

Article

Real-Time Littering Activity Monitoring Based on Image Classification Method

Nyayu Latifah Husni ¹, Putri Adelia Rahmah Sari ¹, Ade Silvia Handayani ^{1,*}, Tresna Dewi ¹,
Seyed Amin Hosseini Seno ², Wahyu Caesarendra ^{3,*}, Adam Glowacz ^{4,*}, Krzysztof Oprzędkiewicz ⁴
and Maciej Sułowicz ⁵

- ¹ Electrical Engineering, Politeknik Negeri Sriwijaya, Jalan Sriwijaya Negara, Bukit Besar, Palembang 30139, Sumatera Selatan, Indonesia; nyayu_latifah@polsri.ac.id (N.L.H.); putdelia123@gmail.com (P.A.R.S.); tresna_dewi@polsri.ac.id (T.D.)
- ² Department of Computer Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Azadi Square, Mashhad 9177948974, Iran; hosseini@um.ac.ir
- ³ Faculty of Integrated Technologies, Universiti Brunei Darussalam, Jalan Tungku Link, Gadong BE1410, Brunei
- ⁴ Department of Automatic Control and Robotics, Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, AGH University of Science and Technology, Al. A. Mickiewicza 30, 30-059 Kraków, Poland; kop@agh.edu.pl
- ⁵ Department of Electrical Engineering, Cracow University of Technology, Warszawska 24 Str., 31-155 Cracow, Poland; maciej.sulowicz@pk.edu.pl
- * Correspondence: ade_silvia@polsri.ac.id (A.S.H.); wahyu.caesarendra@ubd.edu.bn (W.C.); adglow@agh.edu.pl (A.G.)



Citation: Husni, N.L.; Sari, P.A.R.; Handayani, A.S.; Dewi, T.; Seno, S.A.H.; Caesarendra, W.; Glowacz, A.; Oprzędkiewicz, K.; Sułowicz, M. Real-Time Littering Activity Monitoring Based on Image Classification Method. *Smart Cities* **2021**, *4*, 1496–1518. <https://doi.org/10.3390/smartcities4040079>

Academic Editors:
Aneta Poniszewska-Maranda and
Witold Maranda

Received: 18 October 2021
Accepted: 7 December 2021
Published: 13 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: This paper describes the implementation of real time human activity recognition systems in public areas. The objective of the study is to develop an alarm system to identify people who do not care for their surrounding environment. In this research, the actions recognized are limited to littering activity using two methods, i.e., CNN and CNN-LSTM. The proposed system captures, classifies, and recognizes the activity by using two main components, a namely camera and mini-PC. The proposed system was implemented in two locations, i.e., Sekanak River and the mini garden near the Sekanak market. It was able to recognize the littering activity successfully. Based on the proposed model, the validation results from the prediction of the testing data in simulation show a loss value of 70% and an accuracy value of 56% for CNN of model 8 that used 500 epochs and a loss value of 10.61%, and an accuracy value of 97% for CNN-LSTM that used 100 epochs. For real experiment of CNN model 8, it is obtained 66.7% and 75% success for detecting littering activity at mini garden and Sekanak River respectively, while using CNN-LSTM in real experiment sequentially gives 94.4% and 100% success for mini garden and Sekanak river.

Keywords: public facilities; human activity recognition; littering; machine learning; CNN; LSTM

1. Introduction

The development of various sectors in Indonesia has been progressing very rapidly, one aspect of which is infrastructure development. However, this development is not equal to the consideration shown in maintaining their continuity. It is not commensurate with the level of public awareness to maintain it. Some public facilities cannot survive because of human indifference and irresponsible actions. To fix the damage, the government has to spend a lot of money, energy, and time.

Palembang, as the capital of South Sumatra, has infrastructure development programs as well, including the revitalization and restoration of the Sekanak River that is located in the Sekanak area. This area is included as a cultural heritage area [1,2] which has many historical buildings, including Sekanak Market, Kantor Ledeng, the Jacobson Building, KBTR, and HokTong. In addition, this area is also close to Benteng Kuto Besak (BKB), Sultan Mahmud Badarrudin Jayo Wikramo Great Mosque, Musi River, industrial jumputan

and songket center, pempek center, and Palembang mattress center. All of these are part of the cultural heritage and local wisdom of Palembang City.

The restoration of the 11 km long Sekanak River is continuing (currently only 800 m has been completed) so that, in 2023, it is intended that the Sekanak River will not only be restored but will also become a new tourist destination in Palembang. However, as mentioned above, the development that has cost a lot of money and energy has not been followed by public awareness to maintain it. This can be seen from the fact that several parts of the Sekanak River restoration have been damaged.

The urgency of research to overcome the problems mentioned above has arisen. To protect against improper human actions in the environment and existing public facilities, the author was inspired to propose a monitoring device that can classify human activities in a public environment. Using this device, humans will be forced to obey the rules and so will be reluctant to commit criminal action. The existence of coercion will force the humans to realize the importance of protecting the surrounding environment. In addition, it will cause them to become accustomed to do so.

This study uses information technology that is integrated into smart village technology, where the devices offered not only utilize the IoT system but also use artificial intelligence in its application. The specific purpose of this research is to monitor the situation in the Sekanak area, which includes monitoring the Sekanak River (27 Ilir district) and monitoring the mini garden near the Sekanak market (28 Ilir district). Monitoring the mini garden is necessary due to there being many residents who are littering in that location.

Smart Village technology forms part of a smart city, which aims to provide flexibility for a village to solve its own problems intelligently. This technology is used by researchers in solving problems of damage and cleanliness in the villages of 27 Ilir and 28 Ilir (as stated above). This research was initiated not only to focus on the sophistication of the technology offered but to change the condition of the local community into a better, safer, and more prosperous state, as well as to raise public awareness of the importance of innovation and creativity to maintaining and to developing their village.

The contributions of this research are the datasets for the image classification, especially regarding the activity of littering and the implementation of the system in the real environment, so that the pioneer smart village in the 27 Ilir Palembang can be achieved.

2. Related Work

Smart Village technology is one of the concepts that is used to solve the problems of villages. This technology has been widely used in the areas of agriculture [3], health, government, transportation, and security [4]. Smart village technology in this study is combined with Internet of Things (IoT) technology, where the data generated by several sensors will be used for certain services [5], in this case for monitoring human activities. IoT will allow people and things to be connected. It will be connected in "Any time", "Anywhere", with "Anything" and with "Anyone". In addition, ideally, the connection is made using "Any" paths/networks and "Any" services. This IoT technology allows the formation of new services or the re-establishment of existing services in a previous smart village [6]. In addition to IoT, this research also uses artificial intelligence that is connected to the camera. With this intelligence, criminal acts can be prevented [2].

According to [7], Smart village research has been applied in many areas, such as: smart health and education systems, smart energy management systems, smart safety systems, etc. Moreover, smart village technology has been successful applied in improving the sustainability of the rural environment, as stated in [8–11]. The researchers used the smart village concept to achieve sustainability and resilience of rural areas. In addition, they also used that concept to strengthen the relations among the rural communes to closed cities and towns [8]. The concept also covers the management of the centralization power of the government [11], while in [10] the smart village concept focused on the role of technology in building governance and public services. The smart village [12] concept has also been successful in its application for revitalizing the demographic of a community. In

this research, the smart village concept is applied to protect the beauty of the infrastructure that has been built by the government, especially in Sekanak area.

Besides smart villages, one more thing that relates to this research is human activity classification. The recent research on human activity recognition and classification that was conducted by the researchers have been applied in many parts, for instance as presented in Table 1. The usage of the EFTS and IMU in paper [13] helped the authors to analyze the activities of the football players, such as remaining stationary, walking, jogging, running, slow turning, and fast turning. In references [14,15], the author conducted research by detecting the fall by using Channel State Information (CSI) to recognize the activity of falling [14] and used wearable sensors, such as an accelerometer and a groscope to detect the fall [15]. They have a simulation accuracy of about 93.2%. Xinyu Li in [16] used CNN-LSTM to recognize concurrent activities. Human activity was also investigated by Thomas Stadelmayer in [17]. They used radar to help the authors record the data of daily human activities. Their proposed work has been successful in reaching an accuracy of 99.5%. Besides the implementation above, the research in [18] applied CNN to detect the driving activity. The safety of the driver was enhanced by using a camera and the method proposed. In [19], the author conducted an overview of the use of deep learning methods to solve the problem of human recognition. They also highlighted future issues that can be analyzed. One of their ideas that is really interesting is how to predict future activities. Djamila et al. in [20], tried to conduct human activity recognition using a vision based method. They explored so many articles regarding human activity recognition and presented a lot of methods and steps that can be used in solving human activity recognition, such as detection, tracking, and classification.

Table 1. Recent research on activity recognition and classification.

No.	Implementation	Auxiliary Components	Method	Ref.
1.	Football activities	EFTS and IMU	DCNN	[13]
2.	Fall detection	CSI	CNN	[14]
3.	Recognizing concurrent activities	Wearable sensor	k-NN, SVM, DT, EC	[15]
4.	Human Activity	Multiple Sensors	CNN-LSTM	[16]
		Radar	Parametric CNN	[17]
5.	Driving recognition	Sensor-based	Deep Learning	[19]
		Vision-based	Machine Learning	[20]
		Camera	CNN	[18]

From Table 1 it can be concluded that the CNN becomes one of the most useful solution in differentiating human activities. The application of the CNN has also been used in detecting the sports activity [21], classifying the posture of sows [22], surveillance video [23], violence video [24], micro RNA [25], human activity recognition [26–28]. The researchers in [29] conducted the research of activity recognition to help the elderly to manage their lives by themselves. The author focused on the activities that were conducted by the users in the small kitchen in their laboratory. In that paper, they claimed that they were the first who combined three inputted data that came from videos, inertial measurements Units (IMUs), and ambient sensors. However, due to the need for sensors, their usage is limited. They could only be used for monitoring the activity of certain persons who were wearing the device and cannot be applied to public users who are always exposed to fully uncertain environments.

The researchers in paper [30] also used IMUs in their research. They combined those sensors with smart cigarette lighters, proximity sensors, and respiration sensors to compose complete systems of monitoring smoking behavior. The smoking activity was analyzed from the movement of the hand sequentially and the pattern of the breath. They claimed that the method of CNN LSTM that they proposed is robust enough to analyze the puffing with an accuracy of 78%. However, this research is almost the same as research in [17]

in which it was not successful when it was run without the help of the sensors that were attached to the user's body.

The performance of the CNN is improved when more depth is added to the CNN; however, this improvement leads to low accuracy of the system [19]. However, they stated that additional attention should also be paid to the weight layers of the networks [31]. Human activity recognition has also been conducted by Ankita [20] et al. They have proven that the use of CNN LSTM was successful in reaching an accuracy of 97.89%. They claimed it was superior compared [32] to the Feed Forward Convolutional Network (FFCN) and Principle Component Analysis-Bidirectional Long-Short Term Memory (PCA-BiLSTM) methods that have accuracy of 97.64% and the Convolutional Neural Network (CNN) that has accuracy of 97.01%. However, this research has not been implemented in a real environment.

3. Materials and Methods

The location that is the focus of this research is close to several historical buildings, as shown in Figure 1.

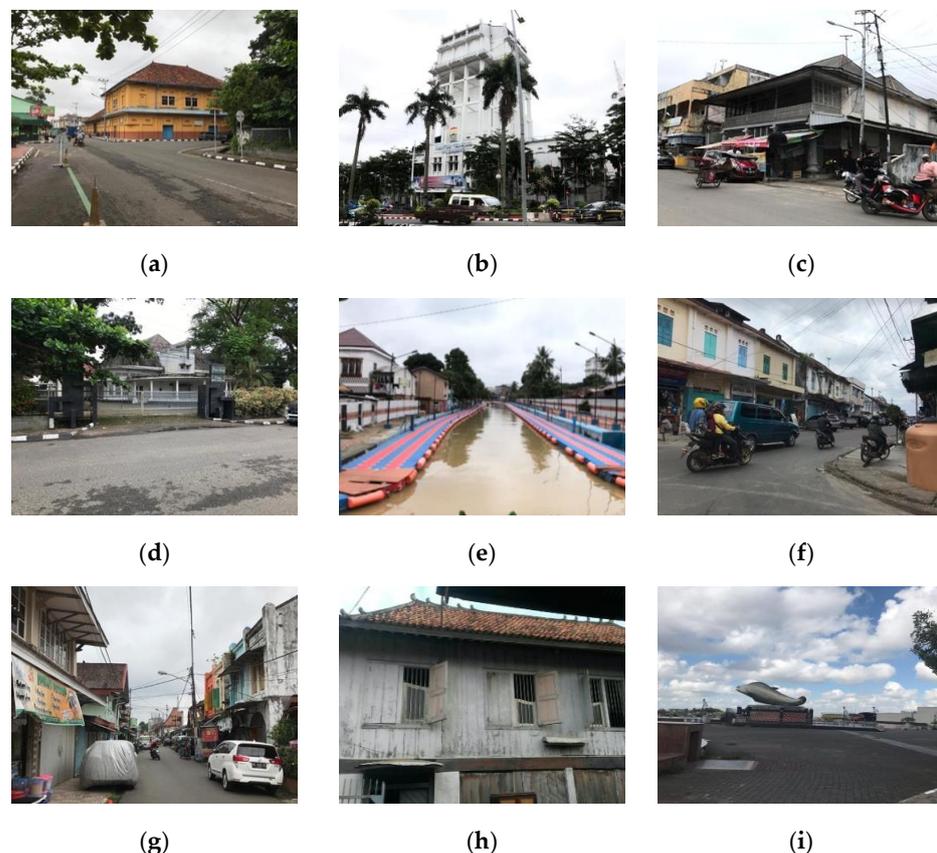


Figure 1. Historical buildings near Sekanak area. (a) Jacobson Building, (b) Kantor ledeng, (c) HokTong, (d) KBTR, (e) Sekanak River, (f) Sekanak market, (g) Dutch building, (h) Limas, (i) Benteng Kuto Besak.

There is the Jacobson building that functioned as a trading Dutch company in 1960, as presented in Figure 1a, Kantor Ledeng in Figure 1b that functioned as the water reservoir building which has become the Palembang city official government building, Hok Tong in Figure 1c that functioned as a manufacturer of rubber products, Kuto Besak Theater Restaurant KBTR in Figure 1d that is located behind the Kantor Ledeng, Sekanak River in Figure 1e, Sekanak market in Figure 1f, the Dutch building in Figure 1g, Limas in Figure 1h, and a Palembang traditional house, and Benteng Kuto Besak in Figure 1i that originally functioned as the palace of the Palembang Darussalam Sultanate.

In this research, the littering activity monitoring system was applied in two places (as shown in Figure 2), namely (1) Sekanak River, as shown in Figure 2a, and (2) the mini garden that is located in front of Sekanak market, as shown in Figure 2b. These two spots were included as the Ilir Barat Dua region of Palembang, South Sumatra, Indonesia. These two spots are important for the researchers because these are located at the center of the Sekanak areas. When people walk around in Sekanak, they are likely to pass those places. Thus, the devices in this research should be placed in those areas to maintain the beauty of the Sekanak area.



Figure 2. The location of the littering activity monitoring. (a) Sekanak River, (b) A small garden.

3.1. Hardware

The mechanical design of the monitoring device in this research can be seen in Figure 3a. It consists of two main parts, namely (1) the solar cell, which includes a solar cell components' box and a pole; and (2) the electronic monitoring components box. The solar cell component box as shown in Figure 3b is filled with components, such as: (1) a battery that functions as the power source and the storage of the electric power that has been obtained from the solar cell. This battery has two electrodes that interact with sulfuric acid so that they change into lead sulfate. This produces current flow when the lead electrode lets some electrons free; (2) the SCC or Solar Charger Controller that is used to optimize and to guarantee that the lifetime of the battery can be upgraded. The SCC has 2 important modes, namely charging and operating. In charging mode, the SCC has a responsibility to charge and to maintain the battery so that it is not overcharged, while in the operating mode, the SCC is used to maintain the supply to the load. When the battery is almost empty, the SCC stops the supply; (3) the battery MCB (Miniature Circuit Breaker), which ensures that there is no short circuiting of the battery; (4) the MCB panel that functions as protection and as the guard against the current overload; (5) the MCB inverter that functions as the breaker for the solar panel in order to avoid short circuit and overload; (6) the inverter that converts DC to AC; (7) the LVD (Low Voltage Disconnect) that functions as the battery protection from the over-discharge. It stops the battery load when the battery is low and it automatically connects the battery load when the battery has been charged.

In the electronic monitoring components box (Figure 3c), the components are placed in two parts, namely the cover and the internal part. In the cover, there are 6 components, including (1) the MQ7 that detects the occurrence of dangerous gases; (2) the DHT22 sensor used to detect the surrounding humidity; (3) the webcam that captures the video; (4) the JSN-SR04 ultrasonic sensor that detects the Sekanak River water level; (5) the speaker that notifies people who litter; (6) the LCD that displays the data regarding the temperature, humidity, air quality, and water level. Meanwhile, in the internal part, there are components, including (1) the PC fan that ensures the mini PC remains at its normal temperature; (2) the Arduino Uno microcontroller that functions as the controller of the environmental sensors used in this research; (3) the mini breadboard that functions as the connector of the connecting cables; (4) the mini PC that functions as the signal processor. It has responsibility to send the obtained data to the router; (5) the router-modem that functions as the channel that sends and receives the data; (6) the volume controller that controls the produced voices.

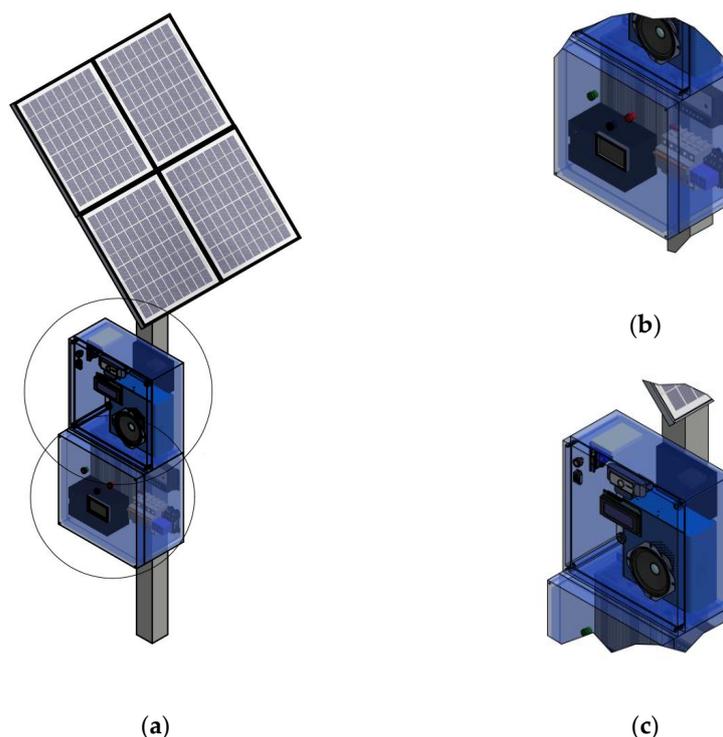


Figure 3. The hardware of the littering activity monitoring system: (a) the full hardware system, (b) the solar cell component box, (c) the electronic monitoring components box.

The block diagrams of the monitoring systems can be seen in Figure 4. Overall, the systems applied in the Sekanak River and mini garden consist of the same connection, as shown in Figure 4. However, the monitoring system in the mini garden has no waterproof ultrasonic sensor, as applied in the Sekanak River monitoring system. The power supply obtained from the solar cell is input into the Mini PC that connects to the Arduino and the webcam, which become the input for the Mini PC. The Arduino is the processor of the inputted sensors' data, i.e., from the Ultrasonic sensor, DHT 22, and MQ7. The data that have been processed by the Arduino is displayed on the LCD and sent to the Mini PC. On the other side, the webcam that captures the human activity near it is also has a connection to the mini-PC. The video captured by the webcam is processed by the mini-PC and then is sent to the cloud server through the Wi-Fi router which, then sends the final data to the users. The users can use their mobile phone, PC, or laptop to monitor the littering activity.

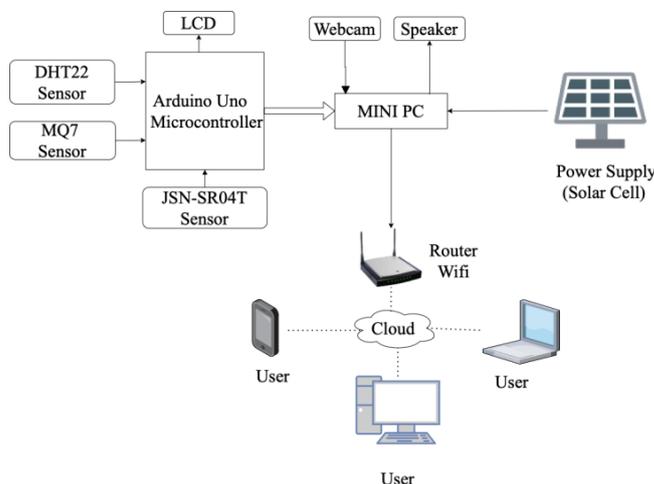


Figure 4. The hardware of the littering activity monitoring system.

3.2. Software

In this research, there were two methods used, i.e., the first using CNN only and the second using CNN-LSTM. The architecture of the CNN can be seen in Figure 5, while the CNN LSTM is presented in Figure 6.

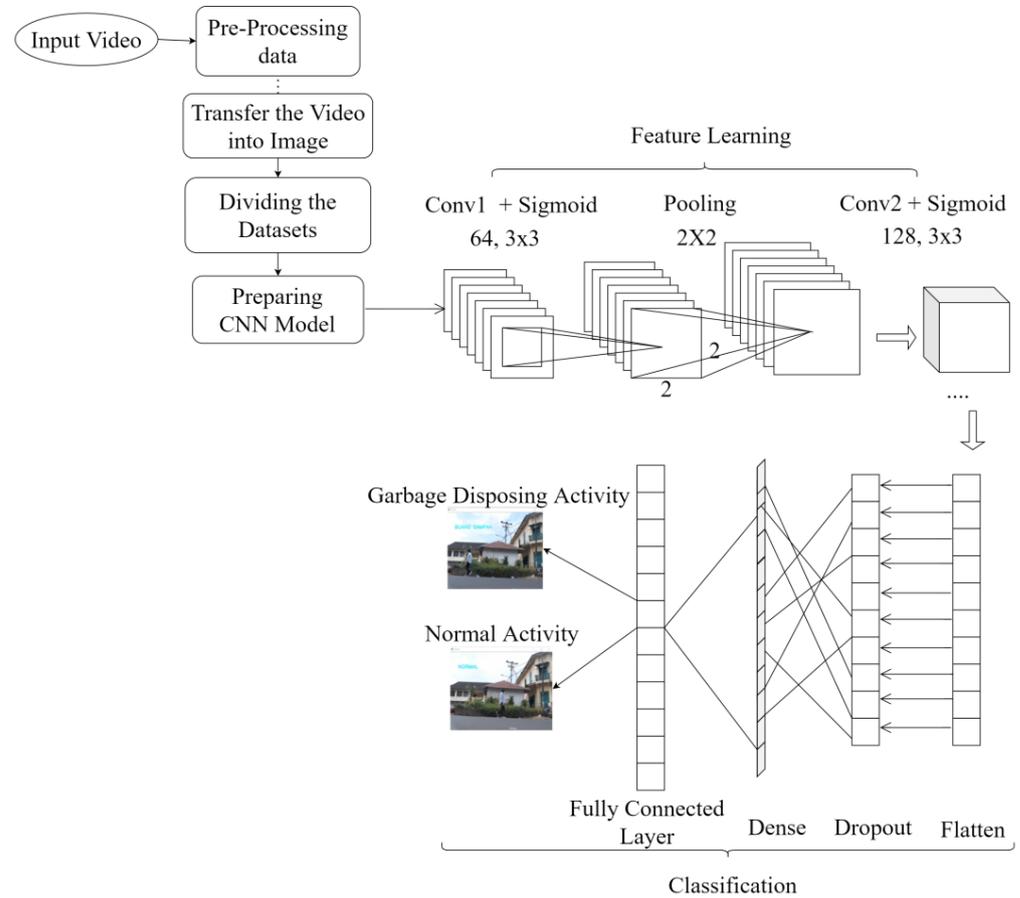


Figure 5. The CNN.

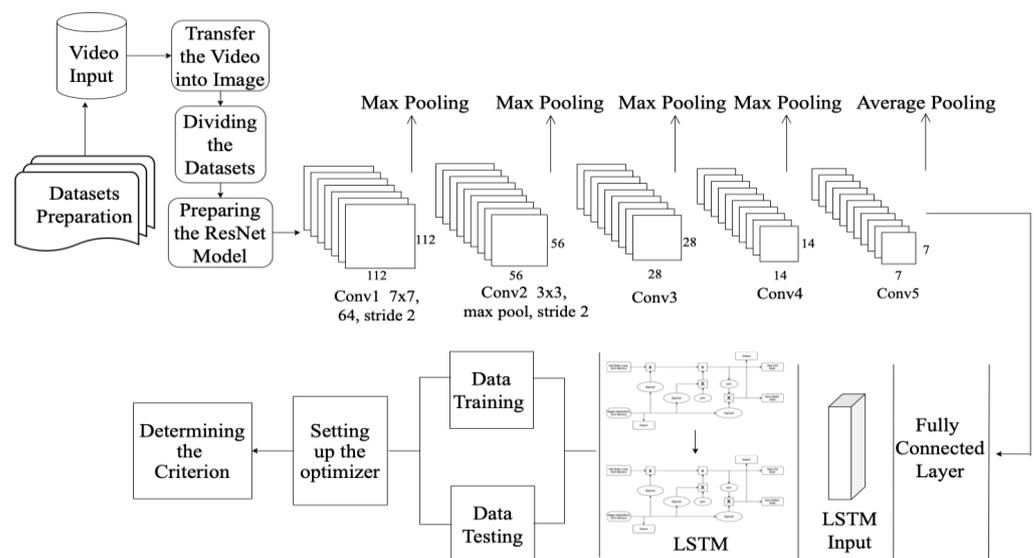


Figure 6. The CNN-LSTM.

The architecture of the CNN shown in Figure 5 consists of 3 main parts, i.e., (1) preparation, (2) feature learning, and (3) classification. The preparation includes inputting the video, conducting the data pre-processing, transferring the video into images, dividing the datasets, and preparing the CNN model. The process is continued to the feature learning, where the pooling is conducted between convolution 1 and convolution 2. After that the classification takes place. In this stage, the data obtained from the pre-processing process are flattened, dropped out, densified, and passed through the fully connected layer so that they can decide what activity is being performed.

In this research, videos that have been collected from two places, i.e., Sekanak River and the mini garden are processed in the data pre-processing. In this process, all video data enter the video extraction stage in which the video is extracted into several images. Videos that have durations up to 5 min are split into 173 jpg images with jpeg format RGB size 427×240 , 22.3 kb. Then, the video extraction results are placed in 2 prepared folders, namely the littering and normal folders. The system adds up all the images results that can be solved by the system. After the image is obtained from the video extractor, the image is separated. After that, the data enter the CNN model process, during which they enter the learning features process. In this process, the input that is ready to become a CNN model performs a convolution stage for 1 layer with a 3×3 kernel and 64 filter. This network activates the sigmoid at each layer. After that, the data are pooled 2×2 and continued to the second convolution using the kernel or filter sigmoid of size 128, 3×3 . Then, they enter the classification process, starting from the flattening process.

When a flat layer is formed, the vector value of 128 channels, size 3×3 is converted into a single vector form. After the multiplication of $3 \times 3 \times 128$ is calculated, there will be 1152 values that will enter the neural network. After the process has finished, it is continued to create a solid layer that is set to be 256 units. The resulting vector of 1152 values is entered one by one into 256 units. Thus, it will give $(1152 \times 256) + 256$ bias = 295,168 parameters. After that, the process is continued to the dropout process. It aims to prevent overfitting and to speed up the learning process. The system then temporarily removes the hidden neurons that have probability value between 0 and 1. The dropout for the previous 256 units is redecorated for the solid layer. Thus, the parameters to be generated are $(256 \times 2) + 2$ bias = 514 parameters. Thus, the total parameters performed by Machine Learning are 295,168. After the calculation is completed, it continues to a dense stage, in which, it is provided by adding a fully connected layer so that the data can be classified. The output is the information regarding normal or littering activity.

For the second method, a hybrid of a Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) was used as the intelligence for detecting the littering activity. In Figure 6, data preparation is carried out. Then, the process is continued with inputting the video. After that, it enters the process of the transferring video into images. Each video produces 173 RGB 427×240 jpeg images, 22.3 kb. Then, the video extraction results are saved into the provided folders, namely littering and normal folders. After that, the system prepares the Resnet, which is a classical neural network. Then, the data enter the feature learning stage process to determine the characteristics of each image that has been solved by the system. The convolution was conducted 5 times to obtain the best results. In this convolution process, every time the system convolutes the data, max-pooling is carried out to retrieve the largest image data. Then, the convolution is carried out again with different sizes. At the end of the convolution, average pooling is carried out, i.e., by calculating the average value of the feature patch obtained. Then, the data can be fully connected so that they become the input of the LSTM (long short-term memory) process. In this process the data are sorted by following the old context on the LSTM with the new context. Then, the data are divided into 2, namely training data and testing data. After that, the system prepares an optimization process and the data are assessed by the class system classifier.

Using CNN-LSTM, the researchers can obtain information from the entire scale of objects so that they can classify objects more accurately. The steps that should be conducted in this research can be described as follows:

- Dataset preparations were obtained by collecting videos of littering activity and non-littering activity. There were 400 videos that consisted of 200 videos for littering and 200 videos for non-littering activities. The non-littering activities in this research were categorized as normal activities.
- Transferring the videos into images.

The next step was to transfer the video into images. Each video obtained in the data preparation was then converted into about 100–300 images as shown in Figure 7.

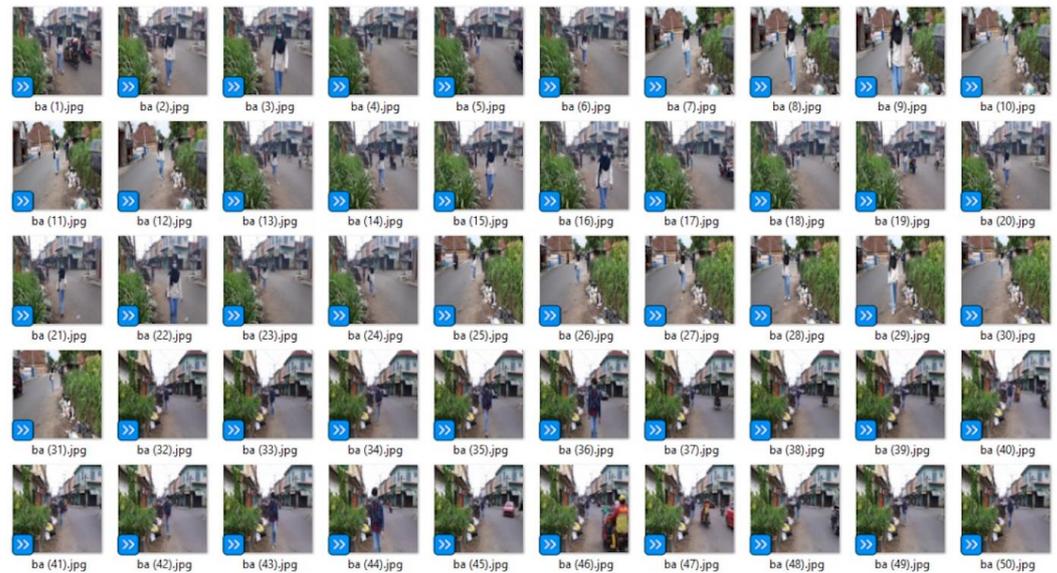


Figure 7. Image samples obtained from the video conversion.

- Dividing the datasets.

The datasets were then divided into the testing data and training data. In this research, the tests data size was 0.2 and the training data size was 0.8.

- Preparing the ResNet model.

The model used was Resnet 101 [33], in which the model used was the Residual CNN for classifying the images obtained before. In Resnet 101, there are 101 layers that are divided into 3-layer blocks. The specification of the architecture layer can be seen in Table 2. ResNet works by inserting the shortcut connection so that the network becomes the version of the counterpart residue. When the input and the output of the networks are in the same dimension, then the identity shortcut, $F(x\{W\} + x)$, can be directly used. However, when the shortcut is different, the system makes the system become identical by increasing the dimension using extra zero entries or the projection shortcut in $F(x\{W\} + x)$ is used in order to match dimensions using 1×1 convolution.

- Setting up the fully connected layer.

In this research, one fully connected layer was set up, in which this layer has 1000 neurons. This fully connected layer has predicted the next image.

- Setting up the LSTM.

The LSTM was designed using 300 inputs, 256 hidden sizes, and 3 layers blocks.

The LSTM architecture used in this research can be seen in Figure 8. The equations used for each element in the sequence are as follows:

$$i_t = \sigma(W_{ii}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi}) \quad (1)$$

$$f_t = \sigma(W_{if}x_t + b_{if} + W_{hf}h_{t-1} + b_{hf}) \quad (2)$$

$$g_t = \tanh(W_{ig}x_t + b_{ig} + W_{hg}h_{t-1} + b_{hg}) \quad (3)$$

$$o_t = \sigma(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho}) \quad (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (5)$$

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$

where:

h_t is the hidden state at the time t ;

c_t is the cell state at the time t ;

x_t is the input at the time t ;

h_{t-1} is the hidden state at the time $t - 1$;

i_t, f_t, g_t, o_t are the input, forget, cell, and output gate;

σ is the sigmoid function;

\odot is the Hadamard product.

- Setting up the loader for the data training and data testing.

The loader used was a data loader from pytorch and is intended for preparation of the data so that they are ready to be trained and tested. The most important thing in this set-up is the dataset that will be processed. In this research, the video that is converted to the pictures becomes the dataset. This dataset is then processed using an iterable-style dataset.

- Setting up the optimizer.

The optimizer used in this research is torch.optim. It accelerates the training and testing process so that they can achieve the effective value quickly. The optimizer object used holds the current state and updates parameters.

- Determining the criterion.

This criterion determination is useful in balancing the training set that is used. The input for this step is the raw data and the target of the criterion is in the class indices of the range $[0, C - 1]$, where C is the number of the class.

Figure 8 below shows the architecture of the LSTM. In Figure 8, the data that have been prepared and have passed through the CNN process are input into the LSTM process. The data are connected to the cell state/long-term memory at the top of the LSTM module. The system performs multiplication and addition operations so that the data become a new cell state. This initial process is assisted by the existence of a sigmoid gate which regulates how much information can pass. Then, the system decides which information can pass after obtaining a new cell state. At this stage, there are 2 parts, namely the sigmoid gate which first decides which value to be updated. Then, the tanh layer generates a new context vector candidate, or a new cell state vector candidate. After that, it combine the two and update the context again. The next step is to update the old context or long-term memory to the new cell state by multiplying the new cell state by sigmoid to determine how many candidates the system will include in the new context. Then, the system adds up the long-term memory with the new cell state. The output value obtained at this stage is based on the context value that has been passed to a filter. The first thing is that the system runs a sigmoid gate to determine which parts of the context the system generates. Then, the system passes through the tanh layer to make the values -1 and 1 . At the end, the system is multiplied by the sigmoid gate output so that the system determines the part that can be disconnected.

Table 2. Specification of the ResNet 101 used in this research.

Layer Name	Output Size	101 Layer
Conv1	112×112	$7 \times 7, 64, \text{stride } 2$
		$3 \times 3, \text{max pool, stride } 2$
Conv2_x	56×56	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
Conv3_x	28×28	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
Conv4_x	14×14	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$
Conv5_x	7×7	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	Average pool, 1000 d-fc, softmax
FLOPs		

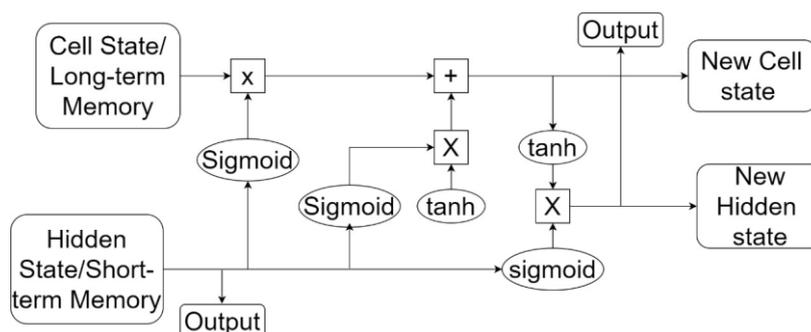


Figure 8. LSTM architecture.

4. Results and Discussion

4.1. CNN Experiment

The test was conducted in several stages to measure the performance of the model proposed. At first, the data model was tested to analyze the accuracy of the system in understanding the collected data. The test only used CNN as the intelligence without including the LSTM in the experiments. The accuracy of the system to the datasets was tested using eight different models. The specification of each model is shown in Table 3. For the first and the second model, the training phase used 100 videos of littering and 100 data of normal activities in the Sekanak River and the mini garden, respectively. However, the training was found to be erroneous despite it having been trained for 14 days. This occurred due to the processor used not supporting the training process. In the next experiment, the training data of model 3 were processed using more robust processor; however, the training that was conducted over 14 days became failed and it produced accuracy that was still very low, 49%, and a high loss, 410.19%. Then, the model was changed using model 4, which then produced an accuracy of 100%; however, the loss was still high, namely 55%. In model 5, the epoch was 500 with about 500 total datasets. However, the output was still an error. The error occurred when the script was edited to add the other layer.

Table 3. Experimental data.

Model	Activation	Epoch	Accuracy (%)	Loss (%)	Duration (Days)	Output
1	ReLU	200	-	-	14	Error
2	Sigmoid	100	-	-	14	Error
3	ReLU	100	49	410.19	14	Failed
4	ReLU	500	100	55	3	Error
5	ReLU	500	-	-	3	Stopped
6	Sigmoid	500	100	49	3	Failed
7	ReLU	500	56	77.1	3	Failed
8	Sigmoid	500	56	70	3	Success

The training was then continued until it reached an accuracy of 100% and loss of 49% using activation of the sigmoid in model 6 with a NVIDIA GEFORCE GTX 1080 TI GPU processor. Table 4 shows the properties of the training models. Although the loss was still high, due to the high accuracy obtained in model 6, it was then tested in the real environment. The experimental data are shown in Figure 9. Figure 9a–c is the implementation of the device in the mini garden, while Figure 9d–f is the pictures in the Sekanak River. However, in these experiments, the device still could not recognize the activity of littering. Although in Figure 9a–f there were littering activities, the device still considered the activity a normal one. Thus, it means that the system has not perform well.

Table 4. Properties of training using model 6, model 7, and model 8.

Parameters	Model 6	Model 7	Model 8
Name:	cnn_model6	cnn_model7	cnn_model8
Epoch:	500	500	500
Activation function:	Sigmoid	ReLU	Sigmoid
Input Shape:	3030 × 300	3030 × 300	3030 × 300
Pooling Size:	2 × 2	2 × 2	2 × 2
Accuracy:	100%	56%	56%
Loss:	49%	77.1%	70%
Time:	23s85ms/step	54s84ms/step	53s84ms/step

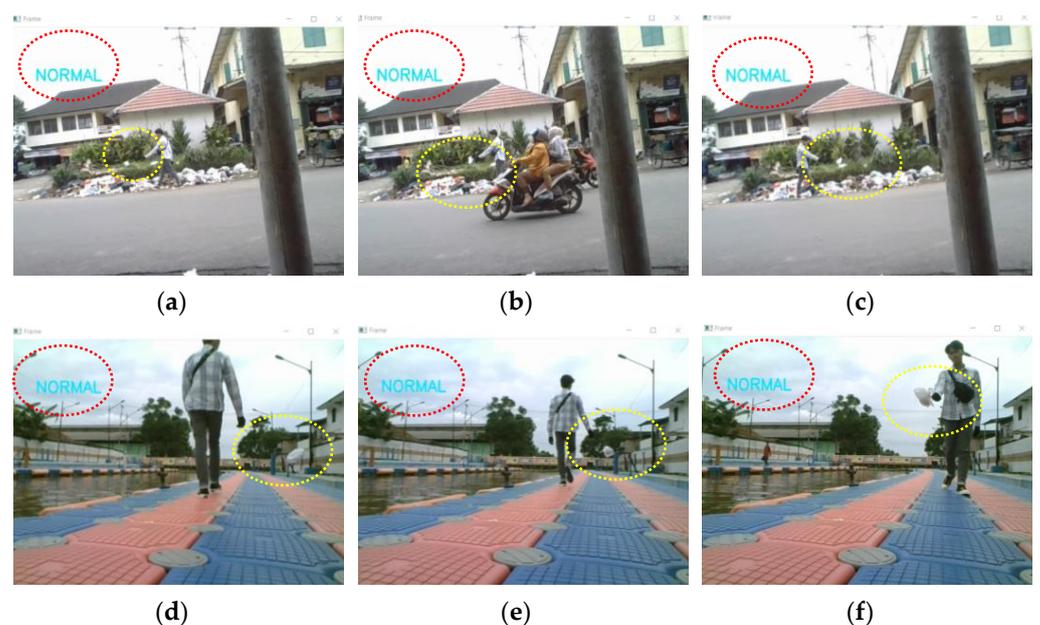


Figure 9. Experiments in the real environment using model 7: (a–c) mini garden, (d–f) Sekanak River. (a) Littering activity, (b) littering activity, (c) littering activity, (d) littering activity, (e) littering activity, (f) littering activity.

The red circles in each group of pictures in Figures 9–11 and Figures 13 and 14 indicate the decision result that is the output of the monitoring device of this research, while the yellow circles show the garbage that was thrown away by the human.

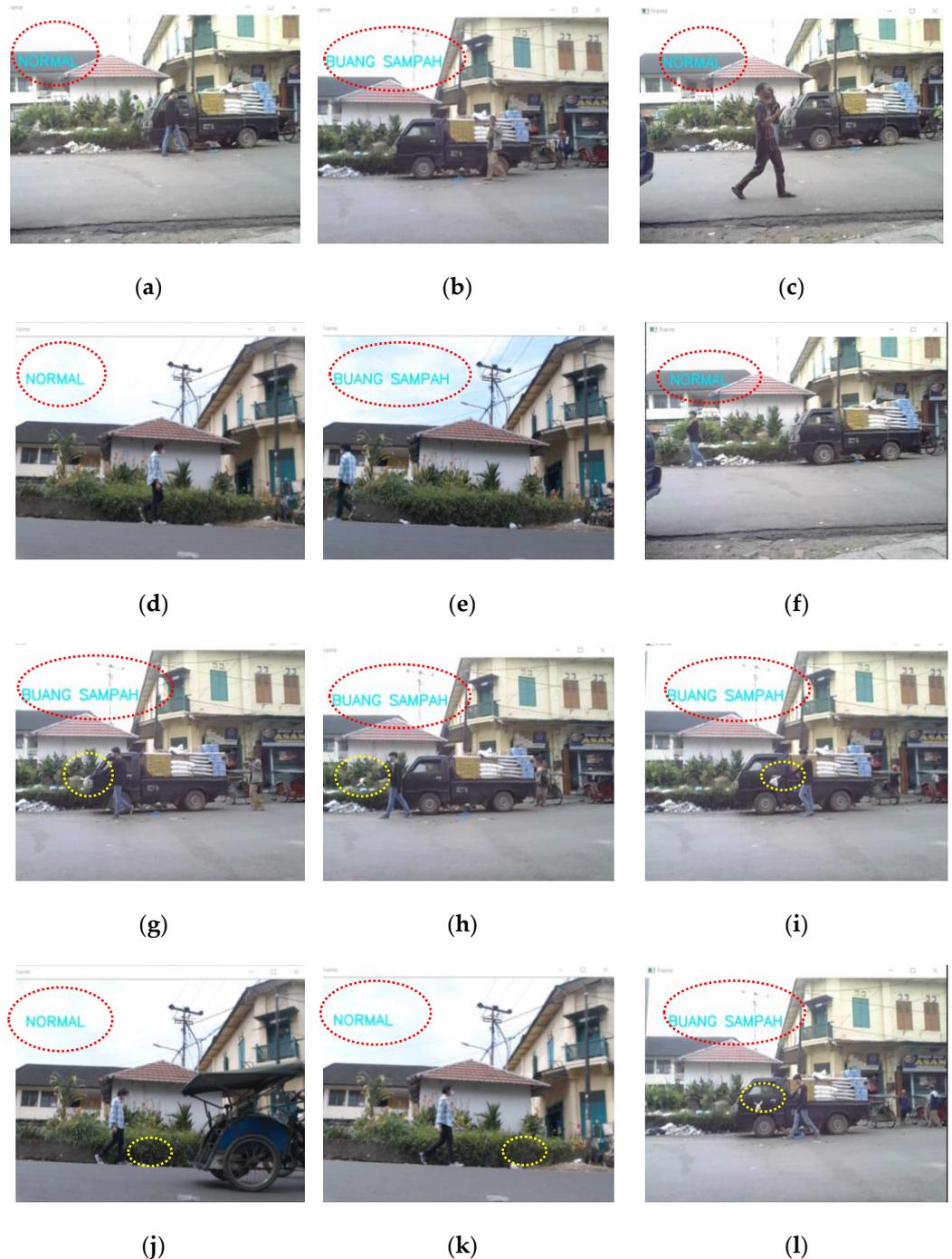


Figure 10. The implementation of the system in the mini garden. (a) Normal activity, (b) normal activity, (c) normal activity, (d) normal activity, (e) normal activity, (f) normal activity, (g) littering activity, (h) littering activity, (i) littering activity, (j) littering activity, (k) littering activity, (l) littering activity. Note: “buang sampah” in Bahasa Indonesia, means littering in English.

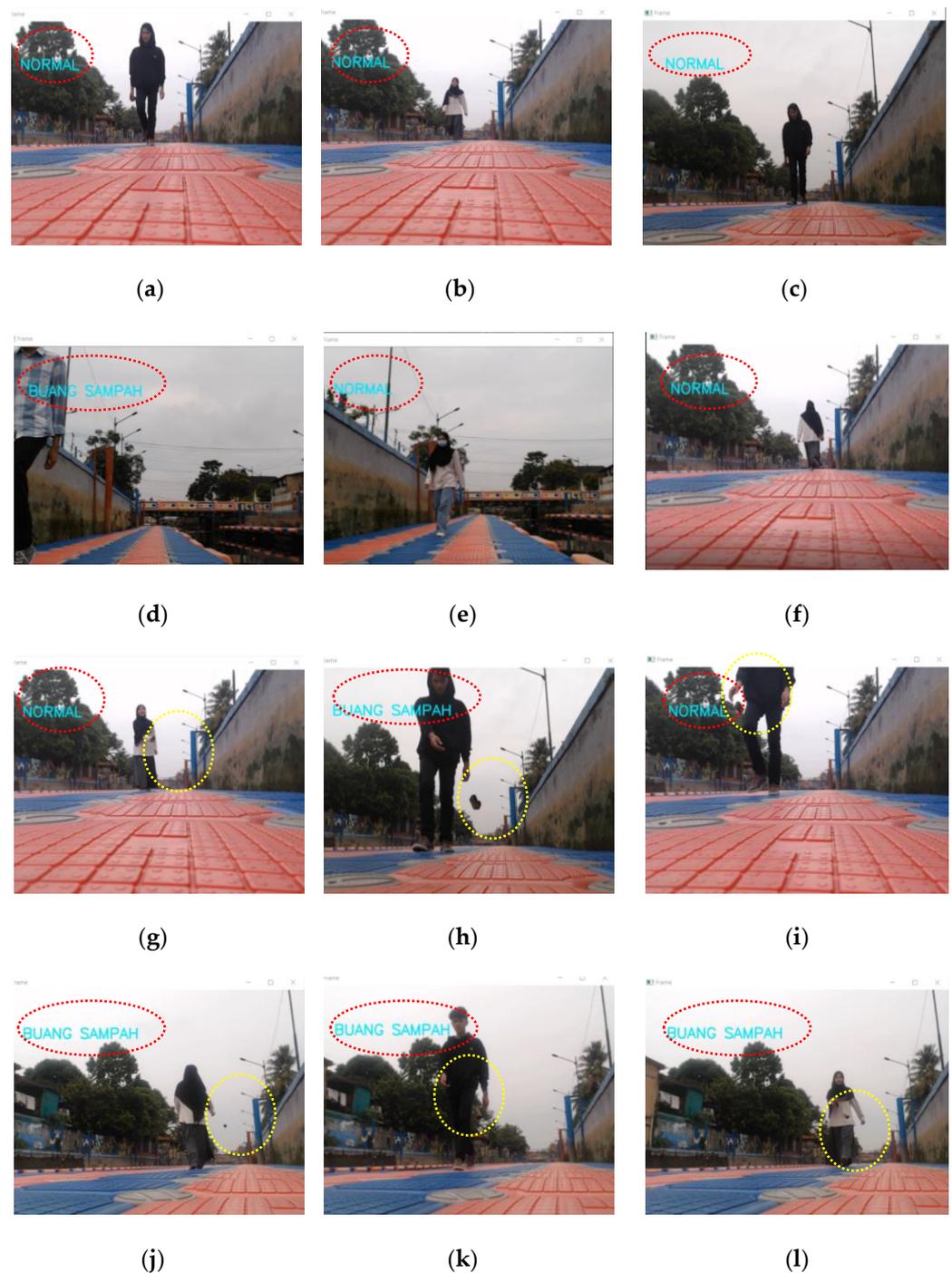


Figure 11. The implementation of the system in the Sekanak River. (a) Normal activity, (b) normal activity, (c) normal activity, (d) normal activity, (e) normal activity, (f) normal activity, (g) littering activity, (h) littering activity, (i) littering activity, (j) littering activity, (k) littering activity, (l) littering activity. Note: “buang sampah” in Bahasa Indonesia, means littering in English.

Due to the system being unable to differentiate between the normal and littering activities in the real environment, the system software was updated using the ReLU activation. The properties of training for model 7 can be seen in Table 4. In this model, the accuracy obtained was not good. As shown in Table 4, the accuracy was only 56% and the loss was very high, 77.1%. Therefore, the activation was then changed to Sigmoid activation and there was an addition of the layers in model 8 (please see Table 4 for more

details of the properties of the training). The training data output in model 8 was then implemented in the real experiments. The result can be seen in Figures 10 and 11.

Figure 10a–l shows the real experiment in the mini garden. In Figure 10a,c,d,f, the system could recognize the action well. The normal activity that was being carried out by the human was detected as normal activity by the system. However, for Figure 10b,e, the actions were not detected correctly. Normal actions in those images were interpreted as littering actions. In Figure 10g–l, the systems were tested to recognize the action of littering.

The system could detect the actions of littering in well, as shown in Figure 10g–i,l. However, the littering activity in Figure 10j,k could not be detected well by the systems; it detected the actions as normal activity. From this experimental result, it can be concluded that the system did not work well.

The experiment was then continued in the Sekanak River, as shown in Figure 11. The system was tested as to whether it could recognize normal activity well, as shown in Figure 11a–f and whether it could recognize littering activity, as presented in Figure 11g–l. From these experiments, the system worked well in detecting five actions of normal activity and four actions of littering activity in Figure 11. In Figure 11d, the system still detected the action as littering (“buang sampah” in Bahasa Indonesia although the person had passed far away from the system, while, in Figure 11g,i, the system still could not detect the actions of littering, although the garbage had touched the ground. From these experiments, it could be concluded that the system still gave the wrong interpretation for about 33.3% of actions in the mini garden and 25% in the Sekanak River. Thus, it still has high error.

To obtain better machine learning results in this research, it needs a large amount of training and testing data. Based on the method used, there were several different layers applied, namely the convolution layer, pooling layer, dropout layer, flatten layer, and dense layer. In addition to those layers, there were also ReLU and Sigmoid activations. In this research, the data process obtained good results when using Sigmoid activation, as was obtained using model 8.

To analyze the experimental data obtained, the confusion matrix of the training data of model 8 presented in Table 5 should be noticed. From this matrix, can be obtained the performance of the proposed model can be obtained. Model 8 shows accuracy of 56.7%, precision of 100%, and recall of 56.7%. The loss of this model was also still high, i.e., 0.70 or 70%. This is because the process of collecting video datasets was still not optimal. Therefore, more video datasets are needed for training and testing. With this adjustment, the machine can distinguish between littering and normal activities. Apart from this, the model selection process in the training and testing data also affected the accuracy value that was obtained. During the detection process, the systems failed to recognize littering activities and normal activities at two locations, as shown in Tables 6 and 7. This is logical due to the accuracy of the system being only 56%; therefore, the system only had around 67.7–75% success in recognizing the normal and littering activity.

Table 5. CNN confusion matrix for model 8.

Predicted Values	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	TP = 1813	FP = 0
Predictive Negative (0)	FN = 1384	TN = 0

Table 6. CNN experimental data in the mini garden.

No.	Reference	Activity	System Detection	Notification	Note
1.	Figure 10a	Normal	Normal	<i>Silent</i>	<i>Success</i>
2.	Figure 10b	Normal	Littering	<i>Sound</i>	<i>Failure</i>
3.	Figure 10c	Normal	Normal	<i>Silent</i>	<i>Success</i>
4.	Figure 10d	Normal	Normal	<i>Silent</i>	<i>Success</i>
5.	Figure 10e	Normal	Littering	<i>Sound</i>	<i>Failure</i>
6.	Figure 10f	Normal	Normal	<i>Silent</i>	<i>Success</i>
7.	Figure 10g	Littering	Littering	<i>Sound</i>	<i>Success</i>
8.	Figure 10h	Littering	Littering	<i>Sound</i>	<i>Success</i>
9.	Figure 10i	Littering	Littering	<i>Sound</i>	<i>Success</i>
10.	Figure 10j	Littering	Normal	<i>Silent</i>	<i>Failure</i>
11.	Figure 10k	Littering	Normal	<i>Silent</i>	<i>Failure</i>
12.	Figure 10l	Littering	Littering	<i>Sound</i>	<i>Success</i>

Table 7. CNN experimental data in Sekanak River.

No.	Reference	Activity	System Detection	Notification	Note
1.	Figure 11a	Normal	Normal	<i>Silent</i>	<i>Success</i>
2.	Figure 11b	Normal	Normal	<i>Silent</i>	<i>Success</i>
3.	Figure 11c	Normal	Normal	<i>Silent</i>	<i>Success</i>
4.	Figure 11d	Normal	Littering	<i>Sound</i>	<i>Failure</i>
5.	Figure 11e	Normal	Normal	<i>Silent</i>	<i>Success</i>
6.	Figure 11f	Normal	Normal	<i>Silent</i>	<i>Success</i>
7.	Figure 11g	Littering	Normal	<i>Silent</i>	<i>Failure</i>
8.	Figure 11h	Littering	Normal	<i>Silent</i>	<i>Failure</i>
9.	Figure 11i	Littering	Littering	<i>Sound</i>	<i>Success</i>
10.	Figure 11j	Littering	Littering	<i>Sound</i>	<i>Success</i>
11.	Figure 11k	Littering	Littering	<i>Sound</i>	<i>Success</i>
12.	Figure 11l	Littering	Littering	<i>Sound</i>	<i>Success</i>

4.2. CNN-LSTM Experiment

For the CNN-LSTM experiment, the simulation was tested using two models, i.e., model 9 and model 10. The properties of the models used are presented in Table 8 and the result of the training is shown in Table 9. The result of model 10 was good, and the accuracy of the system was 97% (see Figure 12).

Table 8. Properties of training in the second CNN-LSTM experiment.

Parameters	Model 9	Model 10
Name:	CNN_LSTM 9	CNN_LSTM 10
Epoch:	1	100
Activation function:	ReLU	ReLU
Layer:	4	4
Input Size:	300	300
Hidden Size:	256	256
Stride:	1 (2 × 3)	1 (2 × 3)
Pooling Layer (Average Pooling and Max Pooling):	2 × 3	2 × 3

Table 9. CNN-LSTM Experimental data.

Model	Activation	Epoch	Accuracy (%)	Loss (%)	Duration	Note
9	ReLU	1	48.3	69.48	30 min	Success
10	ReLU	100	97	10.61	24 h	Success

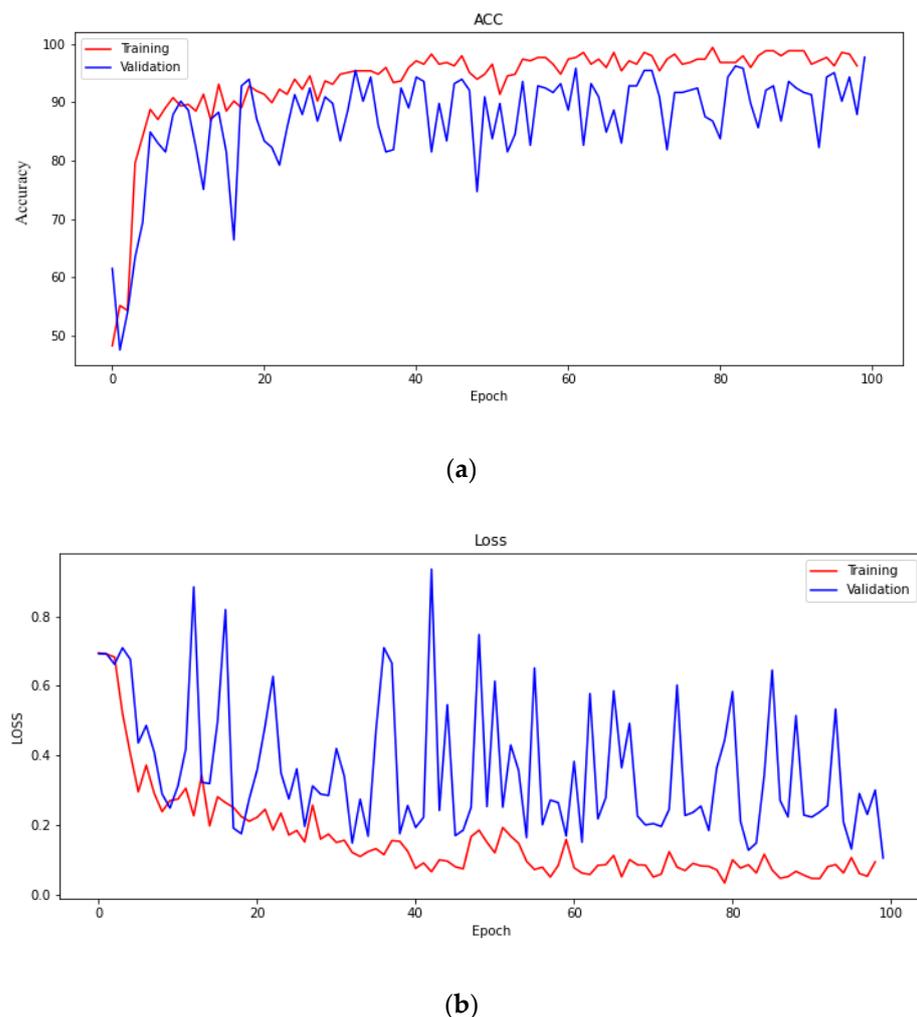


Figure 12. The accuracy and loss of the CNN-LSTM. **(a)** Accuracy, **(b)** Loss.

The implementation of the model to the real experiments can be seen in Figure 13 for the mini garden and Figure 14 for the Sekanak River. Figure 13a–r presents the systems which could differentiate between normal activities and littering activities in the mini garden. This experiment was conducted in multiple scenarios, i.e., the human just kept standing in the mini garden and suddenly threw the garbage, the human walked from the right side to the left side and vice versa, and the human walked across the street. The system could recognize the activity and could differentiate between them well. The normal activity in Figures 12a,b and 13a–i could be classified well. Only one of the activities in Figure 13c could not be predicted well by the system. The normal activity was interpreted as littering activity by the system. However, for the littering activity in the rest of the figures, i.e., Figure 13j–r could be interpreted well by the system.

In Figure 14, the system was implemented to detect normal and littering activity in Sekanak River. The system could classify all activities well. The normal activity in Figure 14a–i could be recognized by the system as normal, and the littering activity, as shown in Figure 14j–r, could be identified as littering. When the system detected the littering activity, the system sent the data to the mini-PC and they were then passed on to the speaker which gave a warning not to litter in that area. However, when the human had just passed by the location, the system did not give a warning through the speaker. Thus, the system showed great performance when using model 10 in this research. The data from the warning system for these two experiments are shown in Tables 10 and 11.

Table 10. CNN-LSTM experimental data in the mini garden.

No.	Reference	Activity	System Detection	Notification	Note
1.	Figure 13a	Normal	Normal	<i>Silent</i>	<i>Success</i>
2.	Figure 13b	Normal	Normal	<i>Silent</i>	<i>Success</i>
3.	Figure 13c	Normal	Littering	<i>Sound</i>	<i>Failure</i>
4.	Figure 13d	Normal	Normal	<i>Silent</i>	<i>Success</i>
5.	Figure 13e	Normal	Normal	<i>Silent</i>	<i>Success</i>
6.	Figure 13f	Normal	Normal	<i>Silent</i>	<i>Success</i>
7.	Figure 13g	Normal	Normal	<i>Silent</i>	<i>Success</i>
8.	Figure 13h	Normal	Normal	<i>Silent</i>	<i>Success</i>
9.	Figure 13i	Normal	Normal	<i>Silent</i>	<i>Success</i>
10.	Figure 13j	Littering	Littering	<i>Sound</i>	<i>Success</i>
11.	Figure 13k	Littering	Littering	<i>Sound</i>	<i>Success</i>
12.	Figure 13l	Littering	Littering	<i>Sound</i>	<i>Success</i>
13.	Figure 13m	Littering	Littering	<i>Sound</i>	<i>Success</i>
14.	Figure 13n	Littering	Littering	<i>Sound</i>	<i>Success</i>
15.	Figure 13o	Littering	Littering	<i>Sound</i>	<i>Success</i>
16.	Figure 13p	Littering	Littering	<i>Sound</i>	<i>Success</i>
17.	Figure 13q	Littering	Littering	<i>Sound</i>	<i>Success</i>
18.	Figure 13r	Littering	Littering	<i>Sound</i>	<i>Success</i>

Table 11. CNN-LSTM experimental data in Sekanak River.

No.	Reference	Activity	System Detection	Notification	Note
1.	Figure 14a	Normal	Normal	<i>Silent</i>	<i>Success</i>
2.	Figure 14b	Normal	Normal	<i>Silent</i>	<i>Success</i>
3.	Figure 14c	Normal	Normal	<i>Silent</i>	<i>Success</i>
4.	Figure 14d	Normal	Normal	<i>Silent</i>	<i>Success</i>
5.	Figure 14e	Normal	Normal	<i>Silent</i>	<i>Success</i>
6.	Figure 14f	Normal	Normal	<i>Silent</i>	<i>Success</i>
7.	Figure 14g	Normal	Normal	<i>Silent</i>	<i>Success</i>
8.	Figure 14h	Normal	Normal	<i>Silent</i>	<i>Success</i>
9.	Figure 14i	Normal	Normal	<i>Silent</i>	<i>Success</i>
10.	Figure 14j	Littering	Littering	<i>Sound</i>	<i>Success</i>
11.	Figure 14k	Littering	Littering	<i>Sound</i>	<i>Success</i>
12.	Figure 14l	Littering	Littering	<i>Sound</i>	<i>Success</i>
13.	Figure 14m	Littering	Littering	<i>Sound</i>	<i>Success</i>
14.	Figure 14n	Littering	Littering	<i>Sound</i>	<i>Success</i>
15.	Figure 14o	Littering	Littering	<i>Sound</i>	<i>Success</i>
16.	Figure 14p	Littering	Littering	<i>Sound</i>	<i>Success</i>
17.	Figure 14q	Littering	Littering	<i>Sound</i>	<i>Success</i>
18.	Figure 14r	Littering	Littering	<i>Sound</i>	<i>Success</i>

Figure 12 shows the accuracy and loss for the CNN-LSTM. The number of epochs was increased from 1 to 100 and shows that the accuracy obtained a good result of 97.7% and the average loss of 0.1. The confusion matrix of model 10 can be seen in Table 12. The calculation of the accuracy, precision, and recall of model 10 obtained results of 97.3%, 96%, and 97.4%, respectively.

Table 12. The CNN-LSTM confusion matrix for Model 10.

Predicted Values	Actually Positive (1)	Actually Negative (0)
Predicted positive (1)	TP = 1338	FP = 35
Predictive negative (0)	FN = 45	TN = 1595

In this research, the sensors were also tested in the real environment. The data are presented in Table 13. All sensors worked well and presented valid data. For the sixth to the tenth experiments, there were no data for the water level. It was due to the water level sensor only being integrated into the systems of river monitoring.

Table 13. Sensors' experimental data.

No.	Temperature (° C)	Humidity (%)	Water Level (cm)	Air Quality (ADC)	Location
1.	32.00	69.00	19	998	River
2.	35.00	67.00	20	1002	River
3.	33.20	67.00	19	1003	River
4.	33.00	66.00	19	999	River
5.	33.00	67.00	19	998	River
6.	36.08	67.40	-	789	Garden
7.	37.08	66.50	-	1003	Garden
8.	32.30	66.00	-	1002	Garden
9.	33.06	67.00	-	1002	Garden
10.	34.08	67.00	-	998	Garden

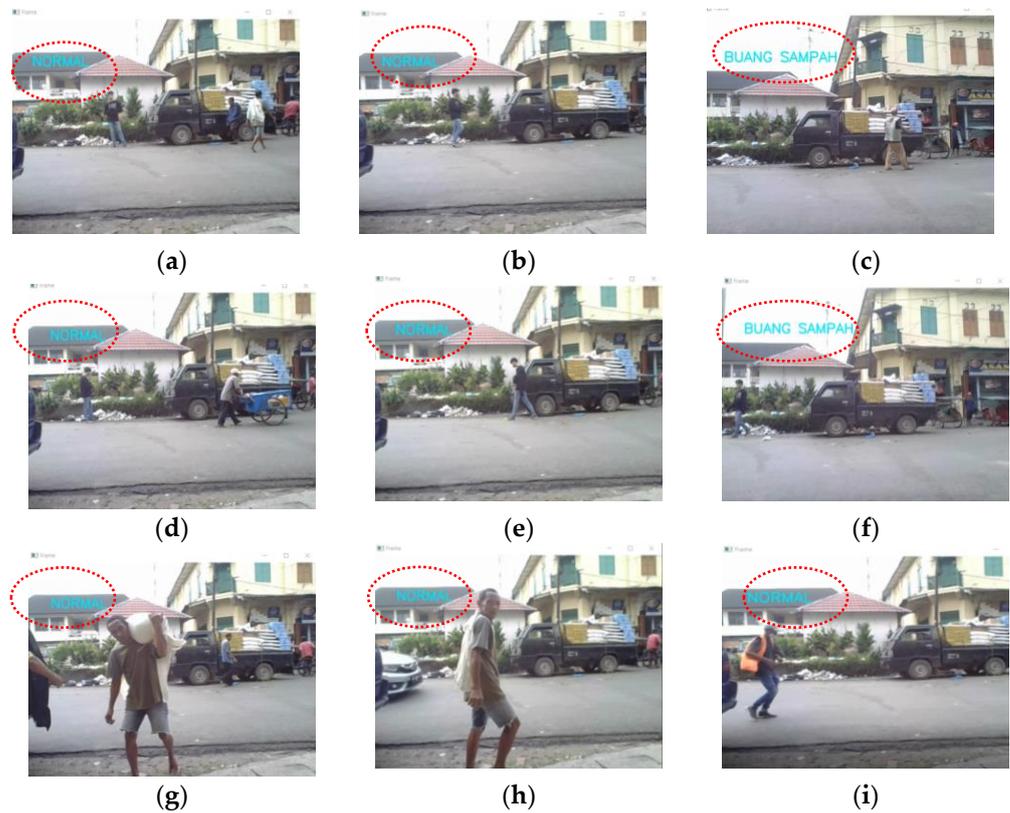


Figure 13. Cont.

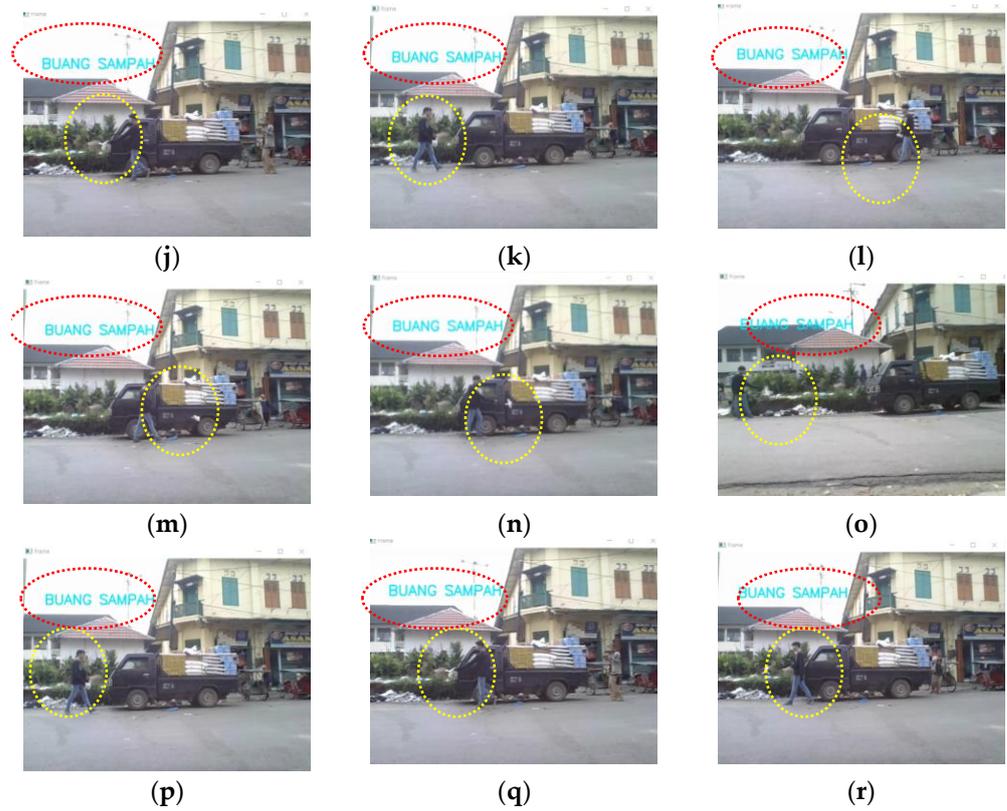


Figure 13. The implementation of the system in the mini garden. (a) Normal activity, (b) normal activity, (c) normal activity, (d) normal activity, (e) normal activity, (f) normal activity, (g) normal activity, (h) normal activity, (i) normal activity, (j) littering activity, (k) littering activity, (l) littering activity, (m) littering activity, (n) littering activity, (o) littering activity, (p) littering activity, (q) littering activity, (r) littering activity. Note: “buang sampah” in Bahasa Indonesia, means littering in English.

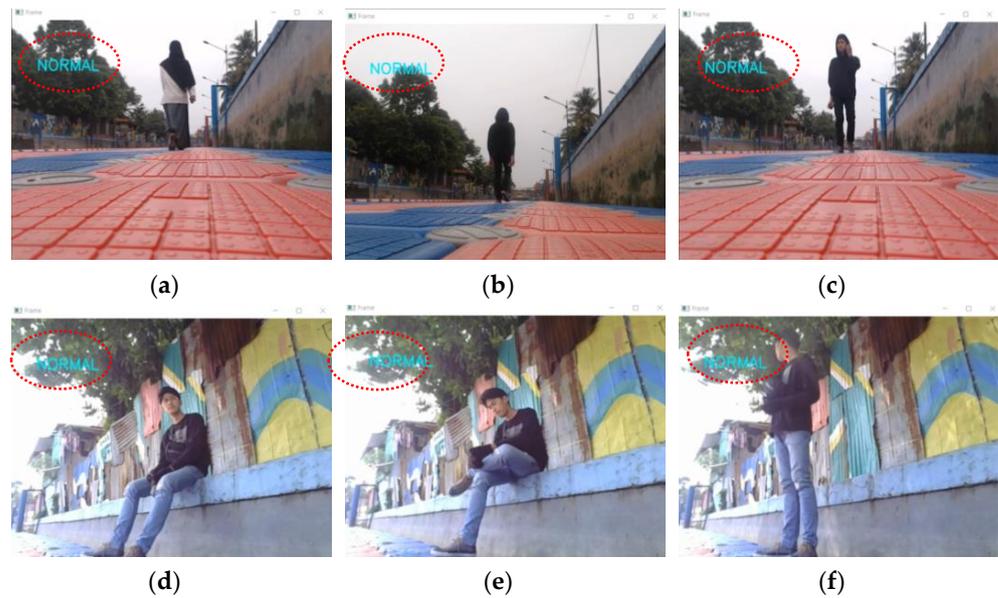


Figure 14. Cont.

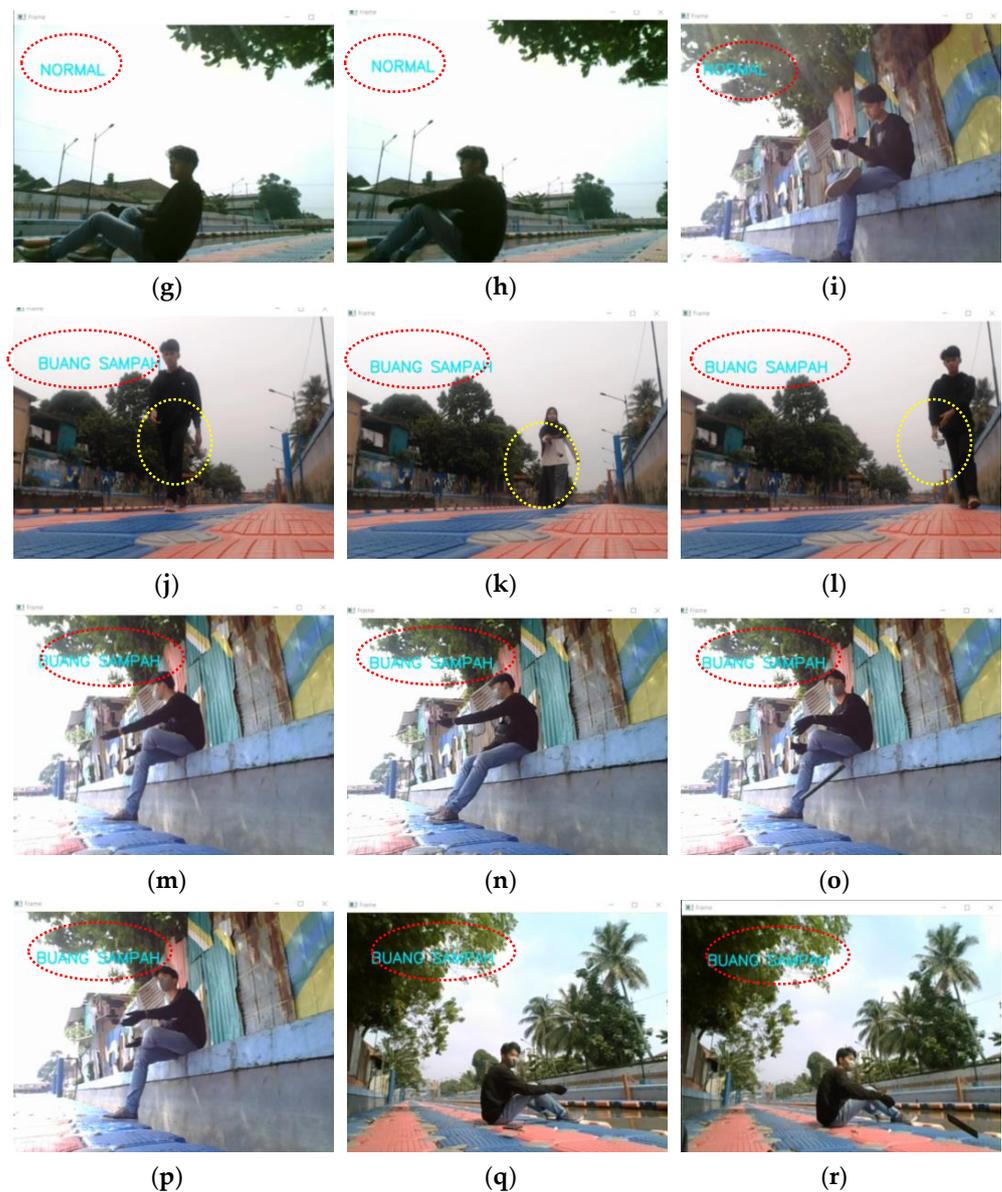


Figure 14. The implementation of the system in the Sekanak River. (a) Normal activity, (b) normal activity, (c) normal activity, (d) normal activity, (e) normal activity, (f) normal activity, (g) normal activity, (h) normal activity, (i) normal activity, (j) littering activity, (k) littering activity, (l) littering activity, (m) littering activity, (n) littering activity, (o) littering activity, (p) littering activity, (q) littering activity, (r) littering activity. Note: “buang sampah” in Bahasa Indonesia, means littering in English.

5. Conclusions

The CNN and the CNN-LSTM that were applied in the system could work well with success rates of around 50% and above. Using the CNN, the system could only recognize the activity in the mini garden and the Sekanak River about 67.7–75% of the time. This was due to the training process of this CNN only being able to achieve 56% accuracy and having high loss value, i.e., 70%. However, by using the CNN-LSTM, the system could perform better. It showed 97.7% accuracy and 10% loss. This method also produced a good result when it was applied in the real experiments, with a percentage of correct classification of around 97.2%, whereas, from 36 experiments in the mini garden and the river, the system only made 1 mistake. It could differentiate between littering and normal activities when it was applied in Sekanak River and the mini garden.

6. Patents

This research project was granted by Ministry of Law and Human Rights of the Republic of Indonesia as Surat Pencatatan Hak Cipta EC00202145092, on 7 September 2021. This project is also on its way to be registered as a patent.

Author Contributions: Conceptualization, N.L.H. and A.S.H.; methodology, N.L.H., A.S.H. and T.D.; software, P.A.R.S. and N.L.H.; validation, W.C. and S.A.H.S.; formal analysis, T.D.; investigation, N.L.H. and P.A.R.S.; resources, A.S.H.; data curation, N.L.H., A.S.H. and P.A.R.S.; writing—original draft preparation, N.L.H. and P.A.R.S.; writing—review and editing, W.C., S.A.H.S., A.G., K.O. and M.S.; visualization, N.L.H. and P.A.R.S.; supervision, W.C.; project administration, P.A.R.S.; funding acquisition, N.L.H., A.G., K.O. and M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by POLITEKNIK NEGERI SRIWIJAYA, grant number 3628/PL6.2.1/LT/2021 and 5831/PL6.2.1/LT/2021.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Acknowledgments: The authors would like to thank the Politeknik Negeri Sriwijaya for its funding and support. The author would like to thank their colleagues in the Artificial Intelligence Laboratory of Electrical Engineering in Politeknik Negeri Sriwijaya. Finally, the authors thank the Intelligence Laboratory of Sriwijaya University and the Cyborg IT Center.

Conflicts of Interest: The authors declare that this research has no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Putri, V.O.; Pratiwi, W.D. Heritage Tourism Development Strategy in Sekanak Market Area of Palembang City. *ASEAN J. Hosp. Tour.* **2021**, *19*, 30–43. [[CrossRef](#)]
- Tripathi, R.K.; Jalal, A.S.; Agrawal, S.C. Suspicious human activity recognition: A review. *Artif. Intell. Rev.* **2017**, *50*, 283–339. [[CrossRef](#)]
- Adesipo, A.; Fadeyi, O.; Kuca, K.; Krejcar, O.; Maresova, P.; Selamat, A.; Adenola, M. Smart and Climate-Smart Agricultural Trends as Core Aspects of Smart Village Functions. *Sensors* **2020**, *20*, 5977. [[CrossRef](#)]
- Cvar, N.; Trilar, J.; Kos, A.; Volk, M.; Stojmenova Duh, E. The Use of IoT Technology in Smart Cities and Smart Villages: Similarities, Differences, and Future Prospects. *Sensors* **2020**, *20*, 3897. [[CrossRef](#)] [[PubMed](#)]
- Goenka, S.; Mangrulkar, R.S. Robust Waste Collection: Exploiting IOT Potentiality in Smart Cities. *i-Manager's J. Softw. Eng.* **2017**, *11*, 10–18.
- Medvedev, A.; Fedchenkov, P.; Zaslavsky, A. Waste management as an IoT enabled service in Smart Cities. In *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*; Springer: Cham, Switzerland, 2015; Volume 9247, pp. 104–105.
- Mohanty, P.S.S.; Mohanta, B.; Nanda, P.; Sen, S. Smart Village Initiatives: An Overview. *Smart Village Technol.* **2020**, *17*, 3–24.
- Adamowicz, M.; Zwolińska-Ligaj, M. The Smart Village as a Way to Achieve Sustainable Development in Rural Areas of Poland. *Sustainability* **2020**, *12*, 6503. [[CrossRef](#)]
- Vaishar, A.; Št'astná, M. Smart Village and Sustainability. Southern Moravia Case Study. *Eur. Countrys.* **2019**, *11*, 651–660. [[CrossRef](#)]
- Aziiza, A.; Susanto, T.D. The Smart Village Model for Rural Area (Case Study: Banyuwangi Regency). *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *722*, 012011. [[CrossRef](#)]
- Zhang, X.; Zhang, Z. How Do Smart Villages Become a Way to Achieve Sustainable Development in Rural Areas? Smart Village Planning and Practices in China. *Sustainability* **2020**, *12*, 10510. [[CrossRef](#)]
- Despotovic, A.; Joksimovic, M.; Jovanovic, M. Demographic revitalization of montenegrin rural areas through the smart village concept. *J. Agric. For.* **2020**, *66*, 125–138. [[CrossRef](#)]
- Kim, H.; Kim, J.; Kim, Y.-S.; Kim, M.; Lee, Y. Energy-Efficient Wearable EPTS Device Using On-Device DCNN Processing for Football Activity Classification. *Sensors* **2020**, *20*, 6004. [[CrossRef](#)] [[PubMed](#)]
- Sharma, L.; Chao, C.-H.; Wu, S.-L.; Li, M.-C. High Accuracy WiFi-Based Human Activity Classification System with Time-Frequency Diagram CNN Method for Different Places. *Sensors* **2021**, *21*, 3797. [[CrossRef](#)] [[PubMed](#)]
- Kerdjidi, O.; Ramzan, N.; Ghanem, K.; Amira, A.; Chouireb, F. Fall detection and human activity classification using wearable sensors and compressed sensing. *J. Ambient. Intell. Humaniz. Comput.* **2019**, *11*, 349–361. [[CrossRef](#)]
- Li, X.; Zhang, Y.; Zhang, J. Concurrent Activity Recognition with Multimodal CNN-LSTM Structure. *arXiv* **2017**, arXiv:1702.01638.
- Stadelmayer, T.; Santra, A.; Weigel, R.; Lurz, F. Data-Driven Radar Processing Using a Parametric Convolutional Neural Network for Human Activity Classification. *IEEE Sens. J.* **2021**, *21*, 19529–19540. [[CrossRef](#)]

18. Yang, L.; Yang, T.-Y.; Liu, H.; Shan, X.; Brighton, J.; Skrypchuk, L.; Mouzakitis, A.; Zhao, Y. A Refined Non-Driving Activity Classification Using a Two-Stream Convolutional Neural Network. *IEEE Sens. J.* **2020**, *21*, 15574–15583. [[CrossRef](#)]
19. Chen, K.; Zhang, D.; Yao, L.; Guo, B.; Yu, Z.; Liu, Y. Deep Learning for Sensor-based Human Activity Recognition. *ACM Comput. Surv.* **2021**, *54*, 1–40. [[CrossRef](#)]
20. Beddiar, D.R.; Nini, B.; Sabokrou, M.; Hadid, A. Vision-based human activity recognition: A survey. *Multimed. Tools Appl.* **2020**, *79*, 30509–30555. [[CrossRef](#)]
21. Sarma, M.; Deb, K.; Dhar, P.; Koshiha, T. Traditional Bangladeshi Sports Video Classification Using Deep Learning Method. *Appl. Sci.* **2021**, *11*, 2149. [[CrossRef](#)]
22. Wang, M.; Oczak, M.; Larsen, M.; Bayer, F.; Maschat, K.; Baumgartner, J.; Rault, J.-L.; Norton, T. A PCA-based frame selection method for applying CNN and LSTM to classify postural behaviour in sows. *Comput. Electron. Agric.* **2021**, *189*, 106351. [[CrossRef](#)]
23. Ullah, W.; Ullah, A.; Hussain, T.; Khan, Z.; Baik, S. An Efficient Anomaly Recognition Framework Using an Attention Residual LSTM in Surveillance Videos. *Sensors* **2021**, *21*, 2811. [[CrossRef](#)]
24. Patel, M.B. Real-Time Violence Detection Using CNN-LSTM. *arXiv* **2021**, arXiv:2107.07578.
25. Tasdelen, A.; Sen, B. A hybrid CNN-LSTM model for pre-miRNA classification. *Sci. Rep.* **2021**, *11*, 1–9. [[CrossRef](#)]
26. Arif, S.; Wang, J.; Siddiqui, A.A.; Hussain, R.; Hussain, F. Bidirectional LSTM with saliency-aware 3D-CNN features for human action recognition. *J. Eng. Res.* **2021**, *9*, 115–133. [[CrossRef](#)]
27. Shiranthika, C.; Premakumara, N.; Chiu, H.-L.; Samani, H.; Shyalika, C.; Yang, C.-Y. Human Activity Recognition Using CNN & LSTM. In Proceedings of the 2020 5th International Conference on Information Technology Research (ICITR), Moratuwa, Sri Lanka, 2–4 Desemeber 2020; pp. 1630–1634.
28. Sarnaik, N.N.J. Human Activity Recognition using CNN. *Int. J. Sci. Res. Publ.* **2020**, *10*, 9804. [[CrossRef](#)]
29. Caetano, P.; Mazzoni, A.; Ranieri, V.; Scott, R.; MacLeod, A.; Mauro, F.; Dragone, R. Activity Recognition for Ambient Assisted Living with Videos, Inertial Units and Ambient Sensors. *Sensors* **2021**, *21*, 768.
30. Senyurek, V.Y.; Imtiaz, M.H.; Belsare, P.; Tiffany, S.; Sazonov, E. A CNN-LSTM neural network for recognition of puffing in smoking episodes using wearable sensors. *Biomed. Eng. Lett.* **2020**, *10*, 195–203. [[CrossRef](#)] [[PubMed](#)]
31. Noh, S.-H. Performance Comparison of CNN Models Using Gradient Flow Analysis. *Informatics* **2021**, *8*, 53. [[CrossRef](#)]
32. Rani, S.; Babbar, H.; Coleman, S.; Singh, A.; Aljhdali, H.M. An Efficient and Lightweight Deep Learning Model for Human Activity Recognition Using Smartphones. *Sensors* **2021**, *21*, 3845. [[CrossRef](#)] [[PubMed](#)]
33. He, K. Deep residual learning for image recognition. In Proceedings of the IEEE Transactions on Circuits and Systems for Video Technology, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.