Preliminary Study of Multi Convolution Neural Network-Based Model To Identify Pills Image Using Classification Rules

By Windra Swastika

Preliminary Study of Multi Convolution Neural Network-Based Model To Identify Pills Image Using Classification Rules

Windra Swastika
Fakultas Sains dan Teknologi
Universitas Ma Chung
Malang, Indonesia
windra.swastika@machung.ac.id

Kestrilia Prilianti
Fakultas Sains dan Teknologi
Universitas Ma Chung
Malang, Indonesia
kestrilia.rega@machung.ac.id

Andrian Stefanus
Fakultas Sains dan Teknologi
Universitas Ma Chung
Malang, Indonesia
311410002@student.machung.ac.id

Hendry Setiawan
Fakultas Sains dan Teknologi
Universitas Ma Chung
Malang, Indonesia
hendry.setiawan@machung.ac.id

***Abstract*— Personal medicine is very important for those who have special health problems. Having several types of pills can make it hard for people to remember every pill especially aged citizen who easily forget his or her own medication. Another problem often encountered is the difficulty of recognizing the drug pills whose labels or the packaging are damaged and hard to read. This research, we developed a multi convolutional neural network (CNN) model to identify pills using classification rules. The idea of using multi CNN model is that almost all type of pills have three main identifiers, namely color, shape and imprint. Three CNNs model are developed to represent each identifier. The number of data collected is 24.000 images, which 95% of the data is used for training purpose and 5% is used for data test. The results of each CNN model is processed with some predefined rules to generate the classes of pills. From the results of different CNN architectures, number of epochs, optimizers and input size experiments, LeNet architecture with input size 64x64 pixels and Adadelta optimization shows the best accuracy up to 99.16%.**

Keywords—convolution neural network, pill identification

I. Introduction

Self-medication is a treatment of health problems to oneself by use of medicines that are bought freely in drug stores for their own initiative without medical advice [1]. The goal of self-medication is to improve health, treat minor illnesses, and routinely treat chronic diseases after doctor's care. Meanwhile, the role of self-medication is to deal quickly and effectively with complaints that do not require medical consultation, reduce the burden of health services on limited resources and energy, and increase the affordability of people far from health services [2].

In many health problems, self-medication requires several type of medicines. Having several types of medicines (drug pills) can make it hard for people to remember every pill especially for aged citizen who easily forget his or her own medication. Another problem often encountered is the difficulty of recognizing the pills whose labels or the packaging are damaged and hard to read. Those problems can lead to mistreatment of health problem.

To minimize the risk of mistreatment caused by the difficulty of recognizing pills, it is necessary to develop a mobile-based application that is able to automatically identify pills obtained from a camera and shows information about the pills. One of the benefit of this application is during a disaster or emergency situation where pill identification is urgently required to safe patient.

A mobile based pill identification application was developed by Lee et al [3] where users only need to take pictures of pills through a smartphone camera. The recognition methods using Hu moment to recognize the shape of the drug, shape invariant feature transform (SIFT) and multi-scale local binary pattern (MBLT) descriptor to recognize the characters contained in the pill and the color histogram to recognize the color of the pill. The accuracy of this application in recognizing pills is 73.17%.

The use of Convolutional Neural Network (CNN) to identify pills was introduced by Zeng et al [4]. They developed a system to identify pills by using the Multi Convolutional Neural Network method consisting of three CNNs. Each CNN has a different task in processing shapes, colors and characters printed on pills such as Color CNN to manage color information, Gray CNN and Gradient CNN to process information on the shape and character of the pill. For the accuracy of the pill identification with only one side using single-CNN was 26.1% while using multi-CNN was 53.1%. The identification using both sides of the pill using single-CNN was 23.1% while using multi-CNN was 74.1%. To identify the pills, the authors use similarity score calculated from cosine distance. Since there are three independent CNNs, similarity score is calculated as the sum of similarity score of color, gray and gradient.

In this research we propose a multi-CNN method based on the three main identifier of a pill namely color (white, yellow, peach and blue), shape dan imprint (character printed on pill). To cover all three identifiers, the developed system will run three CNNs that is responsible for each identifier. The output of CNNs are then fed into a customized classification rule to identify the pill. We also investigate the appropriate CNN architecture and optimizer to obtain the most accurate identification system.

II. Materials and Methods

A. Dataset

The first step carried out in this study was collecting drug information data through online media. There are 8 types of pill are collected consisting of four different types and colors. The drugs used in this study were Amoxicillin, Amoxicillin Trihydrate, Paracetamol, Mixagrip Flu, Decolgen, Spasmal, Sanmol, and Neozep Forte. Each drug information contains information on the name, description, shape, color, type, category, indication, dosage, usage rules, side effects, attention and methods of storing the drug. Fig. 1 shows 8 types of pill used in this study. It also shows four samples of each pill obtained in different distances and angle.

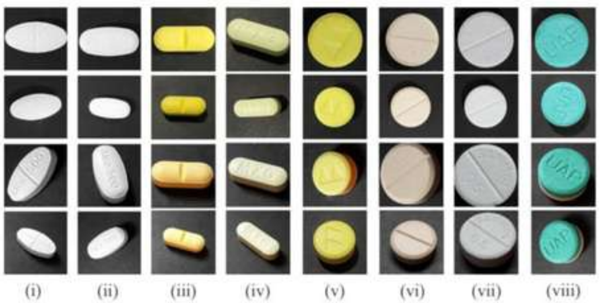

Fig. 1. (i) Amoxicillin, (ii) Amoxicillin Trihydrate, (iii) Paracetamol, (iv) Mixagrip Flu, (v) Decolgen, (vi) Spasmal, (vii) Sanmol, dan (viii) Neozep Forte

Image acquisition is done with a variety of distances, camera quality, lighting, and drug position. In this study we use 13 Mp camera resolution. To simplify the pill localization process in the image, black background colors are used in each image during the acquisition.

Obtained images are distributed for the training process and testing data. Since there are 3 main identifiers to identify pill, we use 3.000 images for each pill as data training for CNN color, 3.000 images for each pill as CNN shape, and 1.600 images for each pill as data training for CNN imprint. Data training for CNN imprint is less than data training for CNN color and CNN shape because only one-side of pill has character on the surface of the pill.

B. Image Preprocessing

The images were pre-processed first before they are trained in the CNNs. The preprocessing is divided into three parts which will be used as input for each CNN, the pre-process includes pre-processes for the shape, color and imprint of the pill image. Tabel 1 shows image preprocessing for each CNN.

TABLE I. Number of Images and Pre-processing for Each CNN

CNN	# *image*	*Pre-processing*
Shape	3.192	Raw Image → Grayscale → Gaussian Blur
Color	3.192	Raw Image → Gaussian Blur
Imprint	1.600	Raw Image → Grayscale → Sharpening

Fig 2 shows an example of a pill that has been preprocessed for each CNN.

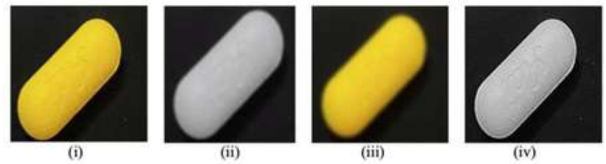

Fig. 2. (i) Original image before pre-processed; (ii) pre-processed image for CNN shape; (iii) pre-processed image for CNN color; and (iv) pre-processed image for CNN imprint

C. Classification Rules

All pre-processed images in the previous step will be resized according to the input size of each CNN architecture. Three CNNs will be used to identify 3 main characteristics of pill, namely color, shape and imprint. The output for CNN color and CNN shape are divided into 4 classes as shown in the Fig 3.

Fig. 3. (i) classification of CNN shape and (ii) classification of CNN color

Each class of CNN shape represent pill as follows:

- Class 0: Amoxicillin (0)
- Class 1: Amoxicillin Trihydrate (1)
- Class 2: Paracetamol (2), and Mixagrip Flu (3)
- Class 3: Decolgen (4), Spasmal (5), Sanmol (6), and Neozep Forte (7)

Each class of CNN color represent pill as follows:

- Class 0 (white): Neozep Forte (7), Sanmol (6), Amoxicillin Trihydrate (1), and Amoxicillin (0)
- Class 1 (yellow): Decolgen (4), Paracetamol (2), and Mixagrip Flu (3)
- Class 2 (peach): Spasmal (5)
- Class 3 (blue): Neozep Forte (7)

The output of each CNN is a percentage of the confidence level in each class. Color classification is divided into 4 classes, namely white, yellow, peach, and blue. Shape classification is divided into 4 classes while the imprint classification is divided into 15 classes as shown in Fig. 4.

Fig. 4. Output for imprint classification

Outputs from CNN shape, CNN color and CNN imprint will be combined based on a classification rules to determine the final output.

The classification rules to identify pill are as follow:

1. Look for classes that have the highest confidence level from the classification results of each CNN shape, color and imprint.
2. Match the output from CNN shape and CNN color. If the output is mismatch, then identification fails. Otherwise, keep all matching pills as potential identification.
3. Match all potential identification from previous step with selected output of CNN imprint.
4. If all pills in the potential identification mismatch with the output of CNN imprint, then identification fails. Otherwise pill successfully identified.

Using the classification rules, there are only two possible output, successfully identified or fail to identify.

D. Evaluation

Confusion matrix is used to evaluate the performance of CNN architecture and optimizer as shown in Table II.

TABLE II. CONFUSION MATRIX USED TO EVALUATE THE PERFORMANCE OF CNN

N		Predicted Class			
		C_1	C_2	...	C_n
Actual class	C_1	TP_1	$E_{C_1C_2}$	$E_{C_1C_{...}}$	$E_{C_1C_n}$
	C_2	$E_{C_2C_1}$	TP_2	$E_{C_2C_{...}}$	$E_{C_2C_n}$
	...	$E_{C_{...}C_1}$	$E_{C_{...}C_2}$	$TP_{...}$	$E_{C_{...}C_n}$
	C_n	$E_{C_nC_1}$	$E_{C_nC_2}$	$E_{C_nC_{...}}$	TP_n

where:

- TP_i: True positive for i^{th} class
- N: the number of data test
- n: number of class
- C_1-C_n: 1^{st}-n^{th} *class*
- $E_{C_iC_j}$: *error recognizing class* C_i *being class* C_j

Some of the metrics that are the focus of the evaluation are True Positive, False Positive, False Negative, True Negative, Accuracy, Sensitivity and Specificity (Table III).

TABLE III. METRICS USED TO MEASURE CNN PERFORMANCE

Metrics	Formula	Evaluation focus
True Positive (TP)	TP_i	Correctly identify i^{th}class (positive class)
False Positive (FP)	$\sum_{j=1}^{i-1} E_{C_jC_i} + \sum_{j=i+1}^{n} E_{C_jC_i}$	Incorrectly identify j^{th} class (negative class) as i^{th} class (positive class)
False Negative (FN)	$\sum_{j=1}^{i-1} E_{C_iC_j} + \sum_{j=i+1}^{n} E_{C_iC_j}$	Incorrectly identify i^{th} class (positive class) as j^{th} class (negative class)
True Negative (TN)	$N\text{-}(TP_i + FP_i + FN_i)$	Correctly identify negative class
Accuracy (acc)	$\frac{\sum_{i=1}^{n} TP_i}{N}$	Correctly identified i^{th}class ratio
Sensitivity (sn)	$\frac{TP_i}{TP_i + FN_i}$	True positive rate of i^{th}class
Specificity (sp)	$\frac{TN_i}{TN_i + FP_i}$	True negative rate of i^{th} class

E. Experimental Parameters

In this study, we use two types of CNN architecture to compare which architectures yield better accuracy. The following are two architecture that will be used in this study:

1. LeNet. LeNet [5] has two convolutional layers followed by max pooling and ends with two fully connected layers. The diagram of LeNet is shown in the Fig. 5.

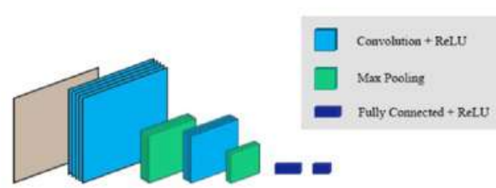

Fig. 5. LeNet Architecture

2. AlexNet. AlexNet [6] has five convolutional layers with an end to three fully connected layers. The diagram of AlexNet architecture is shown in the Fig. 6.

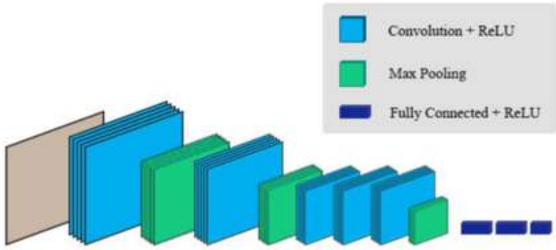

Fig. 6. AlexNet Architecture

Different epochs and image size is also tested to obtain the best parameter. Total images for training process is 24.000 images (each pill has 3.000 images).

III. Results and Discussion

Table IV lists the accuracy of each CNNs with different input of image size using LeNet architecture and Adadelta optimizer.

TABLE IV. IDENTIFICATION ACCURACY USING LENET ARCHITECTURE AND ADADELTA OPTIMIZER

Input Size (pixel)	CNN	Epochs		
		10	*20*	*30*
32x32	CNN Shape	96.13%	99.13%	99.67%
	CNN Color	99.39%	99.85%	99.91%
	CNN Imprint	88.47%	95.29%	98.10%
64x64	CNN Shape	99.24%	99.78%	99.97%
	CNN Color	99.72%	99.73%	99.91%
	CNN Imprint	96.95%	99.78%	99.60%

As shown in Table IV, identification accuracy using LeNet architecture and Adadelta optimizer with 30 epochs reach up to 98% for both 32x32 and 64x64 input size. Input size 64x64 is slightly produce better accuracy compared to input size 32x32.

Table V lists the accuracy of each CNNs with different input of image size using AlexNet architecture and Adadelta optimizer.

TABLE V. IDENTIFICATION ACCURACY USING ALEXNET ARCHITECTURE AND ADADELTA OPTIMIZER

Input Size (pixel)	CNN	Epochs		
		10	*20*	*30*
40x40	CNN Shape	98.73%	99.58%	99.86%
	CNN Color	99.68%	99.59%	99.75%
	CNN Imprint	71.96%	88.65%	94.89%
72x72	CNN Shape	99.77%	99.91%	99.67%
	CNN Color	99.14%	99.78%	99.96%
	CNN Imprint	97.67%	99.30%	99.35%

The results of AlexNet architecture with 30 epochs also shows that the identification accuracy is above 99% (with an exception of CNN imprint with input 40x40 pixel)

From the entire training process, training with the LeNet architecture with input images of 64x64 pixels and AlexNet architecture with 72x72 pixel input images results in a high degree of accuracy. In the training as many as 30 epochs have the optimal level of accuracy throughout the training. These parameters is used for data test.

For the testing purpose, we use 95% (22.800 images) of data as data train and 5% (1.200 images) as data test. Since LeNet architecture outperformed AlexNet architecture, we will use LeNet architecture for testing. Table VI, VII, VIII and IX show our evaluation results of CNN shape, CNN color, CNN imprint class 0-7 and CNN imprint class 8-14, respectively.

TABLE VI. EVALUATION RESULTS OF CNN SHAPE

Evaluation focus	Class			
	0	*1*	*2*	*3*
TP	159	160	318	560
TN	1040	1037	880	640
FP	0	3	0	0
TN	1	0	2	0
Specificity	99%	100%	99%	100%
Sensitifity	100%	99%	100%	100%

TABLE VII. EVALUATION RESULTS OF CNN COLOR

Evaluation focus	Class			
	0	*1*	*2*	*3*
TP	517	444	160	78
TN	682	755	1040	1122
FP	1	0	0	0
TN	0	1	0	0
Specificity	100%	99%	100%	100%
Sensitifity	99%	100%	100%	100%

TABLE VIII. EVALUATION RESULTS OF CNN IMPRINT FOR CLASS 0-7

Evaluation focus	Class							
	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*
TP	80	79	79	76	83	83	78	78
TN	1119	1120	1118	1110	1117	1117	1121	1121
FP	0	0	0	0	0	0	0	0
TN	1	1	3	14	0	0	1	0
Specificity	98%	98%	96%	84%	100%	100%	98%	100%
Sensitifity	100%	100%	100%	100%	100%	100%	100%	99%

TABLE IX. EVALUATION RESULTS OF CNN IMPRINT FOR CLASS 8-14

Evaluation focus	Class						
	8	*9*	*10*	*11*	*12*	*13*	*14*
TP	79	79	78	70	77	80	79
TN	1119	1119	1119	1117	1122	1120	1119
FP	2	1	3	13	1	0	1
TN	0	1	0	0	0	0	1
Specificity	100%	98%	100%	100%	100%	100%	98%
Sensitifity	99%	99%	99%	98%	99%	100%	99%

As shown in Table VI-IX, the average specificy of each class is above 98%. It means that the network is able to identify each class with error less than 2%. Likewise, the average value of the specificity of each class is above 99%, which means that it can distinguish one class from another class. Overall performance of identification accuracy for LeNet and AlexNet architecture is shown in Table X.

TABLE X. IDENTIFICATION ACCURACY USING 1.200 DATA TEST

Architecture	CNN			Classification rule
	Shape	*Color*	*Imprint*	
LeNet	99.83%	99.50%	99.08%	99.16%
AlexNet	100%	99.66%	97.75%	98.75%

Using 1.200 images, the network using LeNet architecture is able to correctly identify 1.198 images for CNN shape, 1.194 images for CNN color and 1.189 images for CNN imprint. The final identification performance after implementing classification rule is able to correctly identify 1.190 images.

For AlexNet architecture with 72x72 pixels, after implementing classification rule, the network is able to correctly identify 1.185 images and fails to identify 15 images. LeNet architecture shows better performace accuracy compared to AlexNet architecture although the input size of LeNet is 64x64 pixels. One of the contributing factors is the performance of CNN imprint is lower compared to the performance of the CNN imprint of the Lenet architecture.

IV. CONCLUSION

We have developed a multi CNN based model to identify pill image. There are three main identifiers of a pill that needs to be identified, ie shape, color and imprint. Three CNNs model are developed based on those three main identifier and a classification rule to combine the results of CNNs model are applied.

Using 1.200 data test, The highest identification accuracy 99.16% is achieved by LeNet architecture with 30 epochs, image size 64x64 pixels and uses the Adadelta optimization. The results of specificity and sensitivity in the evaluation also demonstrated the ability of the network to distinguish between the classes above 98%.

For the future work, we will apply this model in an android-based applications to identify pill image obtained from a camera.

REFERENCES

[1] Tan, H. T., and Kirana Rahardja. "Obat-obat sederhana untuk gangguan sehari-hari." *Jakarta: PT Elex Media Komputindo*(2010).

[2] Supardi, Sudibyo, et al. "Kajian Peraturan Perundang-undangan Tentang Pemberian Informasi Obat Dan Obat Tradisional Di Indonesia." Indonesian Pharmaceutical Journal 2.1: 20-27.

[3] Lee, Young-Beom, et al. "Pill-ID: Matching and retrieval of drug pill images." Pattern Recognition Letters 33.7 (2012): 904-910.

[4] Zeng, Xiao, Kai Cao, and Mi Zhang. "MobileDeepPill: A small-footprint mobile deep learning system for recognizing unconstrained pill images." Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services. ACM, 2017.

[5] LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324.

[6] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

Preliminary Study of Multi Convolution Neural Network-Based Model To Identify Pills Image Using Classification Rules

ORIGINALITY REPORT

5%

SIMILARITY INDEX

PRIMARY SOURCES

1	"Rough Set Theory: A True Landmark in Data Analysis", Springer Science and Business Media LLC, 2009 Crossref	19 words — 1%
2	Windra Swastika, Martien Febriant Ariyanto, Hendry Setiawan, Paulus Lucky Tirma Irawan. "Appropriate CNN Architecture and Optimizer for Vehicle Type Classification System on the Toll Road", Journal of Physics: Conference Series, 2019 Crossref	17 words — 1%
3	Xiao Zeng, Kai Cao, Mi Zhang. " ", Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services - MobiSys '17, 2017 Crossref	12 words — < 1%
4	polen.itu.edu.tr Internet	12 words — < 1%
5	Pejaver, Vikas, Wei-Lun Hsu, Fuxiao Xin, A. Keith Dunker, Vladimir N. Uversky, and Predrag Radivojac. "The structural and functional signatures of proteins that undergo multiple events of post-translational modification : Structural and Functional Signatures of PTM Crosstalk", Protein Science, 2014. Crossref	11 words — < 1%
6	pure.uva.nl Internet	10 words — < 1%
	"Engineering Applications of Neural Networks", Springer Science	

7	and Business Media LLC, 2011 Crossref	10 words — < 1%
8	de.scribd.com Internet	10 words — < 1%
9	Kestrilia R. Prilianti, Samuel P. Yuwono, Marcelinus A. S. Adhiwibawa, Monika N. P. Prihastyanti et al. "Automatic leaf color level determination for need based fertilizer using fuzzy logic on mobile application: A model for soybean leaves", 2014 6th International Conference on Information Technology and Electrical Engineering (ICITEE), 2014 Crossref	10 words — < 1%
10	Windra Swastika, Yoko, Hendry Setiawan, Moch. Subianto. "Android based application for recognition of Indonesian restaurant menus using convolution neural network", 2017 International Conference on Sustainable Information Engineering and Technology (SIET), 2017 Crossref	10 words — < 1%
11	arxiv.org Internet	9 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE MATCHES	OFF
EXCLUDE BIBLIOGRAPHY	ON		