Long Short-Term Memory (LSTM)-Based Dog Activity Detection Using Accelerometer and Gyroscope

Ali Hussain 1, Khadija Begum 1, Tagne Poupi Theodore Armand 1, Md Ariful Islam Mozumder 1, Sikandar Ali 1, Hee Cheol Kim 2 and Moon-Il Joo 1,*

1 Institute of Digital Anti-Aging Healthcare, Inje University, Gimhae 50834, Korea
2 College of AI Convergence, Institute of Digital Anti-Aging Healthcare, u-AHRC, Inje University, Gimhae 50834, Korea
* Correspondence: joomi@inje.ac.kr; Tel.: +82-55-320-4027

Abstract: Dog owners are extremely driven to comprehend the activity and health of their dogs and to keep tabs on their well-being. Dogs’ health and well-being, whether as household pets or service animals, are critical issues that are addressed seriously for moral, psychological, and economical reasons. Evaluations of a dog’s welfare depend on quantitative assessments of the frequency and variability of certain behavioral features, which are sometimes challenging to make in a dog’s normal environment. While it is challenging to obtain dogs’ behavioral patterns, it is nearly impossible to directly identify one distinct behavior when they are roaming around at will. Applications for automatic pet monitoring include real-time surveillance and monitoring systems that accurately identify pets using the most recent methods for the classification of pet activities. The suggested method makes use of a long short-term memory (LSTM)-based method to detect and classify the activities of dogs based on sensor data (i.e., accelerometer and gyroscope). The goal of this study is to use wearable sensor data and examine the activities of dogs using recurrent neural network (RNN) technology. We considered 10 pet behaviors, which include walking, sitting, down, staying, feeding, sideways, leaping, running, shaking, and nose work. As dog activity has a wider diversity, experimental work is performed on the multi-layer LSTM framework to have a positive influence on performance. In this study, data were collected from 10 dogs of various ages, sexes, breeds, and sizes in a safe setting. Data preprocessing and data synchronization were performed after the collection of data. The LSTM model was trained using the preprocessed data and the model’s performance was evaluated by the test dataset. The model showed good accuracy and high performance for the detection of 10 activities of dogs. This model will be helpful for the real-time monitoring of dogs’ activity, thus improving the well-being of dogs.

Keywords: pet activity detection; dog activity; accelerometer; gyrometer; wearable sensors; LSTM; classifier; RNN

1. Introduction

The practice of identifying a pet’s activity through the analysis of video or sensor data can be termed as Pet Activity Detection (PAD). It serves as a helpful indication of animal health and welfare, which benefits pet owners, veterinarians, and the livestock industry. According to studies, observing what animals do when they are not being watched by humans can reveal a lot about their psychological and physical well-being. Recognizing the behavior of the dog would offer additional details, as surveillance with a widely viable activity tracker provides levels of daily activity and rest.

Scientists studying animals have discovered that specific behaviors can serve as important indications of sickness and discomfort, similar to how atypical behavior indicates illness in people. The slow beginning stages of the emergence of a condition such as
Arthritis that reduces movement are frequently missed by owners. Such a lapse of judgment may result in the animal suffering greatly as a result of being left untreated.

Animal activity recognition has grown in popularity recently because of its applicability in a variety of industries, including remote monitoring, security, and health. Additionally, more people are able to contribute to this subject thanks to the rise in sensor-based technologies. The low cost of numerous sensors and the quick development of technology, particularly sensor devices, make it simple to monitor the everyday activities of pets. Low-voltage pressure sensors provide long-lasting and better performance for advanced sensing furthermore it is and are gaining much attention [1]. Wearable technology is utilized to identify activities. Wearable gadgets are seamlessly enabled in the activity recognition process by inertial sensors and embedded systems. Wearable sensors have been developed for activity detection and are utilized in daily life and sporting activities. They combine embedded systems with acceleration and gyro sensors. These days, they are used in a wide range of everyday contexts, such as animal welfare evaluation, medical monitoring, therapeutic interventions, activity recognition, remote control, and health care [2–11].

The research is clear on the value of accelerometry data for developing supervised machine learning models to categorize diverse animal and livestock behaviors, particularly those of cattle. There are several examples [12–23] of such research. However, there has not been much work performed to implement behavior categorization on a wearable device with an accelerometer and gyro sensors [16,24,25]. Few scientific articles highlighted in recent years [26,27] contend that understanding an animal’s activity patterns might help us better comprehend its health and welfare. Developing an information ecosystem that can collect data and extract insights to monitor animal activity is one method of keeping an eye on animals. It is quite easy to create a hypothesis from an animal’s daily activity if the outcomes of such a sensor device are localized in areas such as farms and vet offices. These theories may be pertinent to future health issues, less stressful conditions for animals, and other issues.

There is also another option for recognizing animal activity, called vision-based systems. Data from a camera placed close to the animal are used by computer vision-based approaches [28]. This method has a number of drawbacks, including a high reliance on light, the need for resolution, a constrained viewing area, high costs, and privacy risks. As such, we used a sensor-based methodology to conduct the tests in this work. Data were gathered for our study from 10 dogs of various ages, sexes, breeds, and sizes. The owners of the dogs granted consent for the data to be collected. The data were collected using wearable sensors that were attached to the dogs’ necks and tails.

The main purpose of this research was to assess how recent approaches to animal welfare translated to pet behavior. Sometimes, dogs experience various health conditions unknown to their owners. We aim to analyze the behavior of dogs by identifying differences in behavior such as walking, sitting, down, staying, eating, sideways, jumping, running, shaking, and nose work, in order to identify special behavioral characteristics using deep learning with long short-term memory (LSTM). Even though academics have put forward a variety of ways, several shortcomings, such as the fact that they mostly employed data only from accelerometer, while others simply used raw data without using feature engineering methods, persist in those methods. In this paper, we took these aspects into account and employed two sensors (an accelerometer and a gyroscope) on the dogs’ necks and tails. We also performed feature engineering and used the LSTM model for predicting the dog’s activity.

There are certain reasons for why we have used LSTM. Since our data are time series data, LSTM gives better performance compared to other neural network architectures. LSTM can store information for a longer time where other neural network technologies cannot do so. For the prediction of future events in time series data, it is important to handle the longer dependencies, so LSTM is best fit for such data. LTS also helps resolve
the vanishing gradient problems which occur while training the model. The following is the summary of our work’s major contributions.

- It studies the usefulness of machine learning methods to categorize the primary behavior patterns of dogs from sensor data.
- Application of popular LSTM deep learning technology to identify the pet activity on collected data from two type of sensors.
- We are among the pioneers to use the LSTM approach to identify dog behaviors that consider information from sensors, including an accelerometer and a gyroscope.
- We examined 10 types of behavior by applying our proposed LSTM model.
- We have employed class weight techniques that have been shown to be suitable for the identification of pet activities to address the imbalance issue.

The remainder of this article is divided into the following sections. Related work is briefly summarized in Section 2. The materials and procedures are described in Section 3. The methodology for activity detection is described in Section 4. The discussion and outcomes of the experiments are covered in Section 5. Section 6 draws our research work to a conclusion.

2. Related Work

Wearable sensors have become more particular for detecting animals (especially dogs) in various types of activity and emotions. From the research perspective, body sensors have attracted a lot of research studies for a large number of applications including monitoring vital signs activities, and emotions. Previously, researchers studied machine learning techniques but nowadays they are using deep learning for obtaining better experimental results. Seabra et al. [29] used accelerometer and gyroscope data for the detection of different activities and emotions. Their project aimed to “augment” the service dog by creating a system that makes up for some of the dog’s shortcomings with a robotic device installed on the harness. Ladha et al. [30] have identified several behaviors that are associated with certain activities that are important for a dog’s wellness. They created a framework for collar-worn accelerometry that records canine behaviors in realistic settings. The framework for statistical classification was utilized to identify dog actions. In an experimental study, they looked at the naturalistic behavior of 18 dogs, and they were able to identify 17 different activities with a classification accuracy of approximately 70%. Massawe et al. [31] used sensors and a smart sensing device has been developed to assist in identifying animal emotions based on physiological indicators collected from sensors affixed to the animal’s body. A body temperature sensor, a galvanic skin resistance sensor, and a heart rate sensor all continually collect signals. The sensor signals are first filtered and amplified, then processed in the microcontroller and wirelessly transferred utilizing GSM modem and ZigBee technologies. The signals that the system sends out are seen and saved in the database, where they are visually examined for patterns. The project’s four primary emotion criteria were joyful (excited), sad, furious, and neutral (relaxed). In this investigation, a pilot study was conducted using dogs. In this study, the result will show animal emotions based on physiological signals. Gerencser et al. [32] created a “behavior tracker” that uses a supervised learning algorithm and a multi-sensor logger device (with accelerometer and gyroscope) to automatically identify the behavior of freely moving dogs. The battery timing of the sensors is also an important factor for the long-time detection of activities and motions. Sung-Chan Kim et al. [33] used 10 GHz band controlled oscillator in the manufacturing of motion detecting sensors. Hammond et al. [34] developed tiny acceleration-logging devices and used them to describe two chipmunk species’ behavior. From confined individuals, they gathered paired accelerometer data and behavioral observations. Then, they developed an automatic system for categorizing accelerometer measurements into behavioral categories using machine learning techniques. Last but not least, they released and retrieved accelerometers from wild, free-ranging chipmunks. According to their research, this is the first precedent in which accelerometers have been
employed to produce behavioral data on small-bodied, free-living mammals. Kumpulainen et al. [35] examined the effectiveness of a 3D accelerometer movement sensor mounted on a dog collar for its ability to categorize seven activities in a semi-controlled test environment involving 24 dogs. The acceleration time series signals contained a number of features that were extracted. With two feature scenarios using all calculated features and the ones provided by the forward selection algorithm the performance of two classifiers was assessed. The seven behaviors had the highest overall categorization accuracy of 76%. The results for enhancing the classification of particular behaviors by very basic algorithms are encouraging. As in previous studies performed by behoora et al. [36], they managed to find an association of emotional states between gestures and emotions with accuracies above 98%. McClune et al. [37] also used acceleration data and analyzed the detection of various activities. The classification task was performed using a traditional learning method. Additionally, Zia Uddin et al. [38] provided a reliable activity recognition method for intelligent health care that makes use of body sensors and deep convolutional neural networks (CNN) by using readings from various body sensors, including ECG, magnetometer, accelerometer, and gyroscope sensors, for medical analysis. A deep activity CNN was trained using the salient features after principal component analysis using a Gaussian kernel and Z-score normalization was used to extract them from the sensor data. The deep CNN was utilized to identify the activities in test data. A standard dataset that is openly accessible was used to test the proposed strategy, and the outputs were compared with other traditional methods. The findings of the experiments showed that the process is trustworthy and could be used for cognitive support in body sensor-based smart health care systems. Researchers have used different approaches for the motion detection such as Zhang, Ziyi [39] used camera for the driver smartphone usage detection. Valletta et al. [40] introduced animal behaviorists unfamiliar with machine learning (ML) for the solution of complex behavior data. Kasnesis et al. [41] developed deep learning to track the dogs in real time. They used signal data to identify the position of a search and rescue (SaR) dog and then to recognize and alert the rescue team whenever the SaR dog spots a victim. They proposed a deep learning-assisted implementation that combines a body device, a base station, and a cloud-based system. To achieve this, they used deep CNN, which was trained using information from inertial sensors such as the 3-axial accelerometer and gyroscope and the wearable's microphone, respectively. Deep CNN was used to categorize sounds and recognize activities. They created deep learning models and installed them on wearable devices. The suggested implementation was evaluated in two separate search and rescue situations, successfully locating the victim in both (achieving an F1 score of more than 99%) and informing the rescue team in real-time. Bocaj et al. [42] used deep convolutional neural networks (ConvNets) to recognize farm animals’ activities, including those of goats and horses. They looked into the possible benefits of ConvNets, which have an F1-score that is more than 7% higher and an accuracy that is approximately 12.5% higher than other machine learning algorithms in the literature. Furthermore, they outlined the benefits of 2D convolution and demonstrated that adding more filters to each layer does not always result in improved classification accuracy. They compared several ConvNet designs and showed how hyperparameter adjustment could improve overall accuracy. Venkatraman et al. [43] used wireless sensor devices that are mild and short enough to be worn by rats and other small animals have been developed. These sensors were used to capture data on three axes of acceleration from animals in a cage as they behave naturally. The collected acceleration data were subsequently processed to extract the animal’s behavior using neural network-based pattern recognition algorithms. Using this method, feeding, grooming, and standing were successfully recognized. They used machine learning algorithms (SVM, KNN, RF, and ANN). The algorithm got 97% accuracy for standing activity, 93% for eating, and 98% for grooming, respectively.
3. Materials and Methods

This section covers the methods and materials used to achieve our goal. We applied several methods and techniques to develop our proposed model such as data collection, data preprocessing, feature engineering, data balancing, model development, and evaluation.

Data used in this research were gathered from 10 dogs of regardless of their ages, sexes, breeds, and sizes. After obtaining the dogs’ owners’ consent, we attached wearable sensor devices to the dogs’ necks and tails for data collection. The accelerometer and gyroscope were the two fundamental sensors utilized in the wearable sensors set at a sampling frequency of 33.33 Hz, enabling the investigation of the dogs’ activities. The sensors were capable of tracking the dogs’ rotations and linear movements throughout all angles without discomforting them. The devices were lightweight as the device that were worn at neck is 16 g with a dimension of 52 × 38 × 20.5 mm. Additionally, the tail-worn device’s weight is 13 g with a dimension of 35 × 24 × 15 mm. The devices were manufactured by Sweet Solution, a company located in Busan, South Korea. The scale factors were −4 to 4 and −2000 DPS to +2000 DPS for the accelerometer and the gyroscope, respectively. For this study, we selected trained dogs placed under the responsibility of trainers capable not only to determine the various pet activities but also to instruct the dogs to carry out a particular activity. The selected dogs were able to carry out the desired activity of the trainer accordingly as instructed. For each activity, the trainers make sure the sensors are well fitted on the dogs before staring to record the video and also the recordings of IMU data. The recording is performed in line with the sampling frequency. The sampling frequency were 33.33 Hz and in a second, there were 33 frames. We collected 493,144 data samples.

Any machine learning or deep learning model must be trained with adequate data to meet its target. Obtaining an adequate dataset requires many techniques among which data preprocessing. Data preprocessing is a significant factor necessary for the optimal performance of a model, its role is to increase the quality of data by reducing noise and irrelevant information. Due to the nature of our data collection process, we have noisy, irrelevant, and inconsistent data from the sensor devices on which we applied a 6th-order Butterworth low-pass filter with a cutoff frequency of 3.667 Hz as a remedy. We decided to apply for this filter order because it blocks maximum noise. Similarly, exploratory data analysis (EDA) were used to choose the cutoff frequency. This method lessens the impact of abrupt changes in accelerometer data while assisting in the filtering of sensor data that are impacted by gravity. It also helps make the data cleaner and less reliant.

It requires feature engineering to raise an AI model’s performance. This process consists of using domain knowledge to select the most relevant features from raw data to train the model aiming to improve its performance. In our case, feature engineering was applied by us onto the data gathered using the sensors. At first, we generated several features considering the number of samples. The generated features included mean, minimum, maximum, interquartile range, energy measure, skewness, kurtosis, standard deviation, and mean absolute deviation.

A class weight technique was utilized by us to address data imbalance in our dataset [44]. Most algorithms are not very effective with biased data; therefore, it is necessary to apply a technique that can balance different weights to both the majority and the minority classes of the dataset. To be more precise, the formula to calculate the weight for each class is expressed as follows:

\[
W_j = \frac{n_{\text{samples}}}{(n_{\text{classes}} \times n_{\text{samples}})}
\]  

(1)

Here, \(W_j\) is the weight for each class (\(j\) signifies the class), \(n_{\text{samples}}\) is the total number of samples or rows in the dataset, \(n_{\text{classes}}\) is the total number of unique classes in the target and \(n_{\text{samples}}\) is the total number of rows of the respective class. We have used equation 1 for calculating the weights for each class the weights are 0.337 for walking,
0.531 for sitting, 0.557 for down, 0.797 for staying, 1.142 for eating, 1.998 for sitting sideways, 4.314 for jumping, 33.592 for running, 26.357 for shaking, 2.3509 for Nosework.

4. Framework for LSTM Model

The overall architecture of our proposed model has been shown in Figure 1. First, we collected data from the dogs using sensors. Second, we synchronize the sensor data with the video recordings. Third, we used data preprocessing techniques to clean the data. Fourth, we developed our LSTM model and evaluated its performance.

An algorithm for detecting dog activity was developed that acquires biosignals via wearable devices, such as an accelerometer and gyroscope. Different preprocessing strategies, such as data filtration and normalization, were used to preprocess the biosignals. An LSTM-based algorithm was used to anticipate the dog’s behavior. This dog activity detection algorithm works in the following way:

Wearable devices with gyroscopes and accelerometers were employed firstly to collect experimental data from the dogs. These sensors were attached to the 10 different breeds’ necks and tails, age groups, and genders of dogs. The use of these sensor devices has allowed for the examination and recording of 10 distinct types of activities.

For preprocessing and filtering the raw data, a butter low-pass filter was utilized that was retrieved from the sensor devices. We obtained more accurate data as a result of the filters’ removal of noise and undesired signals from the data. The dataset underwent data normalization in order to uniformly scale and normalize the range of all the data.
The wearable sensor’s sampling frequency is 33.3 Hz. Consequently, we have a matrix with a dimension of 99 to store the data. Each matrix is structured as a set of $124 \times 99 = 12,276$ feature vectors, where 124 stands for the matrix’s 124 dimensions and 99 for the number of rows in each window or data size. The very first row of the vectors for each element was then labeled with the type for that element.

In conducting this research, we employed a deep learning method. Deep learning, in contrast to conventional artificial neural networks, can both extract features and classify them. Without the aid of a human or using a manual process, it automatically extracts the highly significant features and uses them for classification and detection. Long short-term memory (LSTM) networks are a deep learning technique that may recognize order dependency in situations involving sequence prediction. This behavior is essential in fields with complicated problems, such as voice recognition, machine translation, and behavior analysis. As the data we gathered from the sensors consisted of $x$, $y$, and $z$ values, we decided to convert these data into vector magnitude data, and an LSTM was built using this vector to classify different dog behaviors.

For different application that requires feature extraction and classification, LSTMs are one of the most famous recurrent neural network (RNN) methods. The LSTM model, which included LSTM, dropout, fully connected dense, and a Softmax layers, was used in this work to explore the activities of dogs.

**Input layer:** In the form of vector magnitude, the accelerometer and gyroscope provided the input layer of the model with six-axis data, i.e., three axis from each sensor.

**LSTM layer:** We have used 6 LSTM layers. At first, we used LSTM layers with 256 neurons 2 times, while applying dropout at 30%. After that, we applied LSTM with 128 neurons, with a dropout at 20% for the second time. In the last stage of LSTM layers, we used LSTM with 64 neurons with dropout at 20% again.

**Dropout Layer:** Dropout layers help the model to overcome the overfitting problem. It also prevents the synchronized weight optimization of all neurons in a layer. Dropout layers with dropout values of 0.3, 0.2, and 0.2 were utilized in our model to reduce the complexity of the model.

**Dense/Fully Connected Layer:** Dense layer aids in adjusting the output’s dimensionality so that the model may more easily establish the relationship between the values of the data it is working with. In our model, we have utilized dense layers consecutively 3 times with 128, 64 and 32 neurons.

**Output:** Activation functions are crucial to the prediction of any task in deep learning. A good prediction is an output of wisely selecting the activation function. In the experiment, the Rectified Linear Unit (ReLu) was employed. We utilized the Softmax algorithm to classify all 10 of the dog activities because there are several classes for the 10 dog activities that we have. Application of the adaptive moment estimation (Adam) optimizer was performed using a learning rate of 0.0001. A loss function called categorical_crossentropy was employed to determine the difference between the actual and predicted values. The performance of the model is higher the less significant the gap between the values. Figure 1 depicts the fundamental design of this investigation.

5. Experimental Results and Discussion

The details of the experimental findings are covered in this section. To balance the class labels of dog activities, we have used the class weight method, thus overcoming the imbalanced problem.

5.1. Evaluation Methods

Various performance measures, including accuracy, precision, recall, and ROC-AUC curve, were used to assess the model’s performance.

\[
\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \times 100\%
\]  

(2)
Precision = \( \frac{TP}{TP + FP} \times 100\% \) \tag{3} 

Recall = \( \frac{TP}{TP + FN} \times 100\% \) \tag{4} 

\[ F1 \text{- score} = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100 \] \tag{5}

where TP stands for “true positive”, TN for “true negative”, FN for “false negative”, and FP for “false positive”. The precision of a model’s prediction of an activity’s proper categorization is measured.

5.2. Confusion Matrix

The prediction outcome of a classification model is summarized in the confusion matrix. By showing the number of predictions that were accurate and inaccurate for each class, it displays the model’s performance. It presents details on the model’s actual and predicted classifications. Figure 2 shows the confusion matrix without normalization and Figure 3 shows the confusion matrix with normalization for all the ten dogs’ activities. We have used 80% data for training and 20% for testing of the model.

![Confusion matrix](image)

**Figure 2.** Confusion matrix without normalization.
Figure 3. Confusion matrix with normalization.

5.3. AUC-ROC Curve

The overall performance of the model can be visualized using the Area Under the Curve (AUC) of the Receiver Characteristic Operator (ROC). To put it in another way, it is a measurement of assessment that displays each class performance while drawing a graph between the true-positive rate (TPR) and false-positive rate (FPR). Figure 4 shows the ROC-AUC curve of our proposed model. The ROC curve of running activity is slightly different from the other ROCs because in running activity the data samples are less compared to other activities. In deep learning models, the more we have data, the higher will be the performance of the model. Therefore, for all the other activities, we have enough data samples and the model was well trained and produced smooth ROC curves; however, the running activity produced a slightly different ROC curve.
Figure 4. The ROC-AUC curves for each class.

6. Conclusions

This study showed the classification of dogs’ activities using the long short-term memory (LSTM) model. The wearable sensor devices were used to collect the data from different breeds of dogs. The accelerometer and gyroscope are two types of sensors that were utilized for the collection of the data in our proposed method. To reduce the noise, a butter low-pass filter was utilized when the sensor data were preprocessed. Different crucial features such as the mean, standard deviation, and energy measure were extracted from the raw data using feature extraction. We trained our model using the preprocessed data. The class weight technique was used to train the LSTM model to overcome the imbalance problem. Our proposed LSTM model showed good experimental results with the class weight technique. Our proposed LSTM model achieved good accuracy for all the ten activities, such as walking, sitting, down, staying, eating, sideway, jumping, shaking, running, and nose work. The testing accuracy of our model was 94.25%. This research study will help to monitor and improve the well-being of dogs using state-of-the-art techniques.


Funding: This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1I1A1A01050306).
Data Availability Statement: Private dataset (dog activity sensor data for multi-class classification) is not available online. The data used in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References


