standard deviation are first computed using Eq. (1); then, values below -2 or above $+2$ in the normalized data are judged as outliers and replaced with static values of -2 and $+2$, respectively. An example of the normalization procedure is shown in Fig. 4.

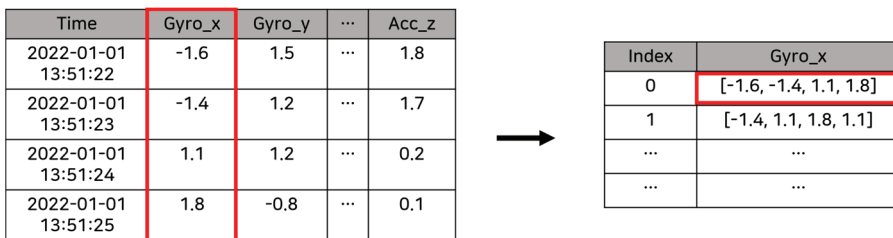
The preprocessed sensor data are filtered using a sliding window for each column to create the sequences. At this time, the length of each sequence is approximately 4 seconds. Then, the sequences are merged to compose the input data required for model training, and labeling is performed on these sequences. A total of five labels are applied, namely lying, running, sitting, standing, and walking. One-hot encoding is then performed on the labeled data to change their structure before being input to model training. One-hot encoding is performed on the sensor values and behavior label data, as shown in Fig. 5.

$$z - score = \frac{x - mean(x)}{stddev(x)} \tag{1}$$

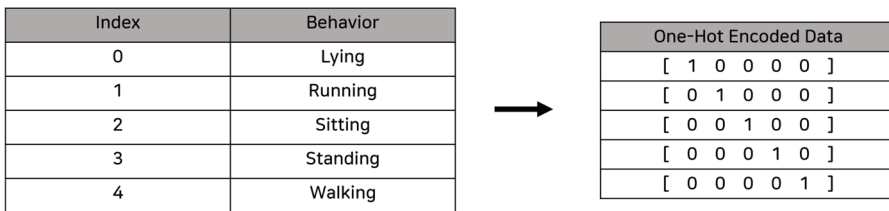


Z-score normalization

Fig. 4. Example showing replacement of outlier values through Z-score normalization.



Create Sequence



One-Hot Encoding

Fig. 5. Examples of sequence generation and one-hot encoding.

3.3 Sensor Data Augmentation

Based on the data obtained after preprocessing, the number of data for each action is identified such that augmentation can be performed on the corresponding time series. Feature extraction from a learning base, such as GAN, does not allow selection of the desired features; however, feature extraction based on a statistical method requires fewer computations and allows selection of specific features [20]. Since the pet wearable device cannot be guaranteed to be in the same position each time it is worn, the sensor data values may change according to differences in the angles. Therefore, among the various augmentation techniques, jittering that generates noise, rotation that creates arbitrary rotation angles, and a combination of these two types are used, as shown in Fig. 6.

3.4 Sensor Data Augmentation

After augmentation, the data are used as input values to the deep-learning-based behavior recognition

model. To capture the irregular features of pet behaviors, we use the CNN-LSTM hybrid model, whose structure is as shown in Fig. 7. A feature map for the entire data sequence is created using the Conv1D layer, which performs the one-dimensional convolution operation, and the features of the time-series elements are extracted using the LSTM layer. At this time, after extracting the three-axis accelerometer values from the Lambda layer via Eq. (2), the time-series features are additionally reflected in the LSTM layer. The extracted features output the values for each behavior through the Dense layer; these values are added, and the probability for each behavior is derived using the softmax function.

$$acc_{feature} = \sqrt{x^2 + y^2 + z^2}. \quad (2)$$

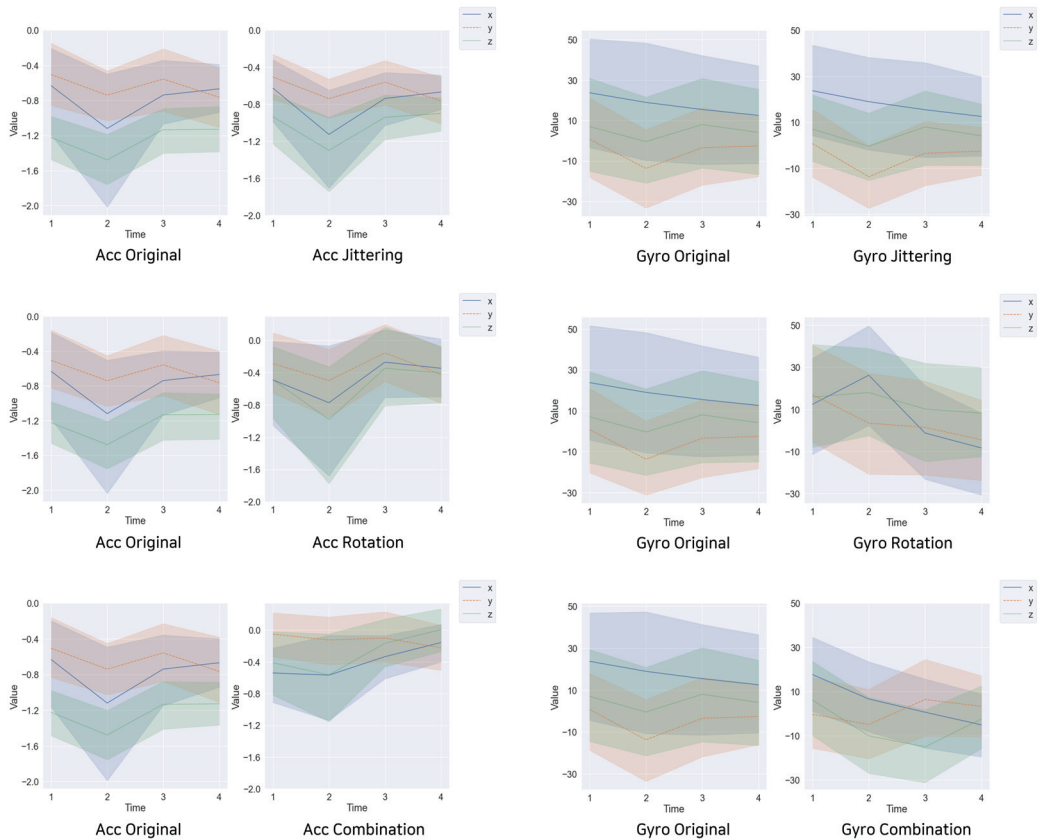


Fig. 6. Sensor data augmentation results by applying statistical methods (jittering, rotation, and their combination).

4. Experiments

Experiments were conducted in compliance with the protocols of the animal ethics committee after completing animal experimentation ethics education. The experimental environment for the model proposed in this study is shown in Table 3.

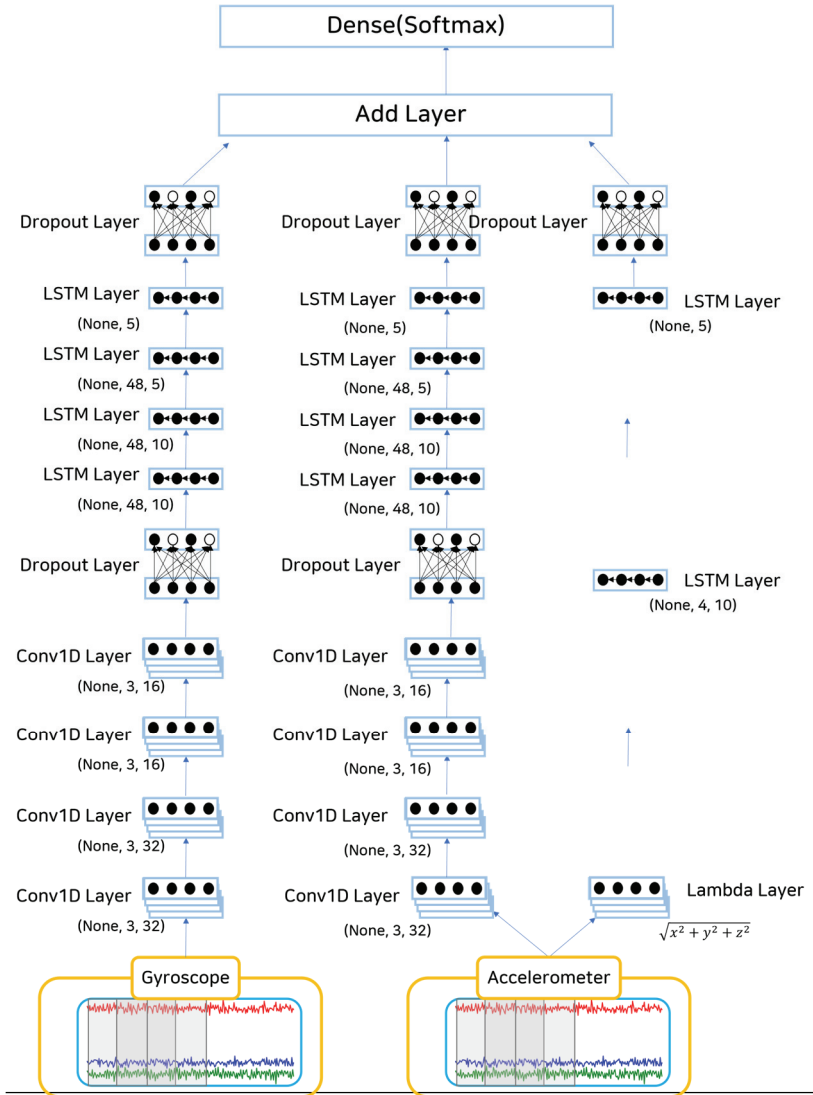


Fig. 7. Structure of the 1D-CNN-LSTM hybrid model for pet behavior recognition.

Table 3. Experimental environment

Type	Value
CPU	AMD Ryzen 9 5900K
GPU	NVIDIA GeForce RTX 3080
Memory	64 GB
Python	3.6.9
TensorFlow	2.5.0
Keras	2.5.0

4.1 Dataset

The data collected by the wearable device are stored in the server database through the user's smartphone application. After preprocessing and augmentation, the data are applied to the deep-learning-

based behavior recognition model. The composition of the collected data is shown in Table 4. Of the total amount of data, 80% was used for training, and 20% was used for testing. About 20% of the training data were used for validation during learning.

Table 4. Data composition for each type of behavior (n=985)

Behavior	Number of data	Proportion of data (%)	Number of training data	Number of test data
Lying	182	18.5		
Running	120	12.2		
Sitting	245	24.9	788	197
Standing	318	32.2		
Walking	120	12.2		

4.2 Experimental Results

From the configured dataset, a suitable model was selected for pet behavior recognition. We experimented with the CNN model to extract features, LSTM model to reflect the time-series features, and hybrid CNN-LSTM model depending on whether lambda was applied to the accelerometer values. In the CNN-LSTM model, the least recently used (LRU) criterion was applied for activation of the Conv1D layer, and tanh was used for activation of the LSTM layer. The number of epochs was 100, batch size was 8, and learning rate was 0.001 with the Adam optimizer.

The loss and accuracy graphs for each model are shown in Fig. 8, and the results of the verifications are shown in Table 5. Verifications were performed using the same dataset, and in most cases, the results with the additional Lambda layer were better than those of the existing model. Among these, the model with the additional Lambda layer in the CNN-LSTM showed the highest accuracy of 88.76%, and it is noted that the differences between loss and accuracy for validation and training are more uniform and smaller than those of the other models.

For accurate performance comparisons, the accuracies of each of the actions for all the models were derived in the form of a confusion matrix, as shown in Fig. 9. From Fig. 9, it is evident that the behaviors showing the greatest differences are walking and running. All six models used in the experiments predicted lying, sitting, and standing with high accuracies. In the case of these models, there were no significant differences between the CNN and LSTM. In particular, the CNN-LSTM model showed high performance with respect to walking and running behaviors as well as better results when used together with the Lambda layer. Table 6 shows the results of the performance evaluations in terms of the calculated precision, recall, and F1-score for the CNN-LSTM model using the Lambda layer along with the best performance based on Eqs. (3)–(5).

$$Precision = \frac{True\ positive}{True\ positive + False\ positive}, \quad (3)$$

$$Recall = \frac{True\ positive}{True\ positive + False\ negative}, \quad (4)$$

$$F1\ -\ score = 2 \times \frac{Recall \times Precision}{Recall + Precision}, \quad (5)$$

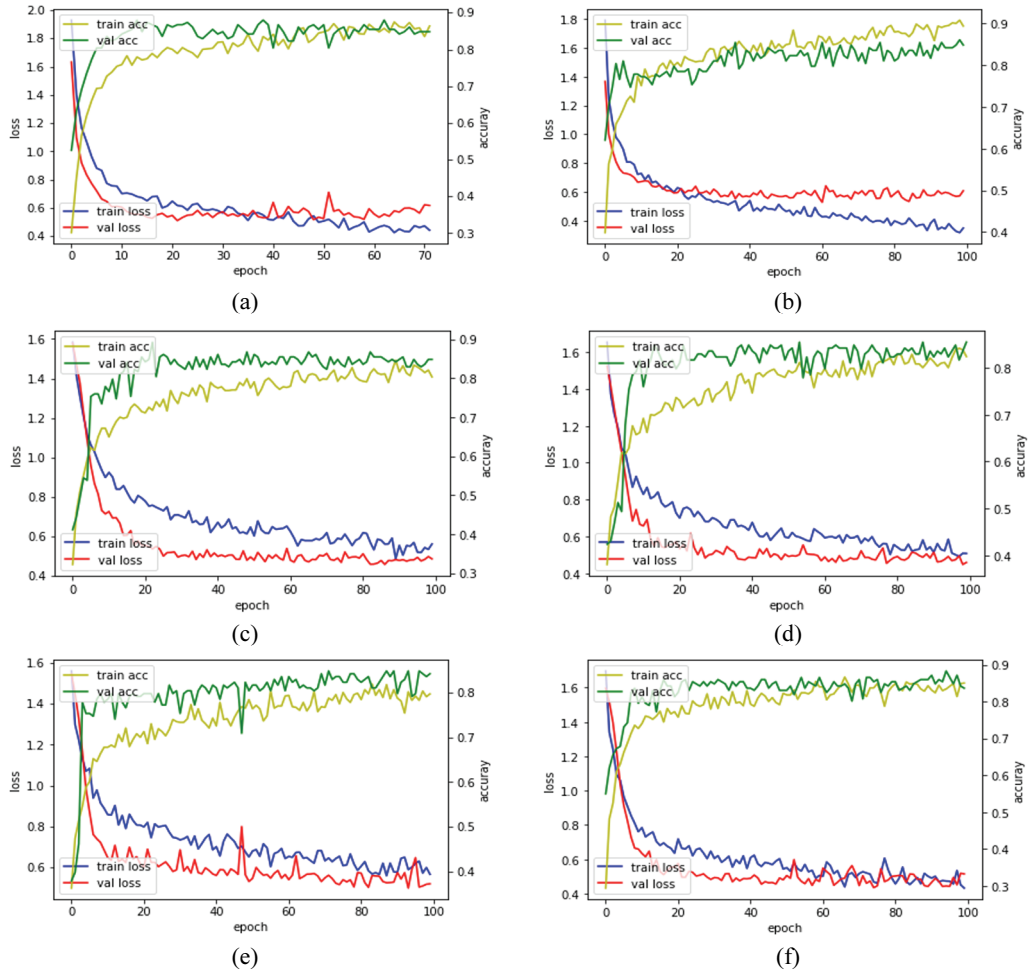


Fig. 8. Comparison of accuracy and loss for each model: (a) CNN, (b) CNN + Lambda, (c) LSTM, (d) LSTM + Lambda, (e) CNN-LSTM, and (f) CNN-LSTM + Lambda.

Table 5. Results of the model losses and accuracies

Model	Loss	Accuracy
CNN	0.5782	0.8325
CNN + Lambda	0.6060	0.8274
LSTM	0.5595	0.8326
LSTM + Lambda	0.4544	0.8477
CNN-LSTM	0.4490	0.8527
CNN-LSTM + Lambda	0.4231	0.8876

The bold font indicates the best performance in each test.

Table 6. Performance evaluation results for each behavior with the CNN-LSTM + Lambda model

Behavior	Precision	Recall	F1-score
Lying	0.92	0.94	0.93
Running	0.85	0.71	0.77
Sitting	0.95	0.84	0.89
Standing	0.83	0.97	0.89
Walking	0.86	0.79	0.83

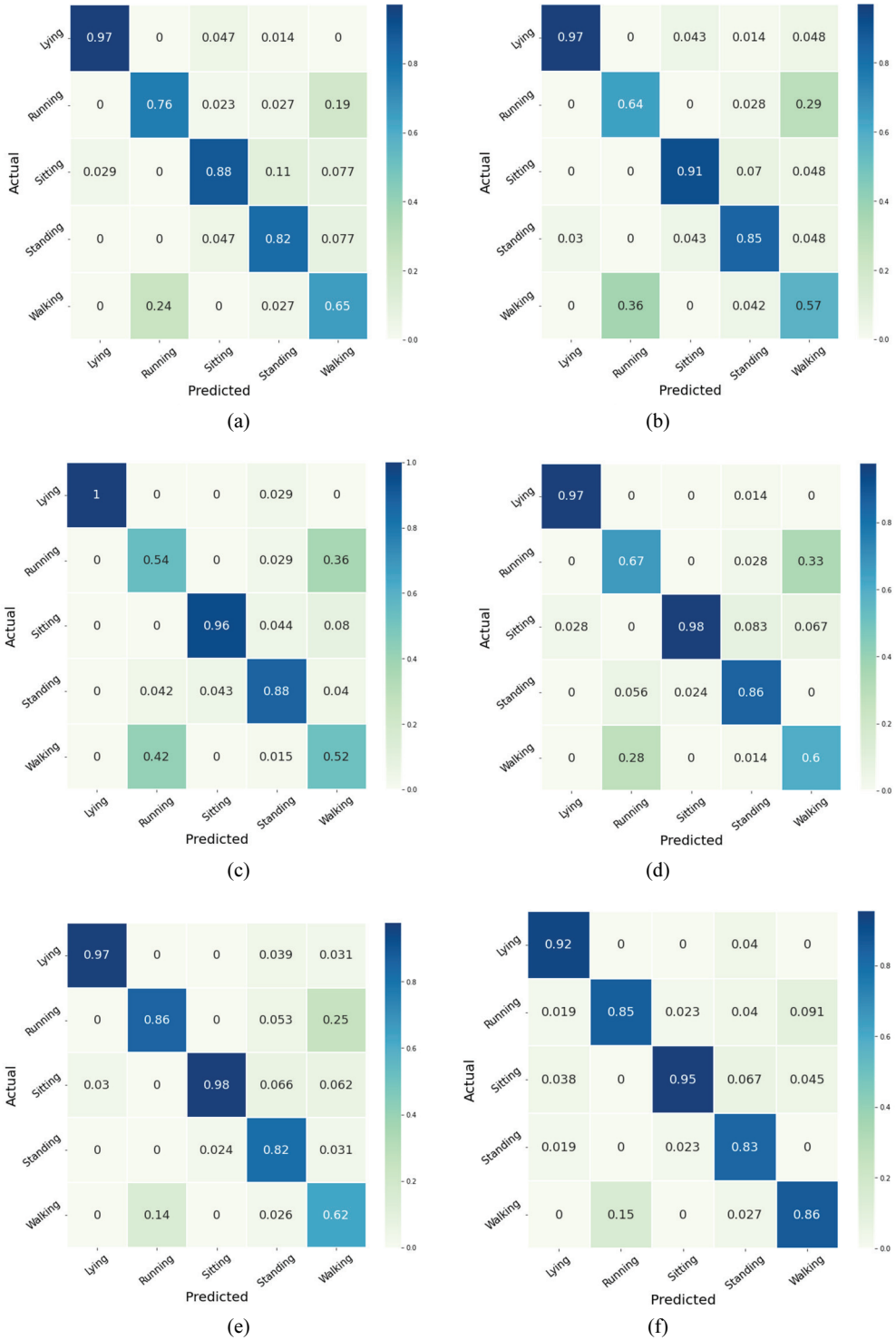


Fig. 9. Confusion matrix for each model for the behavior recognition results: (a) CNN, (b) CNN + Lambda, (c) LSTM, (d) LSTM + Lambda, (e) CNN-LSTM, and (f) CNN-LSTM + Lambda.

5. Conclusion

This paper presents a 1D-CNN-LSTM hybrid-model-based pet behavior recognition scheme with wearable sensor data augmentation. A wearable device based on the Arduino was used to collect accelerometer and gyroscope sensor values from pets through a smartphone application. Thereafter, the collected data were stored in a server database and preprocessed for missing values and outliers. Then, data sequences of length 4 seconds were used as inputs to the proposed model. Since wearable devices cannot always guarantee data collection from the same position, data augmentation was performed to artificially generate noise. Data imbalances from the different behaviors were also considered. Based on the configured dataset, the CNN, LSTM, and CNN-LSTM models with and without a Lambda layer were assessed for performance with respect to five behaviors, namely walking, standing, sitting, running, and lying. Although the overall metrics did not differ significantly between the models, the CNN-LSTM model with the Lambda layer showed the highest accuracy of 88.76%; specifically, running and walking showed better performances with the proposed model than the other models.

However, owing to the lack of sufficient experimental data, this study has limitations in that the proposed model cannot guarantee consistent behavior recognition for various pet species. Therefore, we expect that this study can be extended from pet behavior recognition to recognition of abnormal behaviors. In the future, as the number of recognized behaviors increases in accordance with data collected from large numbers of animals, we intend to expand the present research to classify abnormal behaviors.

Acknowledgement

This paper was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A2C2011966).

References

- [1] L. Morgan, A. Protopopova, R. I. D. Birkler, B. Itin-Shwartz, G. A. Sutton, A. Gamliel, B. Yakobson, and T. Raz, "Human–dog relationships during the COVID-19 pandemic: booming dog adoption during social isolation," *Humanities and Social Sciences Communications*, vol. 7, article no. no. 155, 2020. <https://doi.org/10.1057/s41599-020-00649-x>
- [2] Z. Ng, T. C. Griffin, and L. Braun, "The new status quo: enhancing access to human–animal interactions to alleviate social isolation & loneliness in the time of COVID-19," *Animals*, vol. 11, no. 10, article no. 2769, 2021. <https://doi.org/10.3390/ani11102769>
- [3] G. Cicceri, F. De Vita, D. Bruneo, G. Merlino, and A. Puliafito, "A deep learning approach for pressure ulcer prevention using wearable computing," *Human-centric Computing and Information Sciences*, vol. 10, article no. 5, 2020. <https://doi.org/10.1186/s13673-020-0211-8>
- [4] H. Alshammari, S. A. El-Ghany, and A. Shehab, "Big IoT healthcare data analytics framework based on fog and cloud computing," *Journal of Information Processing Systems*, vol. 16, no. 6, pp. 1238-1249, 2020. <https://doi.org/10.3745/JIPS.04.0193>

- [5] C. Zhu, W. Sheng, and M. Liu, "Wearable sensor-based behavioral anomaly detection in smart assisted living systems," *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 4, pp. 1225-1234, 2015. <https://doi.org/10.1109/tase.2015.2474743>
- [6] Q. Wen, L. Sun, F. Yang, X. Song, J. Gao, X. Wang, and H. Xu, "Time series data augmentation for deep learning: a survey," 2020 [Online]. Available: <https://arxiv.org/abs/2002.12478>.
- [7] X. Cui, V. Goel, and B. Kingsbury, "Data augmentation for deep neural network acoustic modeling," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 9, pp. 1469-1477, 2015. <https://doi.org/10.1109/TASLP.2015.2438544>
- [8] X. Zhao, J. Sole-Casals, B. Li, Z. Huang, A. Wang, J. Cao, T. Tanaka, and Q. Zhao, "Classification of epileptic EEG signals by CNN and data augmentation," in *Proceedings of 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020, pp. 926-930. <https://doi.org/10.1109/icassp40776.2020.9052948>
- [9] J. Cho and N. Moon, "Design of image generation system for DCGAN-based kids' book text," *Journal of Information Processing Systems*, vol. 16, no. 6, pp. 1437-1446, 2020. <https://doi.org/10.3745/JIPS.02.0149>
- [10] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, article no. 60, 2019. <https://doi.org/10.1186/s40537-019-0197-0>
- [11] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," 2017 [Online]. Available: <https://arxiv.org/abs/1712.04621>.
- [12] T. T. Um, F. M. Pfister, D. Pichler, S. Endo, M. Lang, S. Hirche, U. Fietzek, and D. Kulic, "Data augmentation of wearable sensor data for Parkinson's disease monitoring using convolutional neural networks," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, Glasgow, UK, 2017, pp. 216-220. <https://doi.org/10.1145/3136755.3136817>
- [13] D. Van Der Linden, A. Zamansky, I. Hadar, B. Craggs, and A. Rashid, "Buddy's wearable is not your buddy: privacy implications of pet wearables," *IEEE Security & Privacy*, vol. 17, no. 3, pp. 28-39, 2019. <https://doi.org/10.1109/msec.2018.2888783>
- [14] H. F. Nweke, Y. W. Teh, G. Mujtaba, U. R. Alo, and M. A. Al-garadi, "Multi-sensor fusion based on multiple classifier systems for human activity identification," *Human-centric Computing and Information Sciences*, vol. 9, article no. 34, 2019. <https://doi.org/10.1186/s13673-019-0194-5>
- [15] T. Steels, B. Van Herbruggen, J. Fontaine, T. De Pessemier, D. Plets, and E. De Poorter, "Badminton activity recognition using accelerometer data," *Sensors*, vol. 20, no. 17, article no. 4685, 2020. <https://doi.org/10.3390/s20174685>
- [16] S. A. Khowaja, B. N. Yahya, and S. L. Lee, "CAPHAR: context-aware personalized human activity recognition using associative learning in smart environments," *Human-centric Computing and Information Sciences*, vol. 10, article no. 35, 2020. <https://doi.org/10.1186/s13673-020-00240-y>
- [17] A. R. Javed, M. U. Sarwar, M. O. Beg, M. Asim, T. Baker, and H. Tawfik, "A collaborative healthcare framework for shared healthcare plan with ambient intelligence," *Human-centric Computing and Information Sciences*, vol. 10, article no. 40, 2020. <https://doi.org/10.1186/s13673-020-00245-7>
- [18] Z. Xu, J. Zhao, Y. Yu, and H. Zeng, "Improved 1D-CNNs for behavior recognition using wearable sensor network," *Computer Communications*, vol. 151, pp. 165-171, 2020. <https://doi.org/10.1016/j.comcom.2020.01.012>
- [19] S. Mekruksavanich, A. Jitpattanakul, P. Youplao, and P. Yupapin, "Enhanced hand-oriented activity recognition based on smartwatch sensor data using LSTMs," *Symmetry*, vol. 12, no. 9, article no. 1570, 2020. <https://doi.org/10.3390/sym12091570>
- [20] J. Kim and N. Moon, "Dog behavior recognition based on multimodal data from a camera and wearable device," *Applied Sciences*, vol. 12, no. 6, article no. 3199, 2022. <https://doi.org/10.3390/app12063199>



Hyungju Kim <https://orcid.org/0000-0002-3664-8840>

He received B.S. degrees in School of Computer Science and Engineering from Hoseo University in 2021. Since March 2021, he is current with the Department of Computer Science and Engineering from Hoseo University as Master Course. His research interests include deep learning, IoT service, and big data processing and analysis.



Nammee Moon <https://orcid.org/0000-0003-2229-4217>

She received B.S., M.S., and Ph.D. degrees from the School of Computer Science and Engineering at Ewha Womans University in 1985, 1987, and 1998, respectively. She served as an assistant professor at Ewha Womans University from 1999 to 2003, a then as a professor of digital media, Graduate School of Seoul Venture Information, from 2003 to 2008. Since 2008, has been a professor of computer information at Hoseo University. Her current research interests include social learning, HCI and user-centric data, and big data processing and analysis.