Recognizing Driver Activities Using Deep Learning Approaches Based on Smartphone Sensors

Sakorn Mekruksavanich[1(✉)], Ponnipa Jantawong[1], Narit Hnoohom[2], and Anuchit Jitpattanakul[3]

[1] Department of Computer Engineering, School of Information and Communication Technology, University of Phayao, Mueang Phayao, Phayao, Thailand
sakorn.me@up.ac.th

[2] Image, Information and Intelligence Laboratory, Department of Computer Engineering, Faculty of Engineering, Mahidol University, Nakhon Pathom, Thailand
narit.hno@mahidol.ac.th

[3] Intelligent and Nonlinear Dynamic Innovations Research Center, Department of Mathematics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand
anuchit.j@sci.kmutnb.ac.th

Abstract. Human motion detection based on smartphone sensors has gained popularity for identifying everyday activities and enhancing situational awareness in pervasive and ubiquitous computing research. Modern machine learning and deep learning classifiers have been demonstrated on benchmark datasets to interpret people's behaviors, including driving activities. While driving, driver behavior recognition may assist in activating accident detection. In this paper, we investigate driving behavior detection using deep learning techniques and smartphone sensors. We proposed the DriveNeXt classifier, which employs convolutional layers to extract spatial information and multi-branch aggregation transformation. This research evaluated the proposed model using a publicly available benchmark dataset that captures four activities: a driver entering/exiting and sitting/standing out of a vehicle. Classifier performance was evaluated using two common HAR indicators (accuracy and F1-score). The recommended DriveNeXt outperforms previous baseline deep learning models with the most fantastic accuracy of 96.95% and the highest F1-score of 96.82%, as shown by many investigations.

Keywords: Human activity recognition · Deep learning · Smartphone sensors · Driver activities

1 Introduction

The domain of human activity recognition (HAR) in artificial intelligence has seen significant growth in recent years. Current HAR study findings have inspired

O. Surinta and K. Kam Fung Yuen (Eds.): MIWAI 2022, LNAI 13651, pp. 146–155, 2022.
https://doi.org/10.1007/978-3-031-20992-5_13

several applications in medical and related domains, including athletic measuring performance, rehabilitation tracking, and lousy habit identification. Based on the collection of activity data, the development of innovative wearable technology has advanced the progression of HAR research owing to the offering of various and increased activity data. Smartphones and smartwatches are two wearable gadgets that feature sensors such as accelerometers, gyroscopes, and magnetometers that individuals use throughout the globe in their everyday lives.

In the preceding ten years, the HAR research has led to the development of machine learning and deep learning techniques [9]. Nonetheless, machine learning is constrained by the need for individual specialists to extract distinguishing characteristics from raw sensor data. Using convolutional operators as the initial step of recognition models has enabled automatic feature extraction inside deep learning methodologies.

Convolutional neural networks (CNN) and long short-term memory (LSTM) neural networks were determined for the HAR deep learning approaches based on a review of the relevant literature. Several accomplished models have motivated the development of unique architectures for studying computer vision and natural language processing [13], including InceptionTime, Temporal Transformer, and ResNet. Based on these models, unfortunately, recognition performance has been restricted due to a lack of knowledge of the class hierarchy of human activities.

Activity recognition algorithms based on CNNs often use activity labels encoded as one-hot vectors. Because the one-hot encoding considers each class separate from one another, most activity identification models are trained, disregarding the links between activities. Nonetheless, hierarchical linkages between actual actions exist based on sensor data similarity [15]. For instance, when considering four stationary classes, such as walking, ascending, and descending stairs, the three other categories might be regarded as abstract and non-stationary.

Many fields, such as healthcare, sports, tactical awareness, fall detection, and accident identification employ a broad range of HAR solutions [5,10,11]. In order to track vehicle movement for the purpose of accident prevention, current smartphone-based applications and research rely on GPS transceivers [16]. A vehicle is in motion if its GPS coordinates reveal a considerable shift. Nevertheless, these GPS-based systems cannot detect slight displacements, preventing the incident detection approach from activating if GPS coordinates do not move beyond a specific threshold. Therefore, a driver must be spotted as soon as they enter a vehicle, without the car going a significant distance. Multiple benefits might result from this kind of early detection, including the launch of an autonomous or innovative agent-based accident warning system and increased situational awareness [4]. An intelligent agent is a self-aware entity that acts upon its surroundings by observing it using sensors and then actuating it. For instance, a smartphone application uses built-in sensors to detect human behavior or any significant event.

We use deep learning neural networks and smartphone sensor data to solve the abovementioned issues to recognize driver behavior. We unveiled the

DriveNeXt deep learning model, inspired by the ResNeXt image classification framework. To validate the effectiveness of the presented model, we utilized a publicly standard dataset consisting of smartphone sensor data for various driving actions. This paper's essential contribution can be defined as follows: 1) To introduce a unique deep learning classifier based on multi-branch aggregation transformation, 2) To determine the optimal window size for recognizing driver behaviors, and 3) To analyze the effectiveness of several deep learning classifiers using the benchmark dataset.

The remaining parts of the work are arranged as follows. New research of relevance is included in Sect. 2. The study's underlying model, a branch of CNN, is described in Sect. 3. The results of our studies are presented in Sect. 4. The report finishes with a consideration of necessary future studies (Sect. 5).

2 Related Works

The deep learning method has seen widespread implementation to overcome machine learning's shortcomings. When using deep learning, feature extraction is efficient, meaning fewer people need to be involved. Many deep learning models have been presented for identifying human activities, presenting promising findings and a unique learning technique [8,12]. The majority of suggested models use standard CNNs.

According to [18], a CNN model is meant to analyze three-dimensional accelerometer data without considerable preprocessing. Before sending the input to the initial convolution layer, all information is preprocessed using the sliding window approach, and the accelerometer data is normalized. The normalized data are then given to the one-dimensional convolution and max-polling layers. The researcher proposes performing model evaluation using the WISDM standard dataset. Experimental findings indicated that the presented model could achieve significant precision while preserving reasonable computing costs. A CNN with several channels was proposed to unravel the motion detection issue in exercise programs' environment [1]. This study implements a self-collected dataset of 16 events from the Otago training schedule. Multiple sensors are installed on body parts to collect inactivity data for different movements, with individual sensors feeding a distinct CNN channel. After CNN functions, the findings from all sensors will be analyzed individually to establish the optimal placement of sensors for improved lower-limb action recognition. Their findings suggest that many sensor configurations could be more efficient than just one.

A deep HAR network is developed, transforming movement-sensing input to a series of spectrum images before passing these images to two CNN models that have been separately trained [7]. Individually CNN representative incorporates the image sequencing produced by the accelerometer and gyroscope. An ensemble of trained CNNs is used to make an informed guess about the kind of human behavior being observed. This research employs the Real-world Human Activity Recognition (RWHAR) dataset. This dataset includes eight actions: descending and ascending stairs, laying, standing, seated, running, leaping, and

walking. Using the proposed model, an F-score of 0.78 is possible during static and dynamic activities and 0.87 during vigorous activity. The researchers furthermore concluded that the model could effectively process image information. The model generalization is promising, but its accuracy performance is not equivalent to that of the other standard deep learning model. In [2], three ways for using the temporal features of a series of windows are provided. The first technique involves calculating the CNN model's average of the input windows. In another technique, the window series is given to a coincident CNN, which determines the action category established on the intermediate scores. The last approach resembles the second approach, and the learned characteristics are blended using a global intermediate pooling layer to obtain the last forecast.

Compared to using a single CNN classifier, it has been hypothesized that using an ensemble of CNN might improve the accuracy of motion recognition. Zhu et al. [21] introduced a CNN-based framework for HAR by combining several smartphone-based sensors, including a magnetometer, accelerometer, and gyroscope. The suggested technique is an ensemble of two different CNN standards. The first model of CNN is prepared to forecast action categories, and the second CNN is conditioned to concentrate on activity classes with many misclassifications. Employing weighted polling, the result of separate CNN models is then merged to forecast unexpected behaviors. The testing outcome reveals that this suggested model could attain an accuracy of 96.20%.

Also, [19] recommended using an ensemble model with three separate CNN models. The ensemble model computes the final result by averaging the results of the three CNN models. Before assembling each CNN for actual interpretation assessment, researchers investigated the effectiveness of every CNN model. The experimental outcome suggests that the ensemble model outperforms the three CNNs with a precision of 94.00%. This finding demonstrated that this learning approach could generalize how the weak learner's learning influence can be enhanced to increase the overall model. A two-channel model of CNN for action recognition is presented in [14]. The presented approach improves identification accuracy using sensor inputs' frequency and power characteristics. The model's accuracy was 95.30% when experimented on the publicly available UCI-HAR dataset. This technique has the disadvantage of requiring the extraction of specific characteristics to enhance movement detection from sensor data. Applying the attention mechanism module to identify the importance of the features enhances the effectiveness of the CNN model [20]. In order to capture the local features, the three acceleration inputs are transmitted concurrently to three convolutional layers with varying filter sizes. The attention mechanism then calculates how important each feature is to select the most useful ones. The model was validated using the public WISDM dataset, which performed with a 96.40% success rate.

3 Sensor-Based HAR Methodology

Data acquisition, preprocessing, data generation, model training, and evaluation are the four main operational phases in the sensor-based HAR methodology used in this investigation (see Fig. 1).

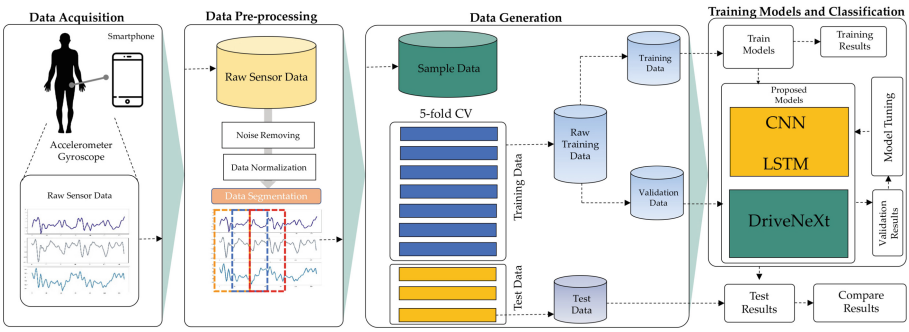

Fig. 1. The proposed HAR methodology

3.1 Driver Activity Dataset

This study uses a public dataset called "Driver entrance into and leaves from a vehicle using smartphone sensors," which records when a driver enters and leaves an automobile while their phone is in their pocket [3]. Participants performed the driving duties of:

- Grabbing the child safety seat (designated IN).
- Seated for some time (designated SITTING).
- Exit the automobile (designated OUT).
- Waiting a little while, perhaps 2 or 3 s, with the phone in the left pocket and the screen towards the thigh as you stand outside the automobile (designated STANDING).

Xiaomi Redmi Note6-Pro smartphones running Android 8.1 were utilized to gather data for the dataset. It features many sensors, including the gyroscope and accelerometer which were employed for data collecting. To acquire these signals, we relied on the Android program Sensor Kinetics Pro, which records data from the three-dimensional sensors at a sampling frequency of more than 400 Hz and provides information on the gravitation, linear acceleration, and spinning of the sensors.

This dataset sampled major features including acceleration, gravitation, direction, linear acceleration, and rotational across all three axes at a rate of 50 Hz.

3.2 Data Pre-processing

In data pre-processing, noise removal and data normalization were performed on unprocessed sensor data. The pre-processed data of the sensor were then separated by utilizing fixed-width sliding windows of 1 to 5 s with a 50% overlap ratio.

3.3 The Proposed DriveNeXt Architecture

In this work, we devised a multi-branch aggregation strategy in response to the ResNeXt model [17]. This approach provides kernel feature maps of varying sizes as a contrast to concatenated in the InceptionNet model [6]. This significantly reduced the number of model parameters, enabling these interconnections to be suitable for edge and low-latency processes.

Three convolutional kernel dimensions are represented in the DriveNeXt model's three components. There are three unique kernel dimensions (1×3, 1×5, and 1×7) in each MultiKernel (MK) device. The sophistication of the network and the number of parameters are further reduced by using 1×1 convolutions before implementing these kernels. DriveNeXt specifications are shown in Fig. 2.

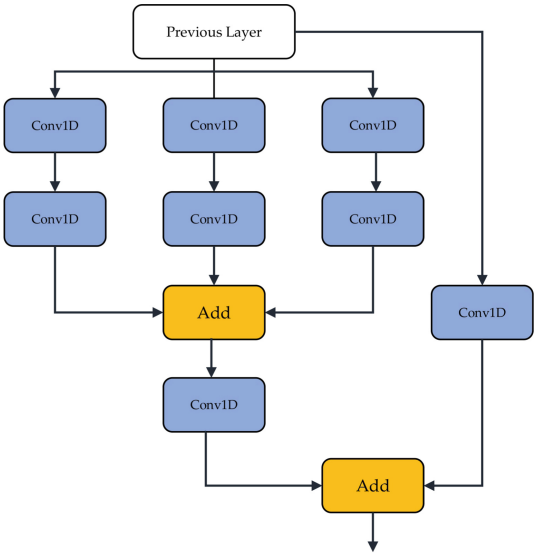

Fig. 2. An overview of MultiKernel component architecture

The DriveNeXt architecture has a minimal number of trainable parameters – just 23,653. The complete model is made up of six MK units, with the number of kernels being reduced to the desired number of classes using a 1×1 convolutional method. The layout of the DriveNeXt model is shown in Fig. 3.

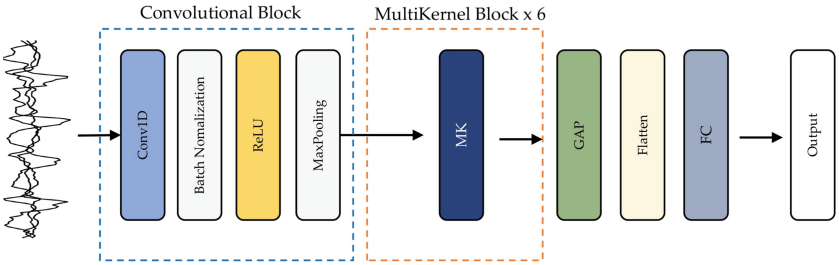

Fig. 3. The DriveNeXt architectural design

4 Experiments and Research Findings

In this section, we discuss the experimental setup and provide the experimental observations used to assess the standard CNN and LSTM models and the proposed DriveNeXt model for driving action detection based on smartphone sensor data.

4.1 Research Setting

Each investigation in this research is carried out on the Google Colab-Pro platform operating a V100 Tesla. Python is used for development in addition to the packages TensorFlow, Keras, Scikit-Learn, Numpy, and Pandas.

4.2 Research Findings

Table 1 displays the average accuracy and F1-score of DriveNeXt and benchmark models. Each model in the table employed varying sizes (1 s to 5 s) of sliding window data to train and test models with a 5-fold cross-validation process.

Based on Table 1, the findings suggest that the DriveNeXt model outperformed CNN and LSTM benchmark models for all window sizes. The DriveNeXt model achieved the most significant performance of window size of 1 s with an accuracy of 96.95%.

Table 1. Identification effectiveness of baseline models compared with the proposed DriveNeXt model

Classifiers	Window size (s)	Identification effectiveness		
		Accuracy	Loss	F1-score
CNN	1	95.49% (±0.411%)	0.29 (±0.058)	95.38% (±0.405%)
	2	94.26% (±1.134%)	0.32 (±0.052)	94.05% (±1.165%)
	3	92.43% (±1.576%)	0.39 (±0.117)	92.09% (±1.607%)
	4	91.63% (±1.047%)	0.41 (±0.072)	90.62% (±1.155%)
	5	90.45% (±2.441%)	0.56 (±0.147)	88.44% (±2.847%)
LSTM	1	96.62% (±0.313%)	0.16 (±0.029)	96.49% (±0.339%)
	2	93.48% (±1.225%)	0.41 (±0.036)	93.27% (±1.225%)
	3	92.74% (±0.793%)	0.32 (±0.070)	92.49% (±0.781%)
	4	92.77% (±1.761%)	0.43 (±0.106)	91.88% (±2.037%)
	5	92.51% (±1.901%)	0.20 (±0.062)	91.12% (±2.086%)
DriveNeXt	1	96.95% (±0.509%)	0.18 (±0.057)	96.82% (±0.549%)
	2	95.57% (±0.446%)	0.21 (±0.089)	95.40% (±0.491%)
	3	95.20% (±1.076%)	0.17 (±0.062)	94.95% (±1.134%)
	4	93.53% (±1.928%)	0.18 (±0.042)	93.09% (±2.141%)
	5	92.81% (±1.779%)	0.37 (±0.127)	91.26% (±2.130%)

5 Conclusion and Future Works

In this study, smartphone sensor-based identification of driving activity was investigated. We proposed the DriveNeXt deep residual model to achieve the study objective for driving behavior identification. Models were trained and tested to measure detection capability using a publicly available benchmark dataset. CNN and LSTM are the two baseline deep learning models used to compare the DriveNeXt model. The experimental findings demonstrate that the DriveNeXt has the most outstanding performance for all sliding window data sizes. The DriveNeXt model is successful at recognizing driver behavior.

In future research, we want to investigate driving action detection using different kinds of deep learning networks, including ResNet, InceptionTime, Temporal Transformer, etc.

Acknowledgment. This research project was supported by the Thailand Science Research and Innovation fund; the University of Phayao (Grant No. FF65-RIM041); National Science, Research and Innovation (NSRF); and King Mongkut's University of Technology North Bangkok with Contract No. KMUTNB-FF-66-07.
The authors also gratefully acknowledge the support provided by Thammasat University Research fund under the TSRI, Contract No. TUFF19/2564 and TUFF24/2565, for the project of "AI Ready City Networking in RUN", based on the RUN Digital Cluster collaboration scheme.

References

1. Bevilacqua, A., MacDonald, K., Rangarej, A., Widjaya, V., Caulfield, B., Kechadi, T.: Human activity recognition with convolutional neural networks. In: Brefeld, U., et al. (eds.) ECML PKDD 2018. LNCS (LNAI), vol. 11053, pp. 541–552. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-10997-4_33
2. Gil-Martín, M., San-Segundo, R., Fernández-Martínez, F., Ferreiros-López, J.: Time analysis in human activity recognition. Neural Process. Lett. **53**(6), 4507–4525 (2021). https://doi.org/10.1007/s11063-021-10611-w
3. Hirawat, A.: Driver entry into and exit from a car using smartphone sensors. https://data.mendeley.com/datasets/3czshz7zpr/1, https://doi.org/10.17632/3czshz7zpr.1. Accessed 01 July 2022
4. Hirawat, A., Bhargava, D.: Enhanced accident detection system using safety application for emergency in mobile environment: SafeMe. In: Das, K.N., Deep, K., Pant, M., Bansal, J.C., Nagar, A. (eds.) Proceedings of Fourth International Conference on Soft Computing for Problem Solving. AISC, vol. 336, pp. 177–183. Springer, New Delhi (2015). https://doi.org/10.1007/978-81-322-2220-0_14
5. Hnoohom, N., Mekruksavanich, S., Jitpattanakul, A.: An efficient resnetse architecture for smoking activity recognition from smartwatch. Intell. Autom. Soft Comput. **35**(1), 1245–1259 (2023). https://doi.org/10.32604/iasc.2023.028290
6. Ismail Fawaz, H., et al.: InceptionTime: finding AlexNet for time series classification. Data Min. Knowl. Discov. **34**(6), 1936–1962 (2020). https://doi.org/10.1007/s10618-020-00710-y
7. Lawal, I.A., Bano, S.: Deep human activity recognition using wearable sensors. In: Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments, PETRA 2019, pp. 45–48. Association for Computing Machinery, New York (2019). https://doi.org/10.1145/3316782.3321538
8. Mekruksavanich, S., Hnoohom, N., Jitpattanakul, A.: A hybrid deep residual network for efficient transitional activity recognition based on wearable sensors. Appl. Sci. **12**(10), 4988 (2022). https://doi.org/10.3390/app12104988
9. Mekruksavanich, S., Jitpattanakul, A.: Deep learning approaches for continuous authentication based on activity patterns using mobile sensing. Sensors **21**(22), 7519 (2021). https://doi.org/10.3390/s21227519
10. Mekruksavanich, S., Jitpattanakul, A.: Multimodal wearable sensing for sport-related activity recognition using deep learning networks. J. Adv. Inf. Technol. **13**(2), 132–138 (2022). https://doi.org/10.12720/jait.13.2.132-138
11. Mekruksavanich, S., Jitpattanakul, A.: Sport-related activity recognition from wearable sensors using bidirectional GRU network. Intell. Autom. Soft Comput. **34**(3), 1907–1925 (2022). https://doi.org/10.32604/iasc.2022.027233
12. Mekruksavanich, S., Jitpattanakul, A., Sitthithakerngkiet, K., Youplao, P., Yupapin, P.: ResNet-SE: channel attention-based deep residual network for complex activity recognition using wrist-worn wearable sensors. IEEE Access **10**, 51142–51154 (2022). https://doi.org/10.1109/ACCESS.2022.3174124
13. Noppitak, S., Surinta, O.: dropCyclic: snapshot ensemble convolutional neural network based on a new learning rate schedule for land use classification. IEEE Access **10**, 60725–60737 (2022). https://doi.org/10.1109/ACCESS.2022.3180844
14. Sikder, N., Chowdhury, M.S., Arif, A.S.M., Nahid, A.A.: Human activity recognition using multichannel convolutional neural network. In: 2019 5th International Conference on Advances in Electrical Engineering (ICAEE), pp. 560–565 (2019). https://doi.org/10.1109/ICAEE48663.2019.8975649

15. Silla, C., Freitas, A.: A survey of hierarchical classification across different application domains. Data Min. Knowl. Disc. **22**, 31–72 (2011). https://doi.org/10.1007/s10618-010-0175-9
16. White, J., Thompson, C., Turner, H., Dougherty, B., Schmidt, D.: WreckWatch: automatic traffic accident detection and notification with smartphones. Mob. Netw. Appl. **16**, 285–303 (2011). https://doi.org/10.1007/s11036-011-0304-8
17. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5987–5995 (2017). https://doi.org/10.1109/CVPR.2017.634
18. Xu, W., Pang, Y., Yang, Y., Liu, Y.: Human activity recognition based on convolutional neural network. In: 2018 24th International Conference on Pattern Recognition (ICPR), pp. 165–170 (2018). https://doi.org/10.1109/ICPR.2018.8545435
19. Zehra, N., Azeem, S.H., Farhan, M.: Human activity recognition through ensemble learning of multiple convolutional neural networks. In: 2021 55th Annual Conference on Information Sciences and Systems (CISS), pp. 1–5 (2021). https://doi.org/10.1109/CISS50987.2021.9400290
20. Zhang, H., Xiao, Z., Wang, J., Li, F., Szczerbicki, E.: A novel IoT-perceptive human activity recognition (HAR) approach using multihead convolutional attention. IEEE Internet Things J. **7**(2), 1072–1080 (2020). https://doi.org/10.1109/JIOT.2019.2949715
21. Zhu, R., et al.: Deep ensemble learning for human activity recognition using smartphone. In: 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP), pp. 1–5 (2018). https://doi.org/10.1109/ICDSP.2018.8631677