Scientific Journal of Informatics
Vol. 11, No. 3, Aug 2024

p-ISSN 2407-7658 https://journal.unnes.ac.id/journals/sji/index e-ISSN 2460-0040

Recognition of Organic Waste Objects Based on Vision Systems Using Attention Convolutional Neural Networks Models

Aradea[1*], Rianto[2], Husni Mubarok[3]
[1,2,3]Department of Informatics, Faculty of Engineering, Universitas Siliwangi, Indonesia

Abstract.
Purpose: High population growth and increasing consumption patterns have resulted in significant organic waste production. The public often does not understand the correct way to deal with the problem of organic waste, including public awareness regarding the need for its management. Therefore, a system is needed to recognize waste objects based on various types. Currently, much research in this field has been studying object recognition, for example, the implementation of the Convolutional Neural Networks (CNN) model. However, there are still various challenges that must be addressed, including objects with diverse visual characteristics such as form, size, color, and physical condition. This research focuses on developing a system that enhances object recognition of waste, specifically organic waste, using an Attention Convolutional Neural Network (ACNN). By integrating attention mechanisms into the CNN model, this study addresses the challenges of recognizing waste objects with diverse visual characteristics. The proposed system seeks to improve the accuracy and efficiency of organic waste identification, which is crucial for advancing waste management practices and reducing environmental impact.
Methods: This research combines a CNN architecture with an attention mechanism to create a better object detection environment called Attention-CNN (ACNN). The ACNN architecture employed consists of one layer input, three convoluted layers, three max-pooling layers, one attention layer, one flattened layer, four dropout layers, and two dense layers arranged in a certain way.
Result: The research result shows that the model CNN with attention mechanism (ACNN) was slightly better at 86.93% than the standard model of CNN, which accounted for 86.70% in accuracy.
Novelty: In general, the current use of CNN architecture to address waste object recognition problems typically employs standard architectures, resulting in lower accuracy for complex waste objects. In contrast, our research integrates attention mechanisms into the CNN architecture (ACNN), enhancing the model's ability to focus on relevant features of waste objects. This leads to improved recognition accuracy and robustness against visual variability. This distinction is important as it overcomes the limitations of standard CNN models in handling visually diverse and complex waste objects, thereby highlighting the novelty and contribution of our research.

Keywords: ACNN, Attention mechanism, CNN, Waste object recognition
Received June 2024 / **Revised** July 2024 / **Accepted** July 2024

INTRODUCTION

Waste has become a never-ending problem in Indonesia. The high population causes high consumption and produces a lot of waste both household organic and non-organic waste. According to Law 18 of 2008, waste is the remains of all forms of human activity in solid form. Based on data from the Tasikmalaya City Environmental Service, around 192.91 tonnes of waste entered the Ciangir landfill in 2021 [1]. This occurred because the majority of people in Indonesia do not understand how to properly handle waste [2]. Facilities and infrastructure, educational levels, and inadequate support from the local government regarding waste processing have become some factors generating public unawareness of managing waste properly[2]. In particular, organic waste is the most widely distributed waste among other existing waste in the environment. This is due to the rate of population growth and community consumption patterns comparable to the rate of waste production, (e.g. organic waste) [3]. Currently, researchers have made various efforts as an alternative to processing organic waste. It aims at reducing the volume of waste scattered in the environment. One way is to convert it into organic fertilizer applied to plants.

*Corresponding author.
Email addresses: aradea@unsil.ac.id (Aradea)
DOI: 10.15294/sji.v11i3.6494

Organic fertilizer is derived from organic waste decomposed to be smaller and more inodorous [4]. The decomposing process is undertaken through both aerobic and anaerobic processes. Compared to chemical fertilizers, organic fertilizers have stable properties in a planting medium allowing beneficial elements to settle and fertilize plants [4]. However, in reality, more chemical fertilizers are employed to decompose the soil structure, loss of organic elements, and environmental pollution periodically [5]. This takes place since chemical fertilizers are easier to obtain and have an instant effect on plants. This research is intended to address the issue of waste, particularly organic waste by developing a model to recognize organic waste objects. The data from this introduction can be adopted as a basis for creating organic fertilizer through computer vision technology.

Artificial intelligence approaches for image processing and classification are rapidly evolving with Convolutional Neural Network (CNN) standing out as a significant milestone in this field. Inspired by the abilities of human brains to differentiate and classify images, CNN is a deep learning algorithm adapted to train machines to recognize various visual objects from input data. The use of larger datasets for machine training leads to higher accuracy in the machine learning model [6]–[8]. CNN has also demonstrated effectiveness in addressing challenges in object recognition, such as distinguishing between organic and non-organic waste [9]. As a result, CNN has become a valuable tool in image classification across various domains as empirically proven by prior studies.

Umam et al. [10] cultivated a CNN model combined with the *Adam optimizer* to specifically identify Hiragana characters. Their research attained a recognition accuracy of 95% for Hiragana characters. Similarly, Nugroho et al. [11] integrated CNN and SVM methods to expand object recognition to identify both Hiragana and Katakana characters. The findings reported that the accuracy reached 88.21%. The advancement of deep learning models with CNNs (e.g. complex modeling, transfer learning, and visual integration) was proposed by Zhang and Shi [12]. This approach was also applied to change detection demonstrated by Zhao et al. [13]. It integrated three types of visuals for object part recognition and utilized the attention module concept for identifying diseases in tomato leaves. Their CNN model development successfully increased accuracy performance compared to standard CNN architecture achieving 99.24% accuracy. On the other hand, Jha et al. [14] integrated the CNN architecture with two U-nets to enhance model efficiency. In particular, it showcased good segmentation accuracy when tested on several datasets, especially on CVC-ClinicDB, namely 82.21%. Ciancetta et al. [15] evolved a new method called nonintrusive load monitoring (NILM) as the advancement of CNN. The proposed algorithm allowed for simultaneous detection and classification of events without requiring dual processing. Additionally, it reached an outstanding accuracy (98%) on the BLUED dataset. Suresh et al. [16] promoted an optimistic CNN capable of real-time face mask detection providing notifications to users if someone was detected without a mask. This proposed model disclosed a high detection accuracy rate, namely 98%. Sandi et al. [17] directly applied CNN to classify organic and inorganic waste. It attained a recognition accuracy of 62% for organic waste and 96% for inorganic waste. Chęciński and Wawrzyński [18] focused on image compression. It extended block transformations inspired by the Discrete Cosine Transform (DCT) and incorporated convolution, non-linear mapping, linear transformation, quantitation, and inverse operations. Their studies have highlighted the potential of CNNs in optimizing data storage and transmission. Furthermore, Thaha et al. [6] utilized Enhanced Convolutional Neural Networks (ECNN) with binary Algae Algorithm (BAT) optimization to increase accuracy in brain tumor segmentation from MRI images. Research related to improving process optimization based on Gradient Descent (SGD) has been conducted by Peng et al. [19]. They improved process optimization based on Stochastic Gradient Descent (SGD) enhancing the training efficiency of CNN models. Further, Pranav et al. [20] advocated a deep convolutional neural network (DCNN) approach for facial emotion recognition. It illustrated the capability of CNNs in emotion detection and human-computer interaction.

However, like many technological advancements, there are still challenges to overcome. One of the main obstacles is the large visual variations in the shape, size, color, and physical condition of organic waste. The ‘attention’ mechanism enables the model to focus on important features in the image similar to how humans focus on relevant aspects. This allows the model to learn to identify the most meaningful features in the image and increase the accuracy and precision of the classification task. The combination of ‘attention’ mechanisms with CNNs has been a significant breakthrough providing several major benefits in invigorating the model's ability to understand and classify images more accurately [21].

Based on recent literature, there are still gaps in future research, notably related to the results of developing CNN and 'attention' models in recognizing organic waste objects based on vision systems. This paper examined this model in the computer vision pattern recognition process. In particular, it aimed at producing a model enhancing accuracy to recognize organic and non-organic waste objects by adopting the Attention Mechanism to the CNN model. Additionally, it functions to create a more efficient computer vision pattern recognition model by implementing an attention mechanism to reduce the number of parameters. The objective of this paper is to develop a waste object recognition model by expanding CNN with an 'attention' mechanism to strengthen accuracy and precision in distinguishing between organic and non-organic waste objects in a computer vision pattern recognition process.

METHODS

The experiment followed the experimental guide in software engineering by Wohlin et al [22]. Table 1 provides a general description of the research design, including purpose, domain, and focus. Research questions function as indicators for assessing objectives and measuring the elements to be examined. On the other hand, variables refer to the metrics or data related to each question to be answered.

Table 1 presents the research design employed in this study. The primary objective of this research was to develop an object recognition model based on vision systems, specifically tailored for waste management applications. This study aimed to evaluate the overall performance of the system to ensure that it can accurately and efficiently identify various types of waste objects, particularly organic waste. Moreover, this study accentuated the domain of organic waste indicating a significant component of current waste management challenges. Accurate recognition and classification of organic waste were essential for effective waste management and recycling processes.

This study consisted of two main foci, namely the strategy for waste object recognition with the developed model and the performance of the system in accurately classifying waste objects. This includes assessing the ability of the model to handle visually assorted and complex waste. The evaluative questions (EQ) in this study comprise the following two questions:

1) EQ1: How can a vision-based waste object recognition system be effectively developed? This involves the design, implementation, and optimization of the recognition model.
2) EQ2: What is the accuracy measure of the developed vision-based waste object recognition system? This involves evaluating the model's performance using metrics such as accuracy, precision, recall, and loss.

Finally, the study considers two main variables for evaluation:

1) V1: Response refers to the recognition strategy of the system, including how the system processes and identifies waste objects.
2) V2: Confusion Matrix encompassed metrics (e.g. loss, accuracy, precision, and recall) utilized to evaluate the performance of the recognition model quantitatively.

Table 1. The research design

No	Element	Description	Value
1	Goal	Developing a vision system-based object recognition model and evaluating system performance as a whole	It provides a purpose for the research and development activities aiming to create a functional and effective object recognition system.
2	Domain	Organic waste	It specifies the scope of the project and allows for tailored solutions that address specific challenges within the organic waste sector.
3	Focus	The Recognizing strategy and system performance	It emphasizes not only the development of accurate recognition algorithms but also the practical effectiveness and efficiency of the system in real-world scenarios.

4	Evaluative Question (EQ)	a. EQ1-How to develop a vision system-based waste object recognition system b. What is the measure of the accuracy of a vision system-based waste object recognition system?	It ensures that the development is assessed on both methodology and performance. EQ1 ensures that the process of building the system is well-documented and replicable, while EQ2 focuses on quantifying the effectiveness of the system crucial for validating its utility and reliability.
5	Variable (V)	a. Response (V1-introduction strategy) b. Confusion matrix (V2- loss, accuracy, precision, recall)	Defining variables provides clear metrics for assessment. V1 allows for evaluating the initial approach or strategy employed in developing the system. On the other hand, V2 offers quantitative measures of the performance of the system through a confusion matrix.

The research development process was based on the Machine Learning/Deep Learning method advocated by John, et al [23], with careful adjustments made to suit our specific research requirements. These adjustments involved integrating the machine with a waste detection model leading to the development steps depicted in Figure 1.

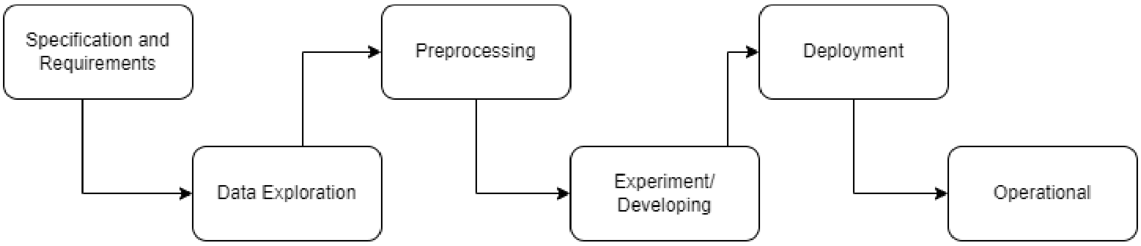


Figure 1. The development stages

The primary objective was to enhance a computer vision-based system for recognizing organic waste objects. The experiments centered on two primary methods, namely CNN and ACNN. A comparison of their performance was conducted to determine if ACNN could improve the accuracy of organic waste identification compared to CNN. The experimental requirements for this research included (1) Google Colaboratory with T4 GPU Processor; (2) Python 3.7; (3) Tensorflow 2.6; (4) Visualhard; (5) Python Imaging Library (PIL); (6) Matplotlib; (7) CV2, and (8) Numpy.

The first step was to select an appropriate dataset for the experiment. Datasets played a crucial role in system testing. The dataset utilized in this experiment was waste classification data (Sashaank Sekar), obtained from: https://www.kaggle.com/datasets/techsash/waste-classification-data/data, under license: CC BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0/). The dataset contained over 25077 images of organic and non-organic waste encompassing a variety of fruits, vegetables, and food waste. Each image was labeled. The dataset was divided into two subsets for valid experiments, namely 85% for training data and 15% for testing data.

The data preprocessing stage in this study was conducted to prepare image data for deep learning model training. The initial step involved normalizing the pixel values using `rescale=1.0/255.0`, converting the image pixel values into the range [0, 1], and enabling data normalization. Subsequently, data augmentation was performed with various defined transformations, including rotation, shift, shear, zoom, and horizontal reversal. This augmentation aimed to create variations in the training dataset, such as changes in object orientation, position, and size. The next step involved determining the batch size with `batch size=32`. It indicated the number of images to be processed in each iteration during training and testing. Finally, a data generator was created using `ImageDataGenerator` to load images from the specified directory, resize the images to uniform dimensions, and classify the images into batches based on the specified sizes. Further, image labels corresponding to classes were defined as one-hot encoding vectors for multi-class classification problems.

After experimenting, the next step was to compare the results of Loss, Accuracy, Precision, and Recall values based on the confusion matrix.

RESULTS AND DISCUSSIONS

The model in this paper adapts the CPS 5C Architecture by Lee et al. [24], and has been adjusted to fit the research objectives. As evidence, the CPS architecture covers Connection, Conversion, Cyber, Cognition, and Configuration Levels. Therefore, a waste object recognition architecture based on a vision system is depicted in Figure 2 subsequently:

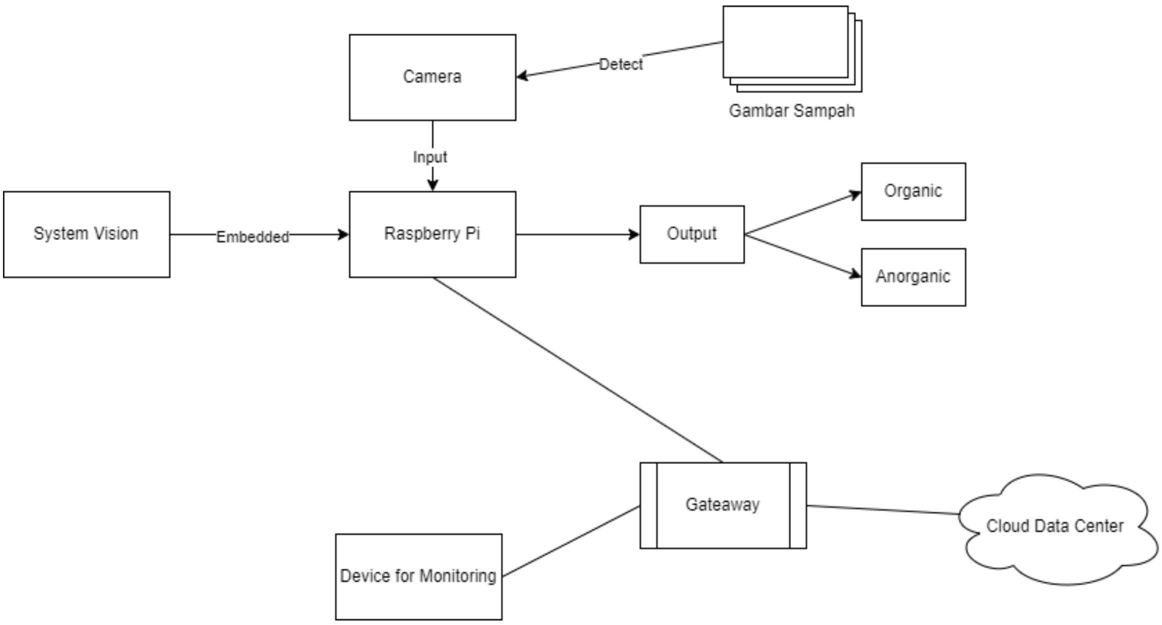


Figure 2. The architecture of waste object recognition

In Figure 2, *the Connection level* or Smart Connection level aspect involved connecting all components of the waste object recognition machine, including a camera, an embedded vision system, and a WiFi adapter directly to the Raspberry Pi. The camera captured trash images processed to identify the type of trash afterward. An embedded vision system adopting the Attention Convolutional Neural Networks (ACNN) algorithm was integrated into the Raspberry Pi for trash detection. Likewise, the WiFi adapter enabled it to connect to a gateway for monitoring the trash detection component with a monitoring device.

The conversion of data to information was undertaken by processing images of trash captured with the camera on the Raspberry Pi. The informative results would be stored in the data center as historical data for later use at the cyber level. The quality of the data-to-information conversion at this stage would determine the self-awareness capabilities of the designed machine. Historical data stored in the data center would be retrieved and compared with current information to accurately predict future system behavior. This process occurs at the cyber level. The activities of the waste object detection component would be continually monitored by monitoring devices. If an error occurred in an existing component, it would be immediately detected. Further, the error data would become historical data stored in the data center for further comparison.

At the *Cognition level*, the focus was situated on monitoring the waste object recognition tools through special devices connected directly to the internet network. This was necessary to obtain real-time monitoring results for both equipment damage and normal operation. The monitoring results would inform decision-making processes undertaken by developers. The decision could also be facilitated with the Decision Support System (DSS) embedded in the monitoring device which provides recommended solutions based on monitoring data from the waste object recognition device.
The decisions made at the *Cognition level* would be executed at the Configuration level enabling the tool to self-configure and adapt. Further, the *Configuration level* serves as a means of organizing all activities performed by the architecture. This includes executing corrective and preventive decisions previously made

at the Cognition level for the waste object recognition system [19]. The utilized vision system is based on the CNN algorithm with the *Attention* Mechanism layer incorporated to form the architecture depicted in Figure 3.

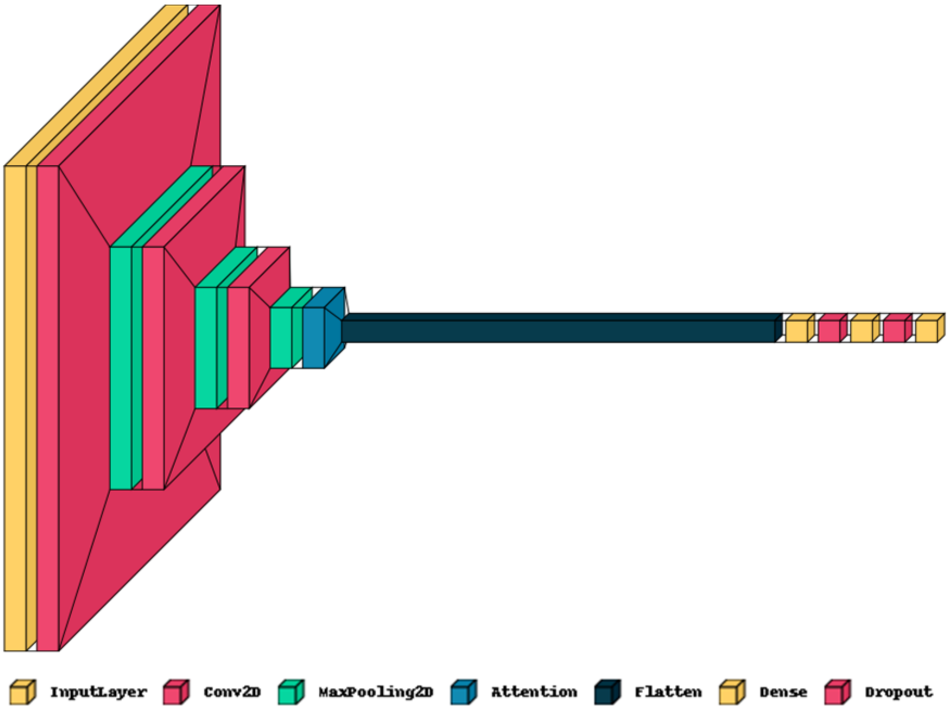


Figure 3. The model architecture

The architecture in Figure 3 is constructed with multiple layers, such as Input Layer, Conv2D Layer, Max Pooling Layer, Attention Mechanism Layer, Flatten Layer, Dense Layer, and Dropout Layer. Table 2 displays the layer types, image size within each layer, number of layers for each type, parameters count, and connectivity between the layers.

Table 2. The model architecture

Layer (Type)	Output Shape	Param #	Connected to
input_2 (InputLayer)	(None, 224, 224, 3)	0	-
conv2d_3 (Conv2D)	(None, 224, 224, 32)	896	['input_2[0][0]']
max_pooling2d_3 (MaxPooling2D)	(None, 112, 112, 32)	0	['conv2d_3[0][0]']
conv2d_4 (Conv2D)	(None, 112, 112, 64)	18496	['max_pooling2d_3[0][0]']
max_pooling2d_4 (MaxPooling2D)	(None, 56, 56, 64)	0	['conv2d_4[0][0]']
conv2d_5 (Conv2D)	(None, 56, 56, 128)	73856	['max_pooling2d_4[0][0]']
max_pooling2d_5 (MaxPooling2D)	(None, 28, 28, 128)	0	['conv2d_5[0][0]']
attention_1 (Attention)	(None, 28, 28, 128)	0	['max_pooling2d_5[0][0]','max_pooling2d_5[0][0]']
flatten_1 (Flatten)	(None, 100352)	0	['attention_1[0][0]']
dense_3 (Dense)	(None, 128)	1284518	['flatten_1[0][0]']
dropout_2 (Dropout)	(None, 128)	0	['dense_3[0][0]']
dense_4 (Dense)	(None, 64)	8256	['dropout_2[0][0]']
dropout_3 (Dropout)	(None, 64)	0	['dense_4[0][0]']
dense_5 (Dense)	(None, 2)	130	['dropout_3[0][0]']

The architecture began with an input layer designated as input_2. It accepted images with a shape of (224, 224, 3) meaning each image was 224 pixels high and 224 pixels wide with three color channels (RGB). This layer did not involve any parameters as it served as the entry point for the data into the network. The first convolutional layer (conv2d_3) followed the input layer. This layer adopted 32 filters of size 3x3 to the input images resulting in an output shape of (224, 224, 32). The total number of parameters in this layer was 896 which includes the weights and biases associated with the convolutional filters. Next, a max-pooling layer (max_pooling2d_3) reduced the spatial dimensions of the output from conv2d_3 by half. Also, it produced an output shape of (112, 112, 32). This operation helps in down-sampling the feature maps reducing the computational load and capturing dominant features.

Following the max-pooling layer is another convolutional layer (conv2d_4) increasing the depth of the network by applying 64 filters of size 3x3. The output shape of this layer was (112, 112, 64) and it involved 18,496 parameters. This layer captured more complex features from the input images. Another max-pooling layer (max_pooling2d_4) was employed next. In a similar vein, it reduced the spatial dimensions of the feature maps to (56, 56, 64). This step continued the process of down-sampling to focus on the most significant features. The third convolutional layer (conv2d_5) increased the depth to 128 filters of size 3x3 resulting from an output shape of (56, 56, 128). This layer has 73,856 parameters and it is responsible for extracting even more detailed features from the input data.

The corresponding max-pooling layer (max_pooling2d_5) reduces the dimensions again to (28, 28, 128). This layer helped in further condensing the feature maps retaining only the most prominent patterns. An attention mechanism (attention_1) was applied next. This mechanism helped the network focus on the most relevant parts of the feature maps with the same input twice resulting in an output shape of (28, 28, 128). *Attention* mechanisms played a crucial role in enhancing the performance of the network by highlighting significant features. The flattened layer, flatten_1, transforms the 3D output of the *attention* layer into a 1D vector of size 100,352. This step prepared the data for the fully connected (dense) layers. The first dense layer, dense_3, connects all 100,352 input features to 128 neurons, resulting in 1,284,518 parameters. This layer helps in learning complex representations by combining the extracted features. A dropout layer, dropout_2, follows to prevent overfitting by randomly setting a fraction of input units to zero during training. This layer does not involve any parameters and maintains the output shape of (128). The second dense layer (dense_4) reduced the dimensionality by connecting 128 neurons to 64 neurons involving 8,256 parameters. This layer further refined the learned features. Another dropout layer (dropout_3) was applied subsequently to continue combating overfitting. Like the previous dropout layer, it maintained the output shape. Conversely, it did not engage any parameters.

Finally, the output dense layer (dense_5), connects the 64 neurons to 2 output neurons, corresponding to the number of classes in the classification task. This layer had 130 parameters and produced the final classification results. *Attention* layers in the aforementioned architecture enhanced feature extraction by directing the focus model is crucial image components for a given task. This led to better feature extraction and increased accuracy in image detection.
The performance results of the two approaches in recognizing organic waste objects are presented below. Figures 4 to 11 provide a visualization of the value comparison results. In particular, table 3 displays the metric values in detail from the comparison results.

Table 3. Experiment result

Metrics	***CNN***	***Attention CNN***
Loss	0.3379	0.3329
Accuracy	0.8670	0.8693
AUC	0.9333	0.9340
Precision	0.8682	0.8689
Recall	0.8660	0.8698
Validation Loss	0.4116	0.3079
Validation Accuracy	0.8500	0.8882
Validation AUC	0.9148	0.9445
Validation Precision	0.8503	0.8879
Validation Recall	0.8496	0.8886

In this experiment, two distinct neural network architectures, namely Convolutional Neural Network (CNN) and Attention CNN were evaluated on a specific task. The performance of these models was assessed using various metrics commonly employed in machine learning. The following metrics were measured for both the CNN and Attention CNN models:

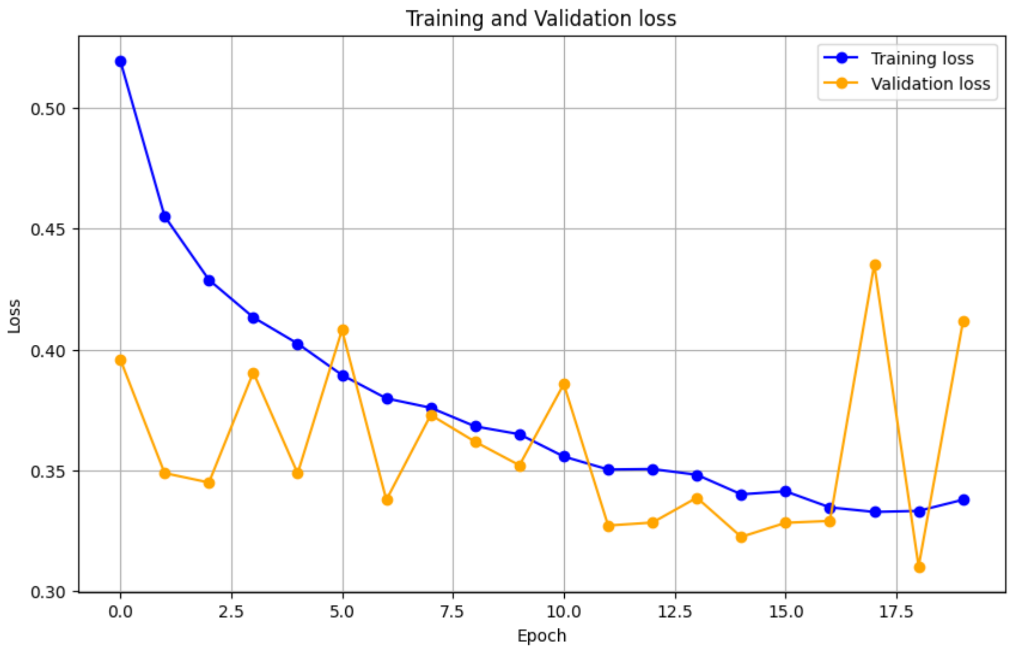


Figure 4. CNN model loss

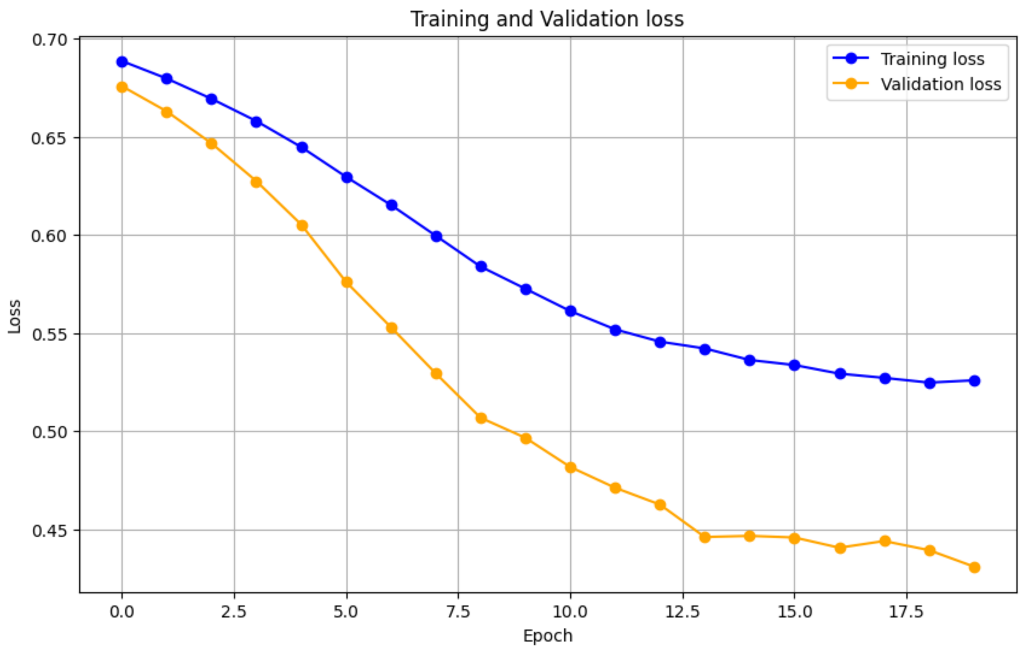


Figure 5. ACNN model loss

The *loss* metric provides an indication of the error during the training process with lower values deciphering better performance. The *attention* mechanism allows the model to focus on relevant parts of the input data which potentially reduces information redundancy and facilitates more efficient learning. This focused *attention* contributes to a more stable reduction in loss during training. Figure 4 reveals the *loss* values for the CNN model. Likewise, figure 5 presents the *loss* values for the ACNN model. The *loss* metric is crucial as it unveils how well the model is performing. On the contrary, lower loss values signify better performance. The comparison of these figures helps in understanding the efficiency of incorporating the *attention* mechanism into the CNN architecture. In particular, the ACNN model represents a slightly lower *loss* value (0.3329) compared to the CNN model (0.3379). In other words, it suggests that the *attention* mechanism reinforces the training efficiency of the model and reduces the error rate.

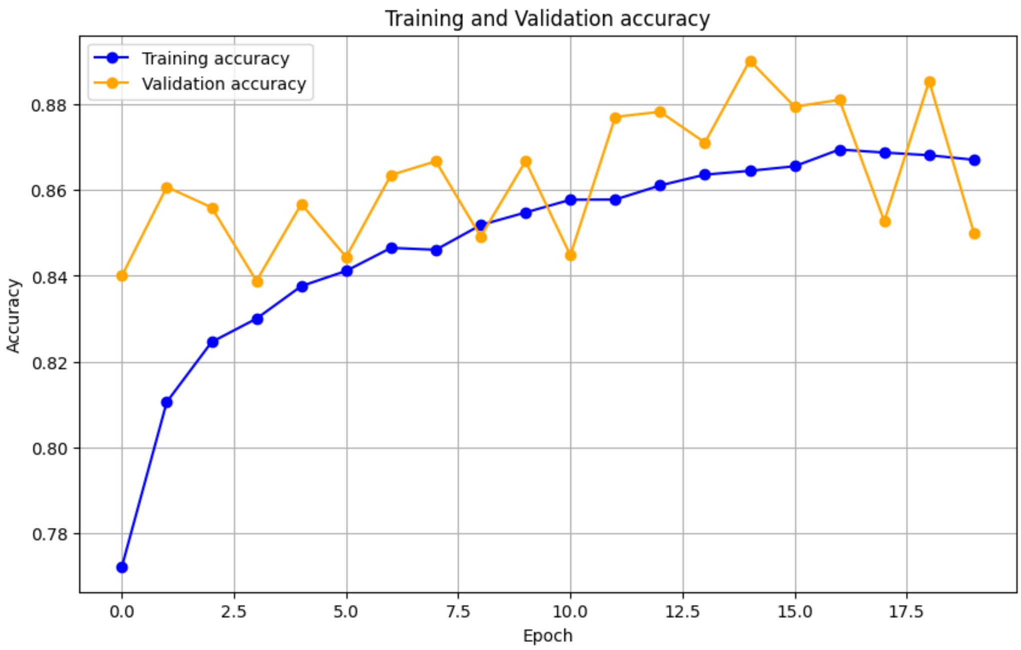
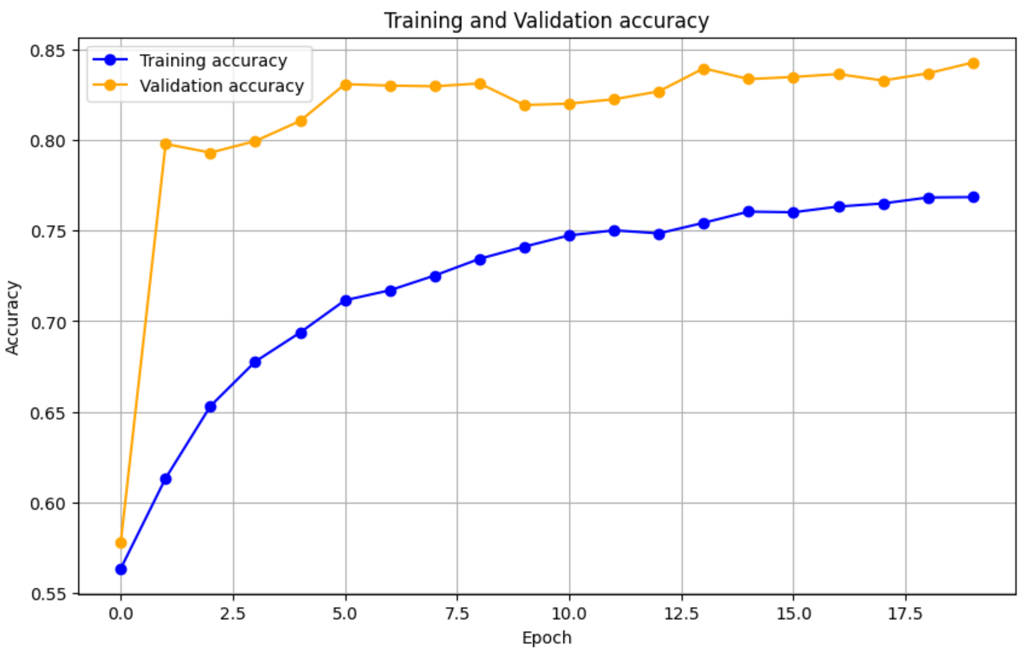


Figure 6. CNN model accuracy



\

Figure 7. ACNN model accuracy

Accuracy measures the overall correctness of the model predictions and is expressed as a ratio of correctly predicted instances to the total instances. *Attention* mechanisms enable the model to selectively attend to important features improving its ability to make accurate predictions. This *attention* to relevant information may lead to a more stable and consistent increase in overall *accuracy* throughout the training process. Figure 6 shows the *accuracy* values for the CNN model. On the other side, figure 7 presents the *accuracy* values for the ACNN model. In other words, *accuracy* measures the entire correctness of the model predictions and it is expressed as a ratio of correctly predicted instances to the total instances. The essence of these images is to disclose how well each model performs in terms of making accurate predictions. The comparison of these figures helps in understanding the efficiency of incorporating the *attention* mechanism into the CNN architecture. Specifically, the ACNN model deciphers a slightly higher accuracy (0.8693) compared to the CNN model (0.8670) suggesting that the *attention* mechanism increases the ability of the model to produce accurate predictions.

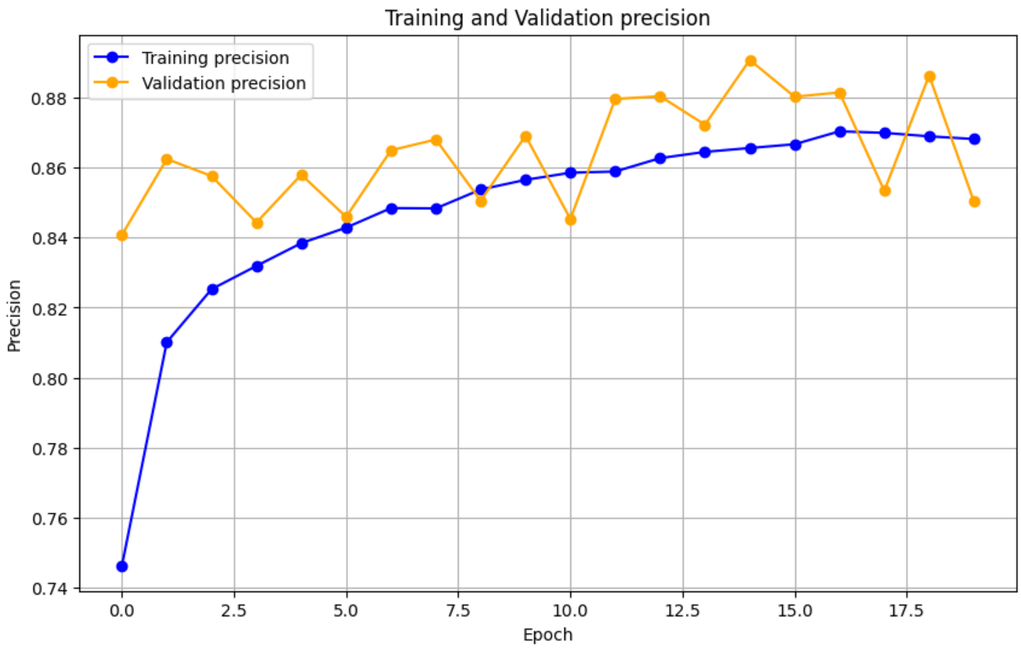


Figure 8. CNN model precision

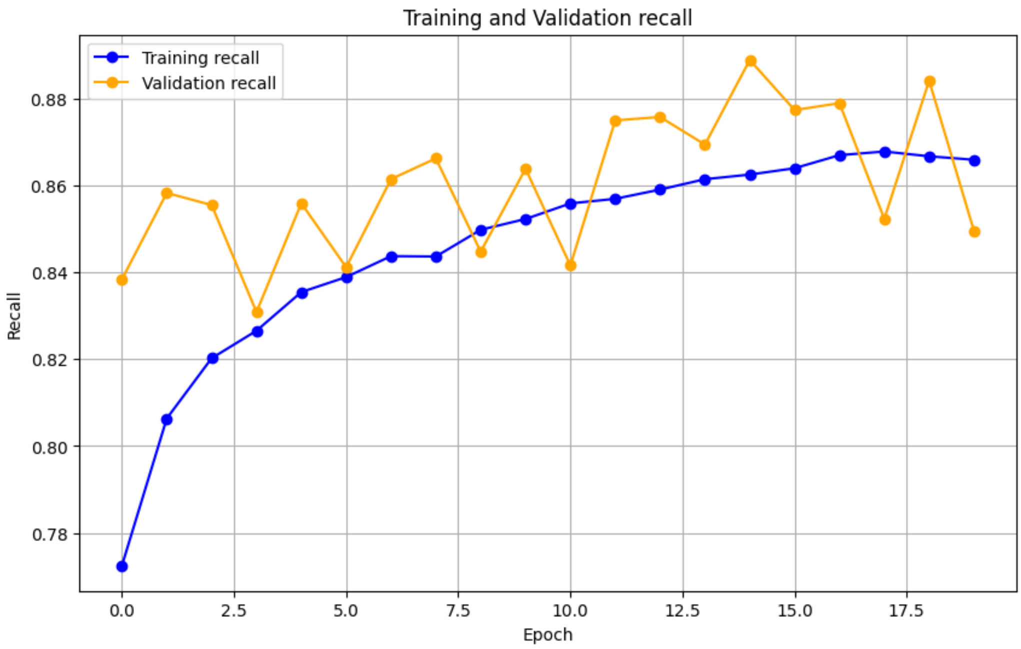


Figure 9. ACNN model precision

Precision evaluates the *accuracy* of positive predictions indicating the proportion of correctly predicted positive instances among all instances predicted as positive. By attending to salient features, the attention mechanism can refine the precision of the model in identifying positive instances. This focused attention may lead to a more stable and higher precision score throughout the training iterations. Figure 8 illustrates the *precision* values for the CNN model. On the other hand, figure 9 provides the *precision* values for the ACNN model. The essence of these images is to delineate how well each model performs in terms of producing accurate positive predictions. By attending to salient features, the *attention* mechanism in the ACNN can refine the precision of the model by identifying positive instances. This focused *attention* may lead to a more stable and higher precision score throughout the training iterations. Thus, these images provide a quantitative evaluation of the models outlining that the ACNN model may outperform the CNN model in terms of *precision* metrics. This is a key indicator of the ability of the model to produce accurate positive predictions during training.

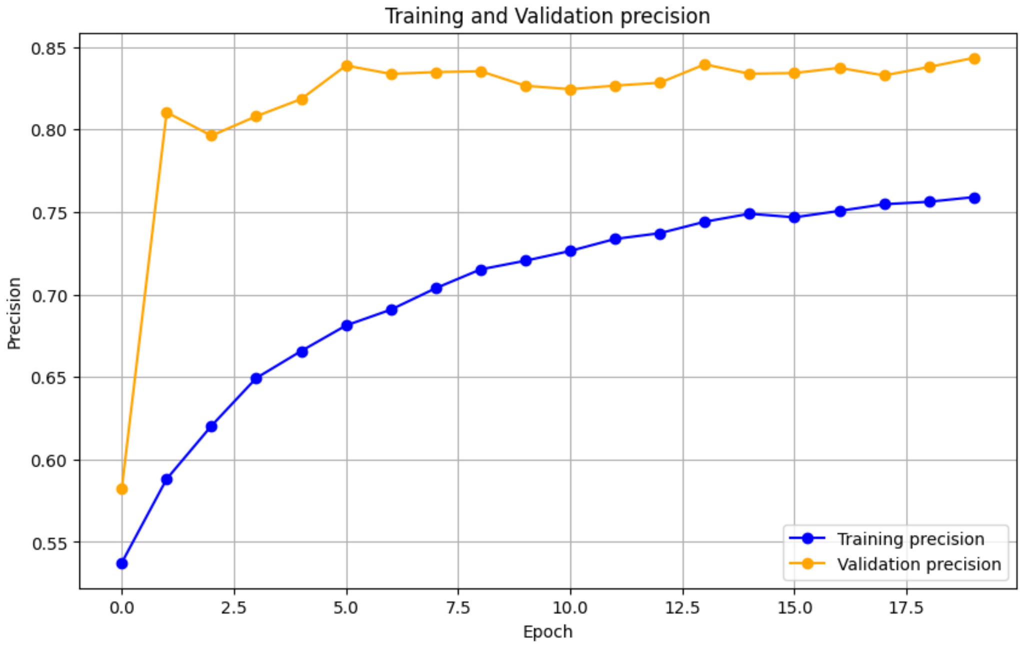


Figure 10. CNN model recall

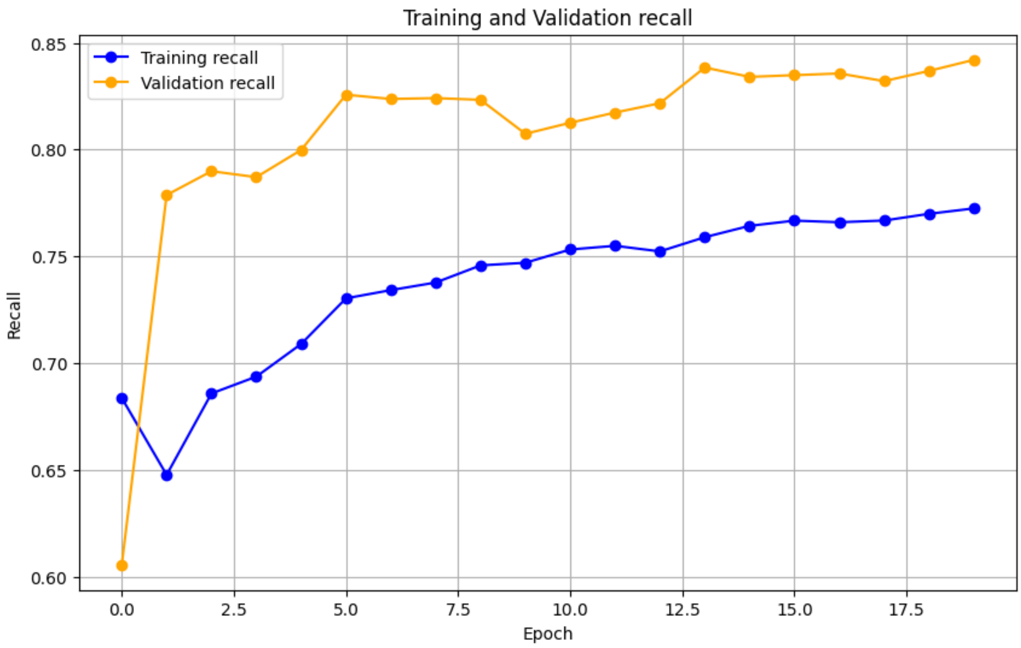


Figure 11. ACNN model recall

Recall, also known as sensitivity or true positive rate, measures the ability of the model to correctly identify all relevant instances. The capability of the attention mechanism to selectively focus on the relevant information can aid in capturing more positive instances. This targeted attention may result in a more stable and consistently higher *recall* rate throughout the training process. Figure 10 displays the *recall* values for the CNN model. On the other hand, figure 11 presents the *recall* values for the ACNN model. The essence of these images is to demonstrate how well each model performs in terms of identifying all relevant positive instances. The attention mechanism in the ACNN can selectively focus on relevant information, potentially assisting in capturing more positive instances. This targeted attention may result in a more stable and consistently higher *recall* rate throughout the training process.

The research model outperforms all metric values in the evaluation results as shown in table 4 and figures 4 to 11. The investigative results demonstrate that the ACNN model reached 86.93% accuracy. In other words, it is slightly higher than the accuracy standard of the CNN model attaining 86.70%. In addition, the evaluation results revealed that the ACNN model also showed improved performance on several metrics, including a lower loss value (0.3329 vs. 0.3379), a higher AUC (0.9340 vs. 0.9333), higher precision (0.8689 vs. 0.8682), and higher recall (0.8698 vs. 0.8660). Statistically, this research confirms a significant improvement and showcases that the ACNN model is more effective in distinguishing complex waste objects, optimizing prediction accuracy, and capturing all relevant occurrences. The enhancement of the ACNN model can significantly increase the efficiency and accuracy of waste object recognition systems by leading to better sorting and recycling processes, reducing environmental impact, and increasing operational efficiency and potential cost savings in waste management. Overall, the superior performance of the ACNN model highlights its potential to revolutionize waste management practices by enhancing scalability and reliability.

CONCLUSION

Test results demonstrate that ACNN displayed slightly better improvements than CNN in terms of loss metrics with a lower loss value (0.3329 vs. 0.3379). This suggests that integrating ‘attention’ into the CNN enhances efficiency in model training, leading to a reduced error rate. Besides, ACNN has a slightly higher accuracy of 0.8693 compared to CNN's 0.8670. Also, it illustrates a slightly better area under the curve (AUC) of 0.9340 compared to CNN's 0.9333. This proves that ACNN is more accurate in distinguishing organic and inorganic waste objects. Furthermore, when evaluating precision and recall, the ACNN model consistently provides a precision of 0.8689 and a recall of 0.8698. It is also evident that the ACNN model requires a longer validation time and has a lower validation loss (0.3079 compared to 0.4116). In other words, this model is more complex and demands more computing resources. Using ‘attention’ with CNNs has been shown to enhance the recognition of organic waste objects resulting in improved accuracy and AUC. Nevertheless, the trade-off between performance and efficiency should be taken into account as there is an increase in training and computing time. In summary, this study has developed a system utilizing the ACNN model for image processing and enabled the detection of organic waste objects. This system has the potential to improve the efficiency of organic waste disposal, lessen the environmental impact of waste, and automate and improve the accuracy of the organic waste separation process.

This paper suggests potential future research directions indicating that initial experiment results could be further explored by delving into various aspects in more detail. For instance, future research may enhance our understanding by examining the use of larger and more diverse datasets and improving the generalization capability of ACNN models. In addition, it is pivotal to consider adjusting the ACNN architecture by adding or extending specific layers to assess their impact on model performance. Moreover, this study demonstrates the promising potential of using ‘attention’ in identifying organic waste objects. Therefore, the recommended next step is to pore over the application of computer vision technology in other fields, such as industry or security. Further, it is crucial to develop practical solutions for implementing the ACNN model, including the need for a more comprehensive investigation into its application in a holistic waste management system (its role in the waste recycling process).

REFERENCES

[1] B. A. Yulianto, Agus, “Pemkot Tasikmalaya Upayakan Kurangi Sampah ke TPA Ciangir,” *repjabar*, 2022.

[2] D. N. Marpaung, Y. Iriyanti, and D. Prayoga, “Analisis Faktor Penyebab Perilaku Buang Sampah Sembarangan Pada Masyarakat Desa Kluncing , Banyuwangi Departement of Health Policy and Administration , Faculty of Public Health , Banyuwangi,” *Prev. Masy.*, vol. 13, no. 1, pp. 47–57,

2022.
[3] Hunaepi *et al.*, “Pengelolaan Sampah Organik dengan Komposter Untuk Mewujudkan NTB Zero Waste,” *GERVASI J. Pengabdi. Kpd. Masy.*, vol. 5, no. 2, pp. 168–183, 2021.
[4] Khawarita Siregar and Christopher W, “Perancangan Tong Sampah Pupuk Dengan Metode QFD Untuk Mengolah Limbah Organik Menjadi Pupuk Serbaguna,” *Talent. Conf. Ser. Energy Eng.*, vol. 2, no. 3, 2019, doi: 10.32734/ee.v2i3.797.
[5] N. D. Fauzan, M. Ardan, A.-N. Izzah Safina, R. Fattur, and E. Octalyani, “Penggunaan Pupuk Organik Cair sebagai Pengganti Pupuk Kimia di Desa Sidomulyo, Kecamatan Air Naningan,” *Altruis J. Community Serv.*, vol. 2, no. 2, 2021, doi: 10.22219/altruis.v2i2.15977.
[6] M. M. Thaha, K. P. M. Kumar, B. S. Murugan, S. Dhanasekeran, P. Vijayakarthick, and A. S. Selvi, “Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images.,” *J. Med. Syst.*, vol. 43, no. 9, p. 294, Jul. 2019, doi: 10.1007/s10916-019-1416-0.
[7] D. Liu, H. Ma, Z. Xiong, and F. Wu, “CNN-Based DCT-Like Transform for Image Compression,” in *MultiMedia Modeling*, 2018, pp. 61–72.
[8] A. H. Mawaddah, C. Atika Sari, D. R. Ignatius Moses Setiadi, and E. Hari Rachmawanto, “Handwriting Recognition of Hiragana Characters using Convolutional Neural Network,” in *2020 International Seminar on Application for Technology of Information and Communication (iSemantic)*, 2020, pp. 79–82. doi: 10.1109/iSemantic50169.2020.9234211.
[9] A. Sharma and G. Phonsa, “Image Classification Using CNN,” *SSRN Electron. J.*, no. Icicc, pp. 1–5, 2021, doi: 10.2139/ssrn.3833453.
[10] C. Umam, D. Andi Krismawan, and R. Raad Ali, “CNN for Image Identification of Hiragana Based on Pattern Recognition,” *J. Appl. Intell. Syst.*, vol. 6, no. 2, pp. 62–71, 2021.
[11] N. E. W. Nugroho and A. Harjoko, “Transliteration of Hiragana and Katakana Handwritten Characters Using CNN-SVM,” *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 15, no. 3, p. 221, 2021, doi: 10.22146/ijccs.66062.
[12] M. Zhang and W. Shi, “A Feature Difference Convolutional Neural Network-Based Change Detection Method,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7232–7246, 2020, doi: 10.1109/TGRS.2020.2981051.
[13] Z. Shengyi, Y. Peng, J. Liu, and W. Shuo, “Tomato Leaf Disease Diagnosis Based on Improved Convolution Neural Network by Attention Module,” *Agriculture*, vol. 11, no. 651, pp. 1–15, 2021, doi: https://doi.org/10.3390/agriculture 11070651.
[14] D. Jha, M. A. Riegler, D. Johansen, P. Halvorsen, and H. D. Johansen, “DoubleU-Net: A deep convolutional neural network for medical image segmentation,” *Proc. - IEEE Symp. Comput. Med. Syst.*, vol. 2020-July, no. 1, pp. 558–564, 2020, doi: 10.1109/CBMS49503.2020.00111.
[15] F. Ciancetta, G. Bucci, E. Fiorucci, S. Mari, and A. Fioravanti, “A New Convolutional Neural Network-Based System for NILM Applications,” *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, doi: 10.1109/TIM.2020.3035193.
[16] K. Suresh, M. B. Palangappa, and S. Bhuvan, “Face Mask Detection by using Optimistic Convolutional Neural Network,” in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, 2021, pp. 1084–1089. doi: 10.1109/ICICT50816.2021.9358653.
[17] K. M. Sandi, A. P. Yudha, N. D. Aryanto, and M. A. Farabi, “Klasifikasi Sampah Organik dan Non-Organik Menggunakan Convolutional Neural Network,” *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 1, pp. 72–81, 2022, doi: 10.28932/jutisi.v8i1.4314.
[18] K. Cheinski and P. Wawrzynski, “DCT-Conv: Coding filters in convolutional networks with Discrete Cosine Transform,” *Proc. Int. Jt. Conf. Neural Networks*, 2020, doi: 10.1109/IJCNN48605.2020.9207103.
[19] Y. Peng, X. He, and J. Zhao, “Object-part attention model for fine-grained image classification,” *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1487–1500, 2018, doi: 10.1109/TIP.2017.2774041.
[20] E. Pranav, S. Kamal, C. Satheesh Chandran, and M. H. Supriya, “Facial Emotion Recognition Using Deep Convolutional Neural Network,” in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2020, pp. 317–320. doi: 10.1109/ICACCS48705.2020.9074302.
[21] T. Xiao, Y. Xu, K. Yang, J. Zhang, Y. Peng, and Z. Zhang, “The application of two-level attention models in deep convolutional neural network for fine-grained image classification,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 842–850, 2015, doi: 10.1109/CVPR.2015.7298685.
[22] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in Software Engineering*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. doi: 10.1007/978-3-

642-29044-2.
[23] M. M. John, H. H. Olsson, and J. Bosch, “Developing ML/DL Models: A Design Framework,” in *2020 IEEE/ACM International Conference on Software and System Processes (ICSSP)*, 2020, pp. 1–10.
[24] J. Lee, B. Bagheri, and H. A. Kao, “A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems,” *Manuf. Lett.*, vol. 3, pp. 18–23, 2015, doi: 10.1016/j.mfglet.2014.12.001.