The Identification of Pill Images Using Convolutional Neural Network (CNN)

Suwat Tangwattananuwat
Business Analytics and Data Science (BADS) , Graduate School of Applied Statistics
National Institute of Development Administration , Bangkok , Thailand.
suwat.tang@stu.nida.ac.th

***Abstract*—Currently, there are many types of tablets that are commonly used in hospitals and pharmacies in Thailand, each with different characteristics. The problems that are often encountered Is the discrepancy in dispensing of pharmacists That is at risk causing danger. Therefore, in order to provide those customers the service to be confident and safe in the treatment. In this paper, studies and techniques are used. Convolutional Neural Network and Transfer learning In order to test how the technique can classify tablets by using Transfer learning 3 models VGG16, Inception-Resnet-V2 and Xception are used as a training model and constructing a Convolutional Neural Network to connect at the end of each Transfer learning model. There are 28 types of tablets and capsules, divided into 2,760 images train data sets and 2 test data sets. The first test data set contains 230 images. The second test data set contains 84 images. Test data set 2 is an image taken from actual use. All 3 models gave 100% accuracy of the test data set in test 1, all 3 models and 71.42%, 82.14% and 77.38% accuracy of the test data set 2. Respectively.**

Keywords*—Convolutional neural network , Pill Identification, Data augmentation, Transfer learning

I. INTRODUCTION

Currently, most tablets are manufactured using chemicals that are used to treat or prevent disease. The tablets are small and hard. There are over ten thousand types of commonly used tablets in Thailand. Therefore, it is difficult for humans to remember all types of drugs. Which, if there is a wrong dispensing of medicine or taking the wrong medicine The situation is getting worse for patients who are allergic to drugs[1].

Therefore, in order to prevent errors in this paper, we will manage the risk by Image Recognition, which is a deep learning process with the Convolution Neuron Network (CNN) Algorithm. The Feature Extraction we use will Divided into shapes, color, and stamps on Imprint [2] by allowing the patient to take pictures of the drug and through the model that we have created to recognize the tablet. Which helps patients, seniors or the general public to use in tablet identification for more accuracy and accuracy. In each type of medication.

II. LITERATURE REVIEW AND THEORY

In the study of research on The Identification of Pills Image Using Convolutional Neural Network (CNN) and Transfer learning. The researcher has studied research theory and related basic knowledge as follows

A. Convolution Neural Network : CNN

CNN is a multilayer neural network that has a unique structure. It is designed to add the feature to extract very complex features from data. CNN is often used to extract features from unstructured data such as images.

CNN will simulate the human vision of the area. And bringing together groups of sub-areas together, looking at the sub-areas of humans, there will be separate feature of that sub-area, such as lines and contrasting colors In which humans know that this area is a straight line or a contrasting color because humans look at both the points of interest and the surrounding area together. There are three types of layers in a convolutional neural network: convolutional layer, pooling layer, and fully connected layer. Each of these layers has different parameters that can be optimized and performs a different task on the input data[3].

- *Convolution Layer*

The Convolutional layer is the first layer that has the process of extracting important features from images. Within Convolution, there will be a Feature Extraction in the Hidden layer. This calculation starts with the configuration in the filter or kernel that helps to retrieve the features used. In recognizing objects Normally, a filter / kernel can pull out one feature of interest. We therefore need to have many filters as well. In order to find various spatial features combined and filters move around of input and the final result of this layers is called feature map as fig.1

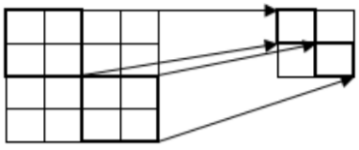


Fig. 1 show Convolutional with input size 4x4 and kernel size 2x2

- *Pooling Layer*

Pooling layers are similar to convolutional layers, but they perform a specific function such as max pooling, which takes the maximum value in a certain filter region, or average pooling, which takes the average value in a filter region. These are typically used to reduce the dimensionality of the network. which in this paper used max pooling as fig.2

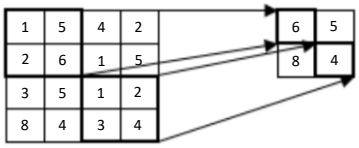


Fig.2 show max pooling filter is (2x2) and stride is 2

- *Fully Connected Layer*

Fully connected layers are placed before the classification output of a CNN and are used to flatten the results before classification.This is similar to the output layer of an MLP as fig.3

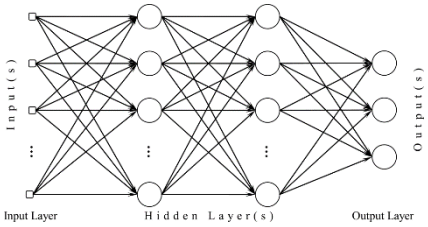


Fig.3 show input , hidden and output layers of fully connected layer.

B. Batch Normalization

BatchNorm is a technique used during Machine Learning to shift, scale and enable the Deep Neural Network's Activation within the Hidden Layer to not be too small, not too big. By comparing the Mean and Standard Deviation of every Activation in the entire Batch layer, it is similar to the Feature Scaling of the Input and is added with Learning Parameter so that the learning model can adjust the Activation as needed .which BatchNorm is considered a regularization method[4].

C. Activation

Activation Function use for modification effective of Convolutional Neural Network (CNN) and get value feature map from convolution layer. In this article, we use 2 types of activation, ReLu and Softmax

- *Rectified Linear Units (ReLU)*

$$f(x) = \max(0, x) = \begin{cases} 0 \ for \ x \geq 0 \\ x \ for \ x > 0 \end{cases} \qquad (1)$$

A Rectified Linear Unit (A unit employing the rectifier is also called a rectified linear unit ReLU) has output 0 if the input is less than 0, and raw output otherwise. That is, if the input is greater than 0, the output is equal to the input .[5]

- *Softmax*

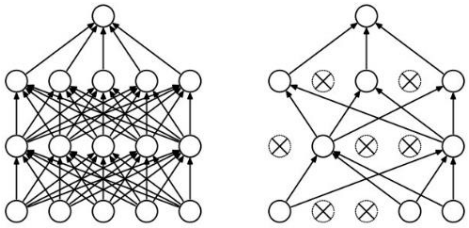
Softmax function is a function that accepts Input as vector of real number then normalize is probability that is equal to 1. and softmax is taken. Frequently used in work. Classification

$$softmax(x_i) = \frac{e^{x_i}}{\sum_j e^j} \qquad (2)$$

When xi as a value of output then exponential and divided by sum of exponential so got softmax that can represent possible what is a class of dataset.

D. Dropout

The idea of Dropout is that we will close some Neuron in the system to not receive any data. Therefore, the closed Neuron will not send the information to other Neuron as fig. 4



(a) (b)

(a) Show normal neural network
(b) Show neural networks that have dropouts.

Fig. 1 Comparison of dropouts in neural networks

Closing some Neuron networks in each layer It is a random shutdown, which sounds superficially, may seem like a bad idea. But in truth, this technique works well. Because we didn't go to close Input x but to turn off some processors Selecting this random shutdown will prevent occurrences. Co-adaptation and is effective in reducing overfitting because it makes the model easy down but does not intentionally press down on any one type of information Is a random click to reduce the complexity of the entire system [6] .

E. Optimization Algorithm

Optimization is an algorithm that improves errors and losses. This will change the weight and bias associated with neurons. which in paper we used Adam algorithm.

- Adam

Adaptive Moment Estimation or Adam is an optimizer that can adjust the learning rates for each parameter and can also solve the decaying of

gradients, and Adam is a popular algorithm in the field of deep learning because it achieves good results fast.

F. Early Stopping

Early stopping is to stop training before the Optimize Converge finds the lowest loss with the assumption that the more training The model becomes more complicated. To the point that it might be too good for you to generalize[6].

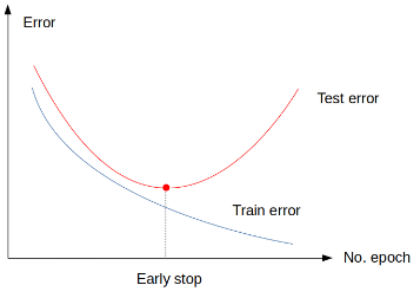


Fig. 2 Show early stopping graphs

You'll see that Optimize normally trains to get the lowest Train error, but sometimes the Test error starts to return to a higher value before finding the lowest point of the Cost function. Therefore, we should stop training at that point. Lowest test error, not the lowest Train error.

G. Transfer Learning (Pretrained model)

Deep convolutional neural network models may take days or even weeks to train on very large datasets.

A way to short-cut this process is to re-use the model weights from pre-trained models that were developed for standard computer vision benchmark datasets, such as the ImageNet image recognition tasks. Top performing models can be downloaded and used directly, or integrated into a new model for your own computer vision problems.[7]

- *VGG16 (2014)*

The folks at Visual Geometry Group (VGG) invented the VGG-16 which has 13 convolutional and 3 fully-connected layers, carrying with them the ReLU tradition from AlexNet. This network stacks more layers onto AlexNet, and use smaller size filters (2×2 and 3×3). It consists of 138M parameters and takes up about 500MB of storage space as fig 3.[8]

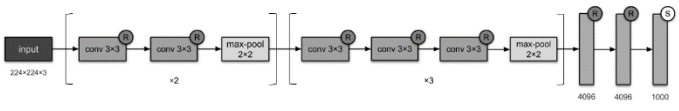


Fig. 3 VGG-16 architecture

- *Inception-Resnet-V2 (2016)*

A family of Inception-ResNet-v1 and Inception-ResNet-v2. The latter member of the family has 56M parameters. and improved from the previous version Inception-v3 . Converting Inception modules to Residual Inception blocks , Adding more Inception modules and Adding a new type of Inception module (Inception-A) after the Stem module as fig 4.[9]

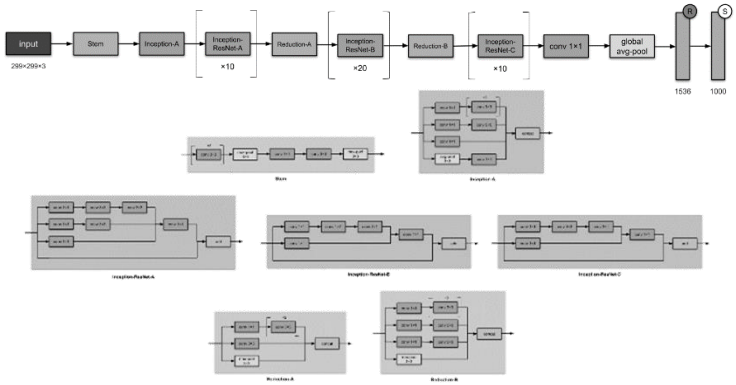


Fig. 4 Inception-ResNet-V2 architecture

- *Xception (2016)*

Xception is an adaptation from Inception, where the Inception modules have been replaced with depthwise separable convolutions. It has also roughly the same number of parameters as Inception-v1 (23M).Firstly, cross-channel (or cross-feature map) correlations are captured by 1×1 convolutions.
Consequently, spatial correlations within each channel are captured via the regular 3×3 or 5×5 convolutions.[10]

Taking this idea to an extreme means performing 1×1 to every channel, then performing a 3×3 to each output. This is identical to replacing the Inception module with depthwise separable convolutions as fig 5.

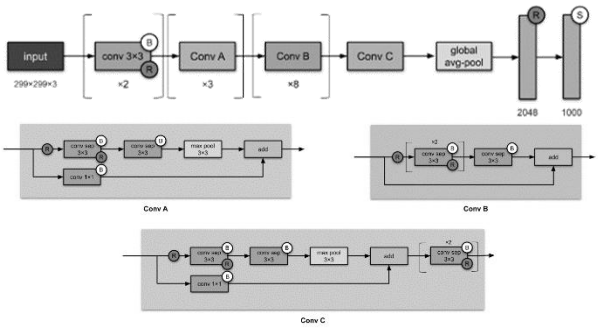


Fig. 5 Xception architecture

III. METHODOLOGY

We have developed an image identification system, with the algorithm consisting of 3 steps. Step 1: We will collect data by taking pictures of 28 tablets and capsules with Iphone 6S plus. The photos will be square shape which has size 1280 x 1280 pixel with the total number of images. 3,074 photos .Step 2 We will preprocess by creating the bounding box and crop the image according to the bounding box. After that, reduce the images size to 224 x 224 pixel and Data Augmentation because the image dataset we have is the size of The tablet is not yet suitable for importing the model, so we have to preprocess it. Step 3 We will add the Pretrained model to connect with the Convolution Neural Network (CNN) that we have created to suit the prediction of 28 types of drugs.

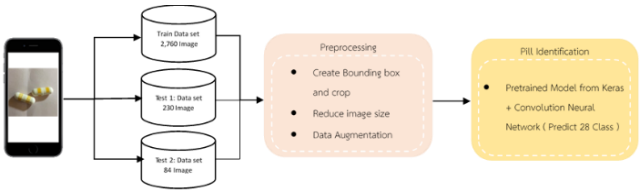


Fig. 6 Overview of pill identification

A. Dataset

In this paper, both the training data and the test data, we used 3,074 image data, including tablets and capsules, with different characteristics.

Trade Name	Generic Name or Ingredient	Strength	Dosage Form
Air-x	Simethicone	80 mg	Tablet
Amlopine	Amlodipine	5 mg	Tablet
Betahistine12	Betahistine Mesilate	12 mg	Tablet
Biocalm	Tolprisone HCL	50 mg	Tablet
Caltab - 600	Calcium Carbonate	600 mg	Tablet
Celebrex	Celecoxib	200 mg	Capsule
Cetirizine	Cetirizine Hydrochloride	10 mg	Tablet
Cinnarizine	Cinnarizine	25 mg	Tablet
Cyproheptadine	Cyproheptadine Hydrochloride	4 mg	Tablet
Chinclonac	Diclofenac Sodium	50 mg	Tablet
Diclofenac25	Diclofenac Sodium	25 mg	Tablet
Dimenhydrinate	Diphenhydramine Theoclate	50 mg	Tablet
Domperidone	Domperidone	10 mg	Tablet
Duran	Ibuprofen	400 mg	Tablet
Dymoxin	Amoxicillin Trihydrate	500 mg	Capsule
Ibrofen	Ibuprofen	600 mg	Tablet
Losartan	losartan potassium	50 mg	Tablet
Merislon	Betahistine Mesilate	6 mg	Tablet
Methycobal	Mecobalamin	0.5 mg	Tablet
Metroclopramide	Metroclopramide HCL	10 mg	Tablet
Miracid	Omepazole	20 mg	Capsule
Nasotapp	Phenylephrine,brompheniramine	14 mg	Tablet
Norgesic	orphenadrine+paracetamol	485 mg	Tablet
Nuosic	Orphenadrine,paracetamol	535 mg	Tablet
P-fen	Ibuprofen	400 mg	Tablet
Pencilin	penicillin V	250 mg	Tablet
Tofago	ergotamine tartrate,caffeine	101 mg	Tablet
Torrent	Orphenadrine,paracetamol	485 mg	Tablet

Table I . Detail of medicine for dataset

B. Preprocessing

In preprocessing, we will divide it into 2 steps which are Create Bounding Box and crop and Data Augmentation

- *Create Bounding Box and crop*

We will pill the localization in the middle of the image by creating a Bounding Box to crop the selected area in the image with just a tablet. The input image initially has a size of 1280 * 1280 pixel, after Crop is 531 * 531 then we have to Reduce images size to 224 * 224 pixel to suit the Input size of our Model.

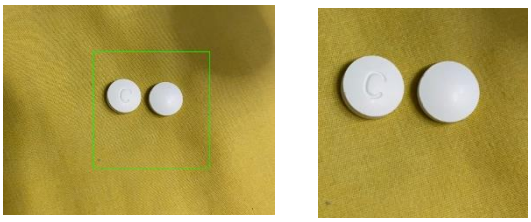


Fig.7 Show Bounding box crop and reduce Images size

- *Data Augmentation*

Data augmentation is a strategy that enables practitioners to significantly increase the diversity of data available for training models, without actually collecting new data [11]. Data augmentation techniques in used such as rotation, brightness, vertical and horizontal flipping are commonly used to train large neural networks

- *Shear Intensity*

Shear is sometimes also referred to as transvection. A transvection is a function that shifts every point with constant distance in a basis direction(x or y). It slants the shape of the image. Here, we fix one axis and stretch the certain angle known as the shear angle. It stretches the image which is different than the rotation technique. we specify the shear_range in the degrees .which in this paper used shear range is 0.2[12] .

- *Zoom*

We use zoom_range argument to specify the values. If zoom_range is less than 1.0 then it magnifies the image and zoom_range greater than 1.0 zooms out

of the image. which in this paper used zoom range 0.8 to 1.2 [12] .

- *Rotation*

We can specify the angle in degrees and this then apply it to a large dataset we can use the rotation_range parameter to specify the range of values which then generates the images in the range of +rotation_range to -rotation_range(in degrees) which inthis paper used rotation range is 90 degrees[12].

- *Brightness*

We can apply the brightness_range technique for randomly picking a brightness shift value from and we can specify a floating-point number between 0.0 to 1.0 which tell us that 0.0 means no brightness & 1.0 corresponds to maximum brightness. Which in this paper used brightness range is 0.2 to 1.0[12] .

- *Horizontal flip*

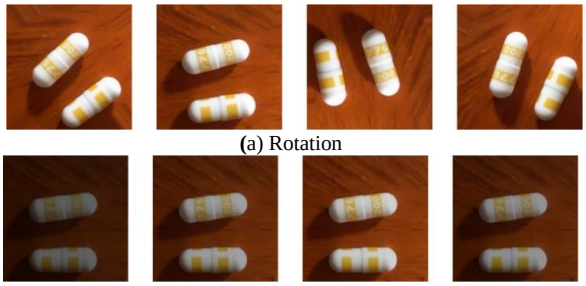
It flips the images horizontally by specifying the boolean value in the horizontal_flip parameter. By specifying true it flips them horizontally[12].

- *Vertical flip*

It flips the images vertically by specifying the boolean value in the vertical_flip parameter. By specifying true it flips them vertically[12].

- *feature standardization*

It is also possible to standardize pixel values across the entire dataset. This is called feature standardization You can perform feature standardization by setting the featurewise_center and featurewise_std_normalization arguments on the ImageDataGenerator class. These are in fact set to True by default and creating an instance of ImageDataGenerator with no arguments will have the same effect[13].



(a) Rotation

(b) Brightness

Fig. 8 Show examples of Data Augmentation

C . Pill Identification

After we perform Image Preprocessing Finished, we will split the trainning dataset into 80% train and 20% validation. The test data is divided into 2 test data sets. Previously. then we will import to the model. The models that we used in Classify are 3 models: Transfer learning of keras with VGG16, Inception-Resnet-V2 and Xception. Within all 3 models, we have cut The back of the model is the Fully Connected section and the Classification section. And add the model we created ourselves to the end, which the model we created will adding layer into tranfer learning 3 models .model have global average pooling, batch normalization with ReLu activation , dense layer, dropout and will adjust the section Classification (Final Layer) softmax activation . Only 28 Classes are needed to suit our predictions. The number of Train cycles is 40 epochs. Early stopping is used to save the model. The monitors are at Val_loss.

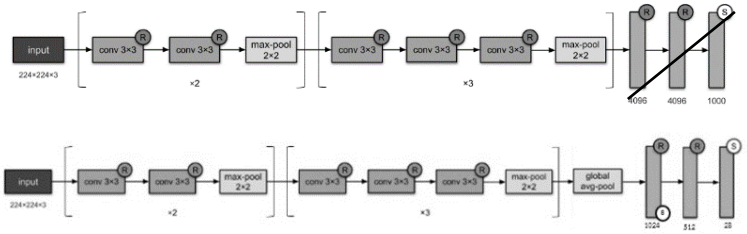


Fig.9 Show model for improvements at the end

IV. RESULT

After input training and testing dataset into 3 models a final result for test dataset 1 have correct for prediction total 230 of 230 pictures is equal to 100% all 3 models and for test dataset 2 totally 84 pictures VGG16 model give accuracy 60 pictures amount to 71.42% and mistake 24 pictures amount to 28.58%. Inception-Resnet-V2 model give accuracy 69 pictures amount to 82.14% and mistake 15 pictures amount to 17.86% and Xception model give accuracy 65 pictures amount to 77.38% and mistake 19 pictures amount to 22.82%. a sample pill images as a correct and incorrectly result of prediction pill images with 3 model represent by table 2.

	Test dataset 1		Test dataset 2	
Model	correct	incorrect	correct	incorrect
VGG16	100 %	0 %	71.42 %	28.58 %
Inception-Resnet-V2	100 %	0 %	82.14 %	17.86 %
Xception	100 %	0 %	77.38 %	22.62 %

Table. 2 Result of test dataset 1 and test dataset 2

V. CONCLUSION AND FUTURE WORK

In this paper , we presented the design , implementation , Image processing followed by a neural network was used to identify pill images. Based on a transfer learning and CNNs architecture . VGG16, Inception-Resnet-V2 and Xception are used

as a training model After our algorithm was evaluated the results showed that the accuracy of pill identification was about 71.42%, 82.14% and 77.38 accuracy of the test data set 2. Respectively. To increase the accuracy, we will develop image processing Allow the bounding box to move along the tablet's position for the model to better classify the tablets.

References

[1] PRARIN CHUPAWA and KIATTISIN KANJANAWANISHKUL.(2015). Pill Identification with Imprints Using a Neural Network. Mahasarakham International ,1(6),30-35.

[2] Xiao Zeng, Kai Cao, Mi Zhang .(2017). MobileDeepPill: A Small-Footprint Mobile Deep Learning System for Recognizing Unconstrained Pill Images. Conference: the 15th Annual International Conference(pp.56-67)

[3] Matthew Stewart, PhD Researcher.(Feb 27, 2019). Simple Introduction to Convolutional Neural Networks(Online).from https://towardsdatascience .com/simple-introduction-to-convolutional-neural-networks-cdf8d3077bac

[4] Keng Surapong ,BUA Labs .(2019). Deep Convolutional Neural Network – ConvNet ep.5(Online) . From https://www.bualabs. com/ archives/2617/what-is-batchnorm-teach-batch-normalization-train-machine-learning-model-deep-convolutional-neural-network-convnet-ep-5/

[5] Himanshu Sharma.(Jan 19, 2019) .Activation Functions : Sigmoid, ReLU, Leaky ReLU and Softmax basics for Neural Networks and Deep Learning(2019)(Online) .From https://medium. com/@himanshuxd/activation-functions-sigmoid-relu-leaky-relu-and-softmax-basics-for-neural-networks-and-deep-8d9c70eed91e

[6] Chitpong Kittinaradon (jan,2020) .Neural Network Regularisation(Online).From https://guopai.github.io/ml-blog18.html

[7] Raimi Karim(Jul 29, 2019) .Illustrated: 10 CNN Architectures(Online).From https://towardsdat ascience.com/illustrated-10-cnn-architectures-95d78ace614d#643c

[8] Karen Simonyan, Andrew Zisserman, University of Oxford(2014). UK. VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION, arXiv preprint arXiv:1409.1556

[9] François Chollet ,Google Inc(2017). Xception: Deep Learning with Depthwise Separable Convolutions . IEEE Conference on Computer Vision and Pattern Recognition (CVPR),(pp.1800-1807)

[10] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, Alex Alemi. Google.(2016) . Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence

[11] Daniel Ho, Eric Liang, Richard Liaw (Jun 7, 2019) . 1000x Faster Data Augmentation(Online) .From https://bair.berkeley.edu/blog/2019/06/07 /data_aug/

[12] Paras Patidar(Jan 6, 2020).Image Data Augmentation- Image Processing In TensorFlow-Part 2(Online). From https://medium. com/mlait/image-data-augmentation-image-processing-in-tensorflow-part-2-b77237256df0

[13] Jason Brownlee(September 13, 2019) . Image Augmentation for Deep Learning With Keras(Online) .From https:// machinelearning mastery.com/image-augmentation-deep-learning-keras/