Analysis of Chatbot-Based Image Classification on Social Commerce LINE@ Platform

Jennifer Gabrielle Nangoy[1], Nabila Husna Shabrina[2]
[1,2]Internet of Things Research Group
Universitas Multimedia Nusantara
Tangerang, Indonesia
[1]gbrlle.jennifer@gmail.com, [2]nabila.husna@umn.ac.id

***Abstract*—The rapid development of internet has influenced people's lifestyle to become more digital. Social commerce is one of the e-commerce categories where sellers offer their products through social media. LINE Messenger provides a place where sellers and buyers can communicate to carry out transaction processes. This paper discusses about the research on a chatbot that is useful for handling picture messages by providing product information replies. The chatbot was built using the Convolutional Neural Network method for the image classification process. The tests conducted found the chatbot to be able to reply to buyer messages with an accuracy level of 0.68.**

Keywords—social commerce, LINE Messenger, chatbot, Convolutional Neural Network, image classification

I. INTRODUCTION

The development of internet affects the buying and selling process. Previously, one had to go to the store to buy and make transactions to get an item, however now the buying and selling processes can be done online. This change brings benefits for both sides; on the seller side, they do not need to pay for operations and on the buyer side, they can make purchases anywhere and anytime with a variety of payment and shipping options.

The development of internet supports the emergence of e-commerce. E-commerce is defined as the process of buying and selling goods online which is divided into 2 categories, the first category is e-tailing which means buying and selling transactions are carried out through a platform that has been provided and the second category is social commerce which means offering goods online through social media with a variety of payment and shipping methods [1].

LINE@ is a feature of the LINE application that supports social trading. LINE@ is a LINE account that can be used by entrepreneurs and companies to spread information about companies, products, and others using the features offered by LINE@ [2]. The LINE@ application is used by sellers and buyers for buying and selling transactions by sending messages that usually begin by asking for the offered product information until the transaction is completed and the buyer provides a product review.

The survey was conducted on 70 male and female respondents with a participation ratio of 1: 1 between 15-40 years resulting in more than 85% of female respondents and 57% of male respondents conducting transactions through LINE @. 41% of respondents are aware of the products offered through the Instagram application, while 17% found out through the promotion of LINE @ Official Accounts or LINE timelines and the rest got recommended by their friends, family and views from the website or Facebook.

Based on data collected from online stores, the buyers will see the products offered through other platforms and will send screenshots of products to LINE@ to request product information as shown in Fig. 1, which followed by a respond from the seller until a transaction occurs

Fig. 1. Example of LINE@ Conversation

However, the seller will be overwhelmed if they receive a lot of messages in one day and it takes them a long time to be able to reply to all messages. The long response time is also the basis for the buyer to determine whether to buy at the store or to find another store. To overcome this, the seller will need to hire more people to reply to messages, however it can only apply for working days and hours.

Image Classification aims to be able to categorize images automatically according to classifications [3]. Research conducted by Rika Rohana et al. found that Convolutional Neural Networks (CNN) are superior in terms of accuracy, sensitivity and specificity compared to other methods such as Support Vector Machine (SVM) and k-Neural Network (kNN) [4]. Another study was conducted by Yiyu Hong and Jongwoen Kim regarding comparison of CNN architecture and Scale-Invariant Feature Transform (SIFT) with painting classification as the case study. Based on the research, it was found that the CNN architecture has a lower error rate of 2% compared to SIFT which has an error rate of 15% [5]. Another research utilizing the CNN method was also carried out by S. Visalini to classify traffic signs and obtained accuracy rate of 85% - 90% [6].

A chatbot is an artificial intelligence that can respond to interactive human conversations with predefined phrases or text. A chatbot can be used for customer service or marketing systems in social media or instant messaging clients [7]. This paper aims to explore image classification using the

2020 7th NAFOSTED Conference on Information and Computer Science (NICS) | 978-0-7381-0553-6/20/$31.00 ©2020 IEEE | DOI: 10.1109/NICS51282.2020.9335874

Authorized licensed use limited to: San Francisco State Univ. Downloaded on July 02,2021 at 09:37:15 UTC from IEEE Xplore. Restrictions apply.

Convolutional Neural Network method implemented on customer service chatbot and to determine the reply according to the image sent by the buyer. This goal of the system is to automize the interaction between the seller and the prospective customer.

II. LITERATURE REVIEW

A. TensorFLow

TensorFlow is an open-source library for high-performance numerical computing that has a flexible architecture providing ease of use for computing on various platforms (CPU, GPU, TPU) [8]. TensorFlow uses data flow graphs for numerical computation. Nodes in the graph represent mathematical operations, while edges in the graph represent the flow of multidimensional data arrays (tensors) between the two [9].

B. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a class of artificial neural networks that uses a convolution layer to filter useful information for the input. Convolution operation involves combining input data (feature map) with the convolution kernel (filter) to form a modified feature map. Filters in the convolution layer are modified based on the researched parameters to extract the most useful information for a particular task. CNN adjusts automatically to find the best features based on the task [10].

CNN accepts the input of feature maps in the form of three-dimensional matrices where the size of the first two dimensions corresponds to the length and width of the image in pixels. The third dimension size is 3 (corresponds to 3 channel color images: red, green, blue) [11].

CNN consists of several layers:

1) Convolutional Layer

This layer is the main process that underlies a CNN. Convolution is a mathematical term that means to apply a function to the output of another function repeatedly. In image processing, convolution can be interpreted to apply the kernel to the image in all possible offsets. The kernel moves from the upper left corner to the lower right.

2) Pooling Layer

Pooling layer is usually after the convolutional layer. In principle, the pooling layer consists of a filter of a certain size and stride that will move around the entire feature map areaThe pooling methods that are usually used consist of Max Pooling and Average Pooling. Max pooling is a technique used to reduce the dimensions of the image by taking over the maximum pixel value of the grid in order to help reducing overfitting and making the model more general. Average pooling will take the average pixel grid value.

3) Fully Connected Layer

The fully connected layer aims to transform the dimensions of the data, as such the data can be classified linearly. Each neuron in the convolutional layer needs to be transformed into one-dimensional data before it can be put into a fully connected layer.

C. Chatbot

A chatbot is a program created to allow for communication with humans which is supported by rules and artificial intelligence. The use of chatbot has been done, among others, as a command control, customer service, virtual assistants, and messaging applications. The general method used chatbot is to interpret the keywords obtained from the input and to respond with the keywords that best match the pattern of words that are similar from the data that has been stored in the database [12].

III. SYSTEM DESIGN AND IMPLEMENTATION

A. System Design

1) General description of the system

The proposed system is built to classify images on chatbots contained on the LINE @ platform. The system is designed to respond to buyer questions regarding product information offered by the seller. Product information is obtained from the administrator control panel. The general picture of the proposed system is shown in Fig. 2.

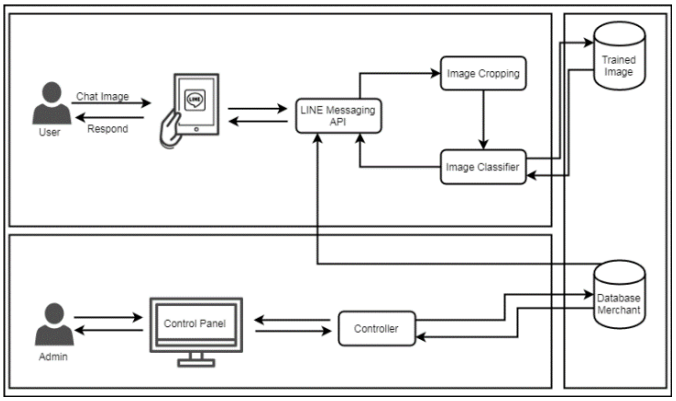

Fig. 2. General descriptions of the system

The first step of building the system is training data. The dataset is collected by making a screenshot on the online shop's Instagram account and cropping the image with a ratio of 1: 1. The image used for the dataset is an image that contain only 1 product in 1 image as shown in Fig. 3. The dataset image is stored in a folder with each other's labels.

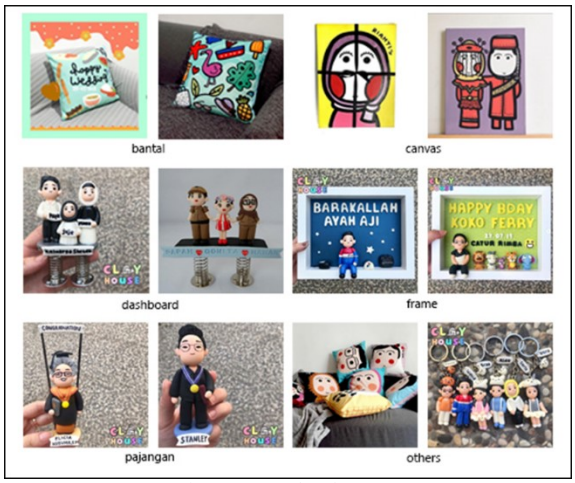

Fig. 3. Example of Dataset

The dataset is augmented to obtain a wider and more variety of number of images. The augmentation results was then resized and be used in the training process. The examples of augmentation results used for the dataset are

Authorized licensed use limited to: San Francisco State Univ. Downloaded on July 02,2021 at 09:37:15 UTC from IEEE Xplore. Restrictions apply.

shown in Fig. 4. The training process was carried out using the CNN architecture with the TensorFlow framework.

Chatbot starts to work when the buyer sends the picture to the LINE @ seller account. The message will be sent to the server by the LINE Messaging API and then stored in a temporary folder. The image will be cropped to focus only on the product. Cropping process is carried out by finding the contour from the image and a box will be made based on the largest area obtained from the contour. Predicted images will be directly predicted to determine the label image. After the image label is obtained, the label will be reversed in the LINE Messaging API to be further matched with product information in the database. If a product is found, then a message will be returned to the buyer in the form of product information including the name, price, stock availability and length of time for the custom product.

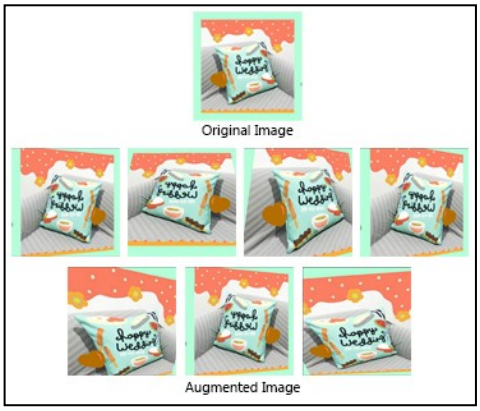

Fig. 4. Augmented Image

The administrator control panel subsystem is utilized by the online shop administrator to add, change or delete products. Administrators can also view transaction data through the control panel. This subsystem was created as a supporter of the application and was not the focus of this research.

2) Flowchart

The system workflow is divided into a chatbot subsystem and control panel subsystem. The chatbot subsystem starts with image training using the CNN method. The chatbot subsystem will also receive picture messages to classify and generate the correct reply message to the buyer.

The training process, as shown in Fig. 5, begins with augmentation of the dataset image, then the image is resized and stored in a folder according to the label. In the next step, CNN architecture modelling will be used for the training process. After the model is created, an image training process is carried out to classify the image from the buyer.

The chatbot process implemented on the LINE @ platform as shown in Fig. 6 begins with webhook authentication using token access channels and secret channels obtained through LINE Developer. Picture messages from the buyer will be saved and followed by carrying out the process of cropping and image classification. After the label from the classification process is obtained, the label is matched with product data in the database to find product information. If the production information is found, the chatbot will give a reply message to the buyer containing product information. If it is found or the image is classified as "others" then chatbot will give a reply message that the image cannot be processed.

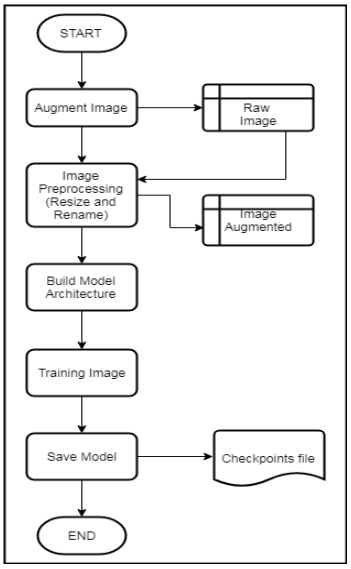

Fig. 5. Training Flowchart

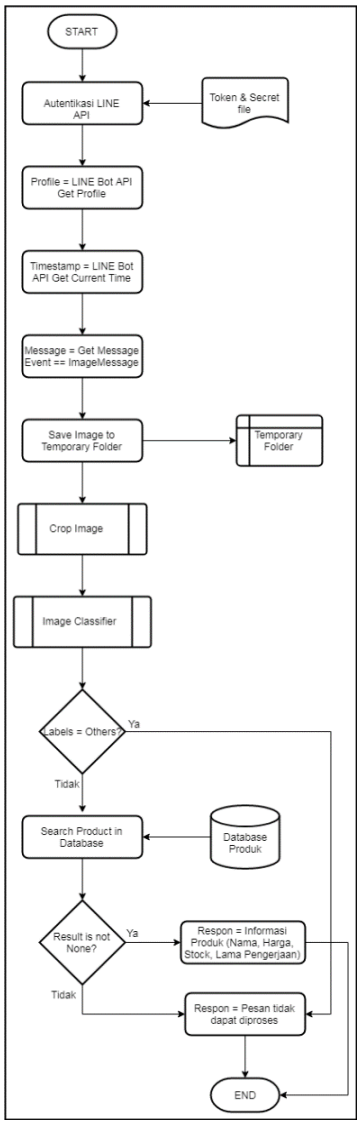

Fig. 6. Chatbot Flowchart

Authorized licensed use limited to: San Francisco State Univ. Downloaded on July 02,2021 at 09:37:15 UTC from IEEE Xplore. Restrictions apply.

The process of cropping the image as shown in Fig. 7 is performed by reading the image stored in the temporary folder. Subsequently, the noise in the image will be reduced and then contour is searched to find the largest area that will be the result of cropping. Cropped images will be stored back to the temporary folder.

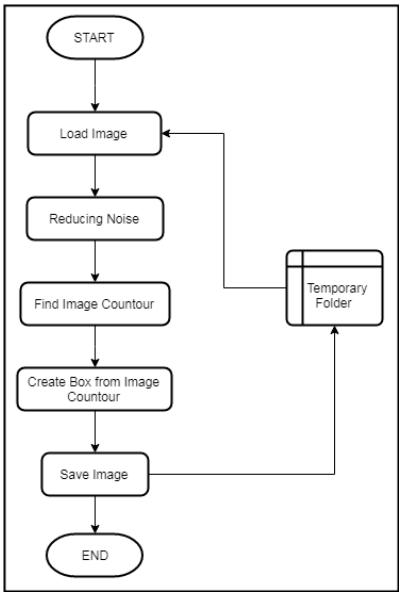

Fig. 7. Cropping flowchart

The image classification process as shown in Fig. 8 starts with taking pictures from the temporary folder which will be resized to a perfect square. By using a model created in the training process, images are sent to the network for classification to obtain image labels. The label is used as a parameter determine he product that the buyer is asking for.

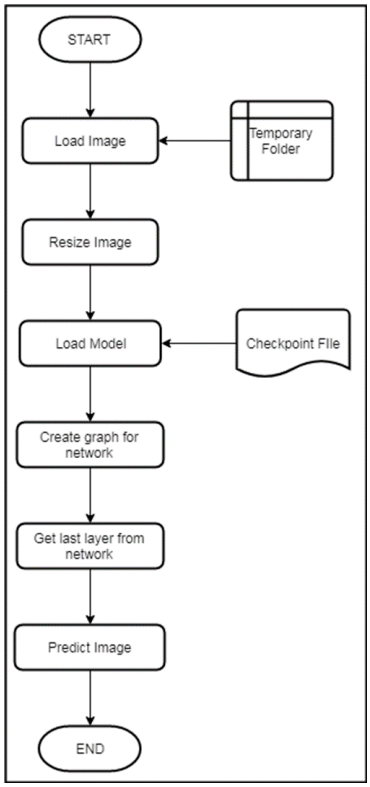

Fig. 8. Image Classifier flowchart

3) Entitry Relationship Diagram

Fig. 9 shows the relation of the tables used in the classification system.

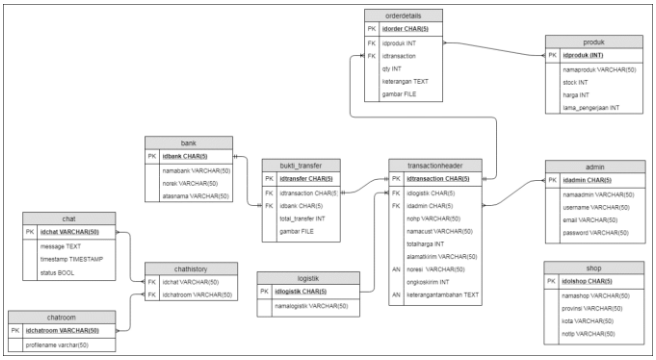

Fig. 9. ERD

The "produk" table contains information about the product information sold by the online shop. The "shop" table contains information about the online shop. The "chatroom" table contains information about the buyer's profile. In this system, only three tables are used.

B. Implementation

This paper carried out an experiment on 4 different CNN models with variations in layers and epochs. Model A uses 3 layers with the number of epoch of 300, model B uses 2 layers with the number of epoch of 40, model C uses 2 layers with the number of epoch of 100 and model D uses 2 layers with the number of epoch of 300. Based on the results obtained from these experiments, the C model results in the best accuracy value compared to other models as summarized in Fig. 10.

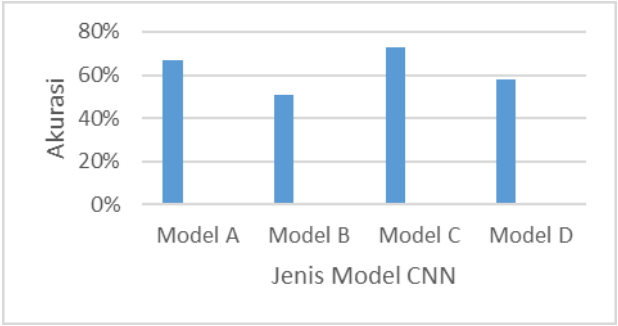

Fig. 10. Comparation of CNN Model

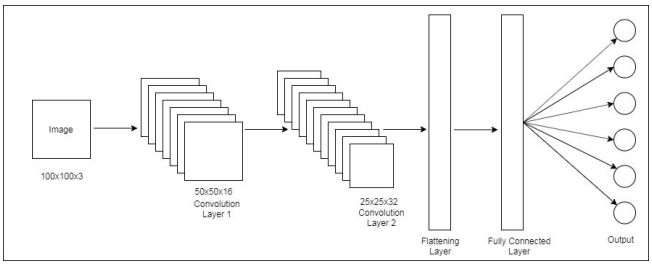

Fig. 11. CNN architecture

Fig. 11 shows the CNN architecture used in this study. The architecture consists of 2 convolutional layers which extract the features of the inputted image. Following the convolution process, the max-pooling process is undertaken to retrieve the largest value from the filter on the feature map

Authorized licensed use limited to: San Francisco State Univ. Downloaded on July 02,2021 at 09:37:15 UTC from IEEE Xplore. Restrictions apply.

and to reduce the size of the feature map. The next process is the activation layer using ReLU. After the image is extracted on the convolution layer, the feature will be reshaped as such it can be an input for the fully connected layer which will determine the image label.

The varying loss values of the training results on the C model architecture can be seen in Fig. 12. The x-axis represents the number of steps taken while the y-axis represents the loss value.

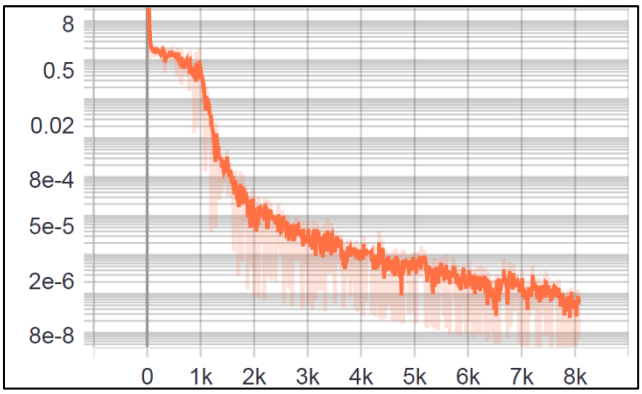

Fig. 12. Loss value

After the training and programming process are finished, testing of the system is completed. Testing is done to obtain the accuracy value of chatbot using a confusion matrix.

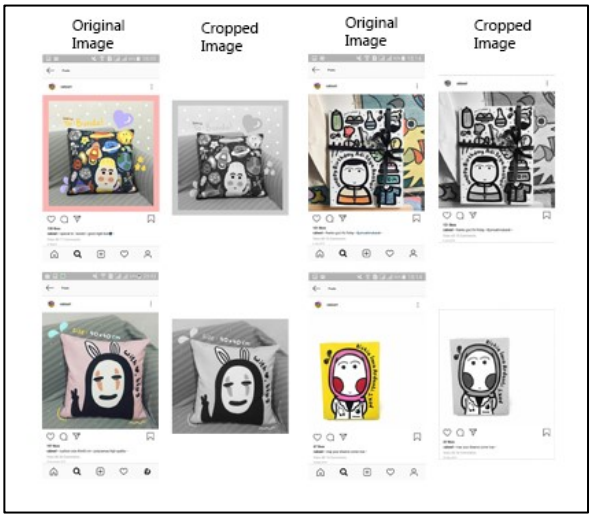

Fig. 13. Cropping result

Testing on the system is undertaken by sending a screenshot scenario from an Instagram online store account via LINE@ with a total of 10 images for each label. The number of correct replies will be used for the confusion matrix calculation for obtaining the overall system accuracy value. Images sent by the buyer will be cropped first. As shown in Fig. 13, the cropping results for some images are unsatisfactory due to the background that is not included in the product. However, the other images are able to be cropped well where only the product is identified.

Table 1 shows the results of the confusion matrix from the image classification results. Equation (1) was used to obtain the accuracy value. TP value = TN because it is the correct number of predictions, hence the attained accuracy value is 0.68.

TABLE 1. CONFUSSION MATRIX RESULT

		True						
		Bantal	Canvas	Frame	Pajangan	Dashboard	Others	Jumlah
Prediction	Bantal	9		1			1	11
	Canvas		10	2		1	2	15
	Frame	1		6	2	2	1	12
	Pajangan				7	1	1	9
	Dashboard				1	4		5
	Others			1		2	5	8
	Total	10	10	10	10	10	10	60

$$Accuration = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (1)$$

Based on the tests result, it was found that some images were incorrectly recognized by the system due to the limitations of images taken from certain online shop accounts, one of the examples is shown in Fig. 14. The used images include similar shapes, such as frames and canvas which have rectangular shapes with sharp elbows and contained Fig.s. Furthermore, images for dashboard displays, frames and displays contain a similarity of having Fig.s in the form of people.

Fig. 14. Example of similar image

In addition to the similarity between one image and another, the background in the product photo also influences the system's ability to recognize images. As seen in Fig. 15, images uploaded by online shop use additional background to enhance the image. It was found that the background becomes the weakness during image extraction as such the image features are not well recognized.

Fig. 15. Image with background addition

Authorized licensed use limited to: San Francisco State Univ. Downloaded on July 02,2021 at 09:37:15 UTC from IEEE Xplore. Restrictions apply.

This chatbot application was created to speed up the process of buying and selling transactions by reducing the time to reply to messages as messages will be replied immediately by the chatbot. The time it takes for a chatbot to reply to a message <1 minute. Fig. 16 shows the time from when the image was sent to the chatbot reply.

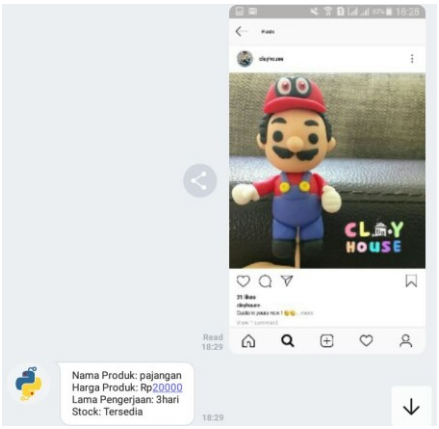

Fig. 16. Chatbot respond time

However, errors in the image classification process may obstruct the process because the provided product information does not match with the sent image, therefore the message needs to wait for the admin to reply with the right message.

IV. CONCLUSIONS AND SUGGESTIONS

A. *Conclusions*

This research concludes that the CNN method can be implemented in chatbots to classify an image and to provide a reply message to the buyer.

The accuracy based on the test results is 0.68. This value is influenced by several factors, namely the similarity between images, background addition to the product image and unsatisfactory cropping results, all of which are causing errors in the image classification process.

Based on the level of accuracy, it can be concluded that the system is unable to speed up the overall transaction time due to image prediction errors that cause incorrect product information delivered and hence the message has to wait for administrators to take over and reply with correct message.

B. *Suggestions*

Suggestions that can be given for future system development are as follows:

1. Research can be developed by conducting supervised learning of incorrectly classified images to allow the system recognize images better.
2. Reviewing the use of methods that can eliminate the background as such the images in the training and testing process become clearer characteristics
3. The use of a plain background is suggested for better image training process in the extraction process in order to recognize the characteristics of the image

REFERENCES

[1] K. Das, T. Tamhane, B. Vatterott, P. Wibowo and S. Wintels, "The digital archipelago: How online commerce is driving Indonesia's economic development," McKinsey&Company, 2018.

[2] "LINE@," LINE Corporation, [Online]. Available: https://at.line.me/id/. [Accessed 24 November 2018].

[3] M. Kumar, "Digital Image Processing," *Satellite Remote Sensing and GIS Applications in Agricultural Meteorology,* pp. 81-102.

[4] R. Rokhana, J. Priambodo, T. Karlita, I. M. G. Sunarya, E. M. Yuniarno, I. K. E. Purnama and M. H. Purnomo, "Convolutional Neural Network untuk Pendeteksian Patah Tulang Femur pada Citra Ultrasonik B–Mode," *JNTETI,* vol. 8, no. 1, pp. 59-67, 2019.

[5] Y. Hong and J. Kim, "Art Painting Identification using Convolutional Neural Network," *International Journal of Applied Engineering Research,* vol. 12, no. 4, pp. 532-539, 2017.

[6] V. S., "Traffic Sign Recognition Using Convolution Neural Networks," *International Journal of Innovative Research in Computer and Communication Engineering,* vol. 5, no. 6, 2017.

[7] "Techopedia," [Online]. Available: https://www.techopedia.com/definition/16366/chatterbot. [Accessed 15 September 2019].

[8] "TensorFlow," [Online]. Available: https://www.tensorflow.org/. [Accessed 26 November 2018].

[9] "Github - TensorFlow," [Online]. Available: https://github.com/tensorflow/tensorflow. [Accessed 26 November 2018].

[10] NVIDIA Corporation, "NVIDIA DEVELOPER," NVIDIA Corporation, [Online]. Available: https://developer.nvidia.com/discover/convolutional-neural-network. [Accessed 26 November 2018].

[11] "Google Developer," [Online]. Available: https://developers.google.com/machine-learning/practica/image-classification/convolutional-neural-networks. [Accessed 26 November 2018].

[12] "Simple, Big Data Made," [Online]. Available: https://bigdata-madesimple.com/how-do-chatbots-work-an-overview-of-the-architecture-of-a-chatbot/. [Accessed 5 Januari 2019].

Authorized licensed use limited to: San Francisco State Univ. Downloaded on July 02,2021 at 09:37:15 UTC from IEEE Xplore. Restrictions apply.