Uluslararası İleri Doğa Bilimleri ve Mühendislik Araştırmaları Dergisi
Sayı 8, S. 243-253, 4, 2024

Araştırma Makalesi

https://as-proceeding.com/index.php/ijanser
ISSN: 2980-0811

International Journal of Advanced Natural Sciences and Engineering Researches
Volume 8, pp. 243-253, 4, 2024

Research Article

The Pharmaceutical Drug Classification using Deep Learning Approaches

Ronak Miteshkumar Patel*, Sanketkumar Dineshbhai Vaghani[2], Thangarajah Akilan[3] and Saad Bin

Faculty of Science and Environmental Studies - Computer Science Department, Lakehead University, Thunder Bay, Canada.

* (*rpate102@lakeheadu.ca)*

(Received:22 May 2024, Accepted: 25 May 2024)

(3rd International Conference on Engineering, Natural and Social Sciences ICENSOS 2024, May 16-17, 2024)

ATIF/REFERENCE: Patel, R. M., Vaghani, S. D., Akilan, T. & Bin, S. (2024). The Pharmaceutical Drug Classification using Deep Learning Approaches. *International Journal of Advanced Natural Sciences and Engineering Researches*, 8(4), 243-253.

Abstract – Accurate drug classification through deep learning approaches enhance medication safety by minimizing errors in drug identification and dosage, ultimately safeguarding patient health. These advanced techniques provide a promising solution to prevent the risks associated with incorrect medication use, ensuring that patients receive the most effective treatment for their condition. Patients may face serious consequences due to improper medication use, including errors in taking the wrong drug or incorrect dosage. In order to mitigate the risk of human error in identifying medications, we employed advanced deep learning models such as Convolutional Neural Networks (CNN), VGG19, and Inception-ResNetV2. These models were trained using a comprehensive dataset comprising over 7000 labeled drug images. Through our study, we achieved a remarkable validation accuracy of 95\% utilizing the CNN model. This demonstrates the potential effectiveness of employing deep learning techniques in accurately classifying drug images, thereby reducing the likelihood of medication errors and improving patient safety.

Keywords – CNN, Inception ResNetV2, VGG19, Deep Learning, Drug Classification

I. INTRODUCTION

Advances in machine learning techniques have changed a number of industries, including healthcare, in recent years[1]. Medical image classification is one such field where machine learning algorithms may help automate the recognition and categorization of medical images, resulting in an efficient and more precise diagnosis. This paper focuses on using machine learning (ML) to categorize drug images, which is an important problem in the pharmaceutical and healthcare industries. In healthcare, pill identification and classification are essential for a number of reasons, such as drug adherence, dosage monitoring, and medication mistake prevention. Conventional techniques depend on manual examination, which is laborious and prone to mistakes. A potential way to accelerate the process is by deep learning classification.

Accurate medication dosage is crucial for patient safety as it ensures the effectiveness of treatment while minimizing the risk of adverse effects. Wrong medications or incorrect dosages can lead to a range of harmful consequences including treatment failure, exacerbation of symptoms, allergic reactions, drug interactions, and even life-threatening complications. Patients may experience delayed recovery, prolonged illness, or irreversible damage to their health. Therefore, precise dosage administration is essential to optimize therapeutic outcomes and safeguard patient well-being. Patients often find it challenging to differentiate between various medications, increasing the risk of self-harm through drug misuse. This difficulty intensifies when pills are relocated to a different containers, mixed up, or stored in communal pill organizers for convenience. Elderly individuals, in particular, are more susceptible to mistakenly identifying medicines, heightening the likelihood of adverse drug events. One solution for this problem can be traditional manual human inspection. Performing a manual search can be laborious, tedious, and time-consuming, especially when handling numerous medications with numerous generic variants. Furthermore, human error is often introduced while reading microscopic imprints on tiny medications. There is little to no room for error when it comes to the medication usage.

Deep learning is a subset of artificial intelligence that mimics the human brain's neural networks to process large volumes of complex data and extract meaningful patterns. Through sophisticated algorithms, deep learning enables machines to progressively improve their performance on tasks such as image and speech recognition, natural language processing, and decision-making. The field of Computer Vision has made significant progress in object detection recently, primarily due to the rapid advancements in Deep Learning, particularly with Convolutional Neural Networks (CNNs). In [2] the authors have conducted several visual classification experiments and concluded that deep learning with CNN should be regarded as the primary candidate in essentially any visual recognition task. With this ability of Deep learning, it can automate pill recognition task and can assist in quickly identifying pills, reduce the likelihood of incorrect pill identification, and give the patient visual assurance.

In this paper, we have trained three Deep learning models, which are CNN, VGG19 and inception-resnetv2. This study is focused on training our models on image data, to accurately classify them into respective classes. We performed evaluations by using the above-mentioned deep learning techniques. The proposed solution discussed which model obtain better performance by considering the complexity of an image.

II. Literature Review

This section is providing review of most well-known models such as CNN, GoogLeNet, fully convolutional network (FCN), ResNet50, MobileNet, SqueezeNet, InceptionV3 and techniques for pill image identification in this section.

In [3], the work was presented to categorized pills using Region of Interest (ROI). They initially identifying the pill in the image by detecting the region with the highest concentration of edges. Three GoogLeNet models specializing in color, shape, and features were trained on an expanded dataset. The method was evaluated using a publicly available dataset from the National Institute of Health, achieving an impressive ROI extraction accuracy of 99.5\%.

A Drug pills detection system was proposed by [4], employing Convolutional Neural Network(CNN) for automated identification. The approach comprised two stages: localization and classification. For localization, a ResNet-based Backbone and Feature Pyramid Network (FPN) were utilized, with FPN manipulates CNN feature hierarchies to construct a feature pyramid merging semantics from high to low levels.The ResNet 50 is used for representation of features map. Regression and classification sub-models are used for bounding boxes estimation. In the second stage, a convolution network model is chosen to predict pill categories.

Another important work is proposed by [5], they utilized a pill localization approach involving a blob-detection neural network and morphological post-processing. They trained a fully convolutional network (FCN) fro the blob detector, employing it as a pixel-segmentation algorithm. This prescription-pill identification method based on FCN enables the removal of background textures from pill images using the predicted segmentation masks. Authors implemented and compared models like ResNet50, MobileNet, SqueezeNet, InceptionV3 on NIH NLM Pill Image Recognition Challenge dataset. Similarly [6], introduced an automatic classification system for pill images focusing on their shape and colour attributes. Employing image processing techniques, they defined an attribute set utilized by Support Vector Machines and Multi-layer Perceptron classifiers. The experiment was conducted on a subset of the NLM PIR dataset from National Library of Medicine. The results indicate that all classifiers perform with an average accuracy above 99.3\%. This high classification accuracy happens even in the presence of unbalanced classes, with precision and recall average scores above 98\%.

Another study[7] focuses on enhancing the drug image retrieval accuracy and efficiency through image segmentation and classification. Their study introduces three neural network (CNN) architectures: two hybrid networks paired with classification methods(CNN+SVM and CNN+kNN), and one ResNet-50 network. Utilizing the National Library of Medicine (NLM) database, the results indicate that their proposed model achieves an accuracy of 90.8\%. For the training of models, attributes like imprints on the pills, shape and colour of the pills are used.

The deep learning allows for the development of automated systems that can rapidly process large numbers of pill images, providing quick and reliable identification, which is especially valuable in settings where manual identification may be time-consuming or prone to error. As learned from the recent contribution, deep learning revolutionizing pill identification by providing efficient, accurate, and scalable solutions to help improve medication safety and healthcare services.

III. METHODOLOGY:

Pharmaceutical drugs encounter various difficulties, including classification into multiple categories, their relatively small sizes, and their similarity in appearance. For instance, within our dataset, ten distinct categories of pharmaceutical drugs exist, with some exhibiting remarkably similar appearances, potentially leading to human errors in identification. To address this issue, we implemented Convolutional Neural Networks (CNNs), including variants like VGG19 and InceptionResNetV2, to recognize images accurately. The primary goal of our proposed model is to comprehend and implement all three models for image recognition. These models have demonstrated significant effectiveness and potential across various tasks in image processing, including handwriting recognition, object detection, and segmentation, within the realm of computer vision[8].

A. *Dataset:*

For our research, we utilized pharmaceutical pill image data which is readily accessible [17]. The access link to this data is provided at the bottom of the page. The dataset comprises over 10,000 images. Of these, 7,000 labeled images were employed for training, while 3,000 images were reserved for testing unlabeled images. The dataset encompasses a total of 10 distinct classes of pill images. As part of our pre-processing steps, all images were resized to 300x300 pixels and normalized pixel intensities were adjusted to a range of 0 to 255 for RGB values. Sample images from the dataset are illustrated in Figure 1. Following this step, data is utilized to train our chosen models. Figure 2 presents the data flow of our methodology.

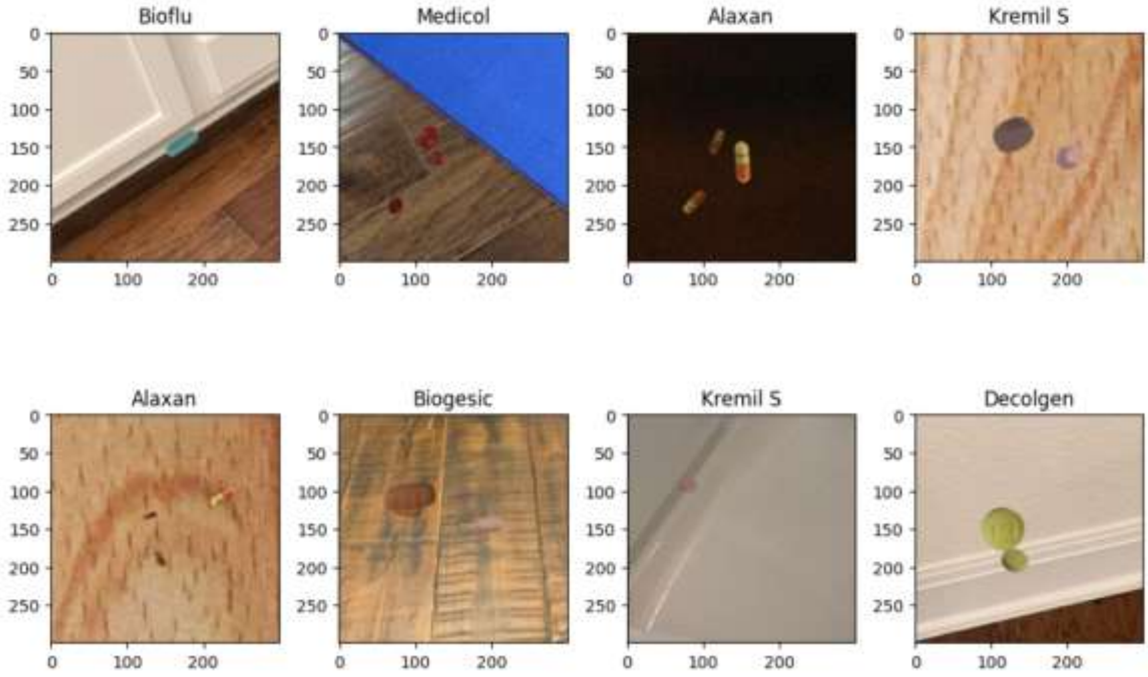

Fig. 1. Pharmaceutical Drug Images

B. *Deep Learning Models:*

1. CNN Architecture:

A convolutional Neural network [9] [10] is made up of different layers that transform one set of information into another using a special function. To create the network, a total of five types of layers are used: Convolutional, Pooling, Dropout, Flatten, and Fully Connected. All these layers are arranged to create a total of 17 layers for our Convo- lutional Neural Network Model.

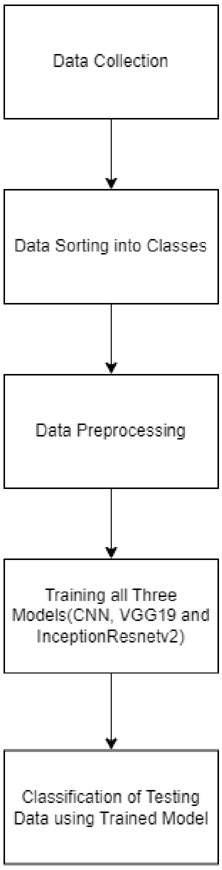

Fig. 2. Methodology

The Convolutional layer is the ReLu(Rectified Linear Unit) activation function with the image input size of n*n = 300*300 and kernel size f*f = 3*3 with some layers having 32 filters and some layers having 64 filters throughout the whole model. This extracts features from the input image and generates 32 and 64 respectively feature maps as output, which serve as input to the subsequent layers in the CNN architecture. The MaxPooling2D layer is a downsam- pling operation used in convolutional neural networks (CNNs) to reduce the input feature maps' spatial dimensions (width and height) while retaining the most important information. The pooling size of every layer is 2*2. The Dropout layer is a form of regularization that randomly deactivates (drops out) a fraction of neurons during training. This means that for each update of the model parameters (weights and biases), a random subset of neurons is temporarily removed from the network, along with all of their incoming and outgoing connections. In every Dropout layer, 30% of the neurons will be randomly deactivated during each training iteration.

2. *VGG Architecture:*

VGG19 is a convolutional neural network introduced by Simonyan et al. [11] [12] that consists of 19 layers, hence the name. The architecture consists of 16 convolutional layers, each followed by a rectified linear unit (ReLU) activation function, and 3 fully connected layers. One notable characteristic of VGG19 is its consistent use of small 3x3 convolutional filters throughout the network, with a stride of 1 and the same padding. This design choice allows the network to learn rich and diverse features from input images at multiple scales. Additionally, max-pooling layers with a 2x2 window and stride of 2 are interspersed between the convolutional layers, effectively reducing the spatial dimensions of the feature maps while preserving im- portant features. VGG19's architecture follows a pattern of repeated convolutional and pooling layers, leading to a deep feature hierarchy. This deep architecture enables the network to learn complex hierarchical representations of the input data, facilitating better discrimination between different classes.

3. *InceptionResNetV2:*

InceptionResNetV2 [13] is a varia- tion of convolutional neural network (CNN) architecture that combines elements from both the Inception and ResNet [4] architectures [14]. This architecture aims to achieve both high accuracy and computational efficiency in image classification tasks. It incorporates residual connections from the ResNet architecture, allowing for more efficient training of deep networks while mitigating the vanishing gradient problem. These residual connections facilitate the flow of gradients during backpropagation, enabling the training of very deep networks without encountering degradation in performance. It employs the Inception module, which consists of multiple parallel convolutional operations with different kernel sizes. This enables the network to capture features at various scales and resolutions, enhancing its ability to discriminate between different classes in complex images. It introduces additional architectural innovations such as batch normalization and factorized convolutions, which contribute to improved training stability and computational efficiency. The model has achieved state-of-the-art performance on various image classification benchmarks, including the ImageNet dataset, demonstrating its effectiveness in extracting meaningful features from images and accurately classifying them into different categories.

C. *Training of Models:*

The dataset was divided into training and validation sets in an 80:20 ratio, respectively. It's important to note that a sep- arate testing dataset without ground truth was also available. All models were trained over 25 epochs, utilizing images with dimensions of 300 by 300 pixels and a total of three channels.

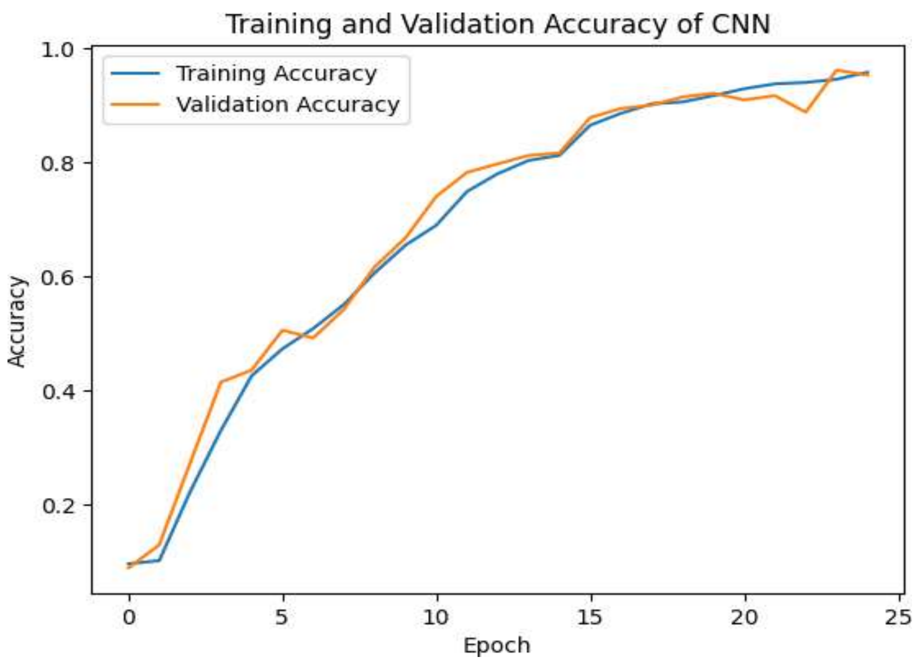

Fig. 3. Training and Validation Accuracy of CNN [9]

In Figure 3, we observe that the training accuracy and validation accuracy of our CNN model exhibit strong perfor- mance almost above 94%, with the validation accuracy closely matching the training accuracy.

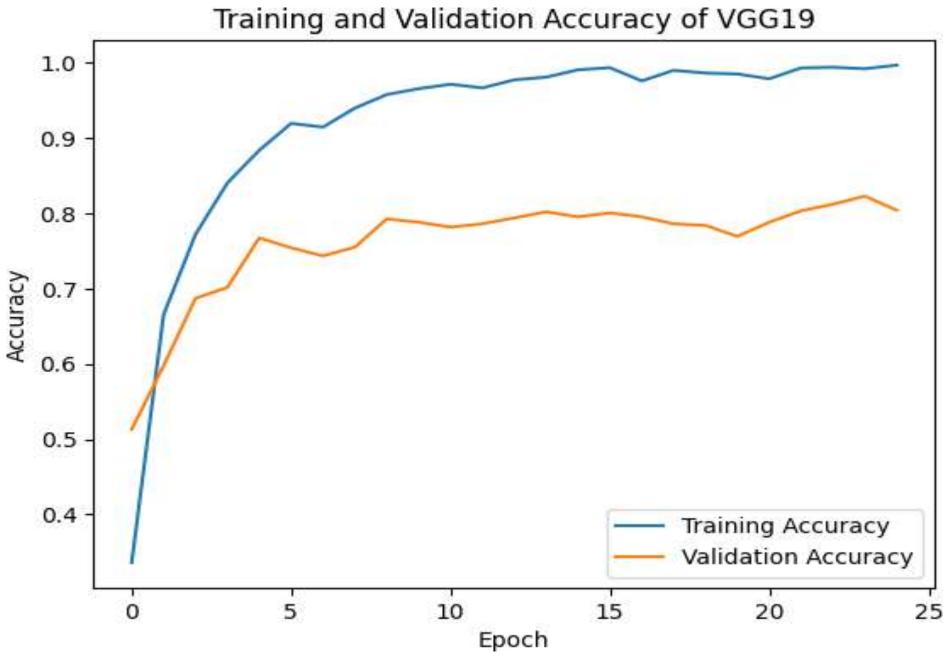

Fig. 4. Training and Validation Accuracy of VGG19 [12]

In Figure 4, during the training process of VGG19 , there is a noticeable disparity between the training accuracy and validation accuracy of the model was observed. While the training accuracy consistently remains above 95%, indicating a strong fit to the training data, the validation accuracy hovers around 80%.

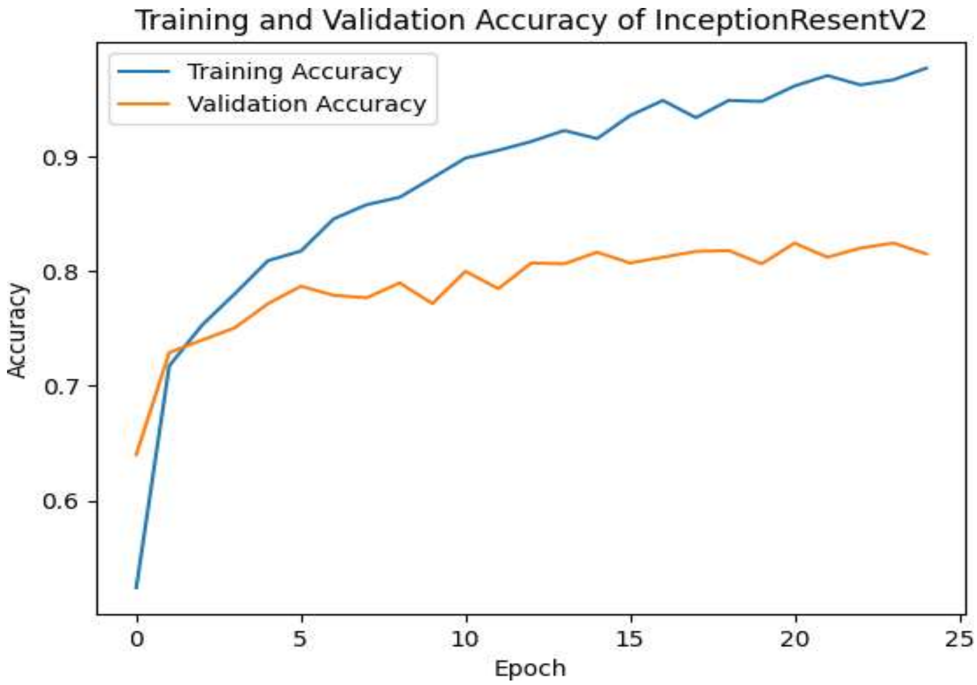

Fig. 5. Training and Validation Accuracy of Inceptionresnetv2 [13]

In Figure 4, during the training process of VGG19 , there is a noticeable disparity between the training accuracy and validation accuracy of the model was observed. While the training accuracy consistently remains above 95%, indicating a strong fit to the training data, the validation accuracy hovers around 80%.

Accuracy was reported as depicted in Figure 5, As the training progresses, there is a notable consistency in the validation accuracy, stabilizing at around 80% after a certain threshold. In contrast, the training accuracy steadily rises, reaching nearly 98% by the end of the training epochs.

Neither VGG19 nor InceptionResNetV2, both varients of CNN, attained the performance level of our CNN model after 25 epoches. These models exhibited lower accuracy and required further fine-tuning, potentially involving more epochs and adjustments to the learning rate, to effectively recognize pharmaceutical drugs and accurately classify them.

VGG19, InceptionResNetV2, and Convolutional Neural Networks (CNNs) [9] are advanced neural network architec- tures specifically crafted for image processing tasks. These models are particularly advantageous in pharmaceutical drug detection applications because they possess the ability to au- tonomously learn complex patterns from images. For instance, both VGG19 and InceptionResNetV2 comprise multiple layers that progressively extract hierarchical features from input images, enabling them to capture details like edges, textures, shapes, and patterns. This hierarchical feature learning capabil- ity is crucial for pill detection, given the significant variability in pill appearances due to factors such as color, size, shape, and texture.

A significant advantage of utilizing pre-trained CNN models such as VGG19 and InceptionResNetV2 lies in the concept of transfer learning. Initially trained on extensive image datasets like ImageNet [15], these models acquire the ability to rec- ognize diverse objects and patterns. By fine-tuning these pre-trained models with specific pharmaceutical drug images, we can capitalize on the knowledge acquired during the initial training phase. This strategy enables the models to achieve su- perior performance even when confronted with limited labeled data for the target task, a common scenario in pharmaceutical drug detection applications. CNNs are also equipped for object localization tasks, such as pinpointing the precise location of

pills within an image. Techniques like object detection can be integrated with CNNs to accurately localize pills, which proves beneficial for applications like pill counting or sorting, where precise pill identification is imperative.

CNNs are computationally efficient and capable of process- ing images rapidly, making them suitable for real-time or near- real-time pill detection applications. This efficiency ensures that the models can analyze large volumes of images effi- ciently, which is crucial for tasks that require quick decision-making, such as drug quality control in pharmaceutical man- ufacturing or pill identification in healthcare settings.

IV. EVALUATION

Given an input pill image to the models, we evaluated how accurately our model is identifying the corresponding pill class. To test the accuracy and different evaluation metrics of our trained models, models evaluated on the testing portion of the data set. For this study, models are evaluated based on standard metrics like Precision, Recall and F1 score [16]. The accuracy is the ratio between the correctly classified samples and the total number of samples in the evaluation dataset. This metric is among the most commonly used in applications of ML in medicine [16].

The accuracy can be calculated using the following formula:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (1)$$

where:

- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative

Using a similar approach, we can calculate the following metrics:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \quad (4)$$

For this study, CNN was able to achieve highest classification accuracy of 97.74%, VGG19 with 82.83% and Inception- ResNetV2 with 88.41%. We can find all the numerical results in below Table-I, includes Classification accuracy, Precision, Recall and F1 score. To our knowledge, there have not been other works that attempt to solve the pharmaceutical drug recognition and classification problem on this dataset.

Table 1. Evaluation Score model

	CNN	VGG19	Inception-ResNetV2
Accuracy	97.74%	80.83%	88.41%
Precision	97.75%	86.12%	88.80%
Recall	97.70%	82.79%	88.38%
F1 Score	97.72%	83.06%	88.35%

The performance evaluation of training accuracy across all three models is illustrated in Figure 6. Each model exhibited strong performance during the training phase, achieving accuracy levels surpassing 94%, indicative of effective learning from the training data. This high level of training accuracy underscores the models' ability to capture and learn from the patterns and features present in the training dataset, suggesting robust training processes across all models. Upon comparing the models among themselves, it becomes evident that the CNN model outperforms the others.

The validation accuracy is depicted in Figure 7, consistently surpasses the 80% mark across all models. This indicates robust performance in both training and validation tasks for all three models. Their ability to maintain high validation accuracy underscores their effectiveness in generalizing well to unseen data. Overall, these results highlight the strong performance and reliability of all three models in pharmaceutical drug classification tasks, demonstrating their potential utility in real-world applications.

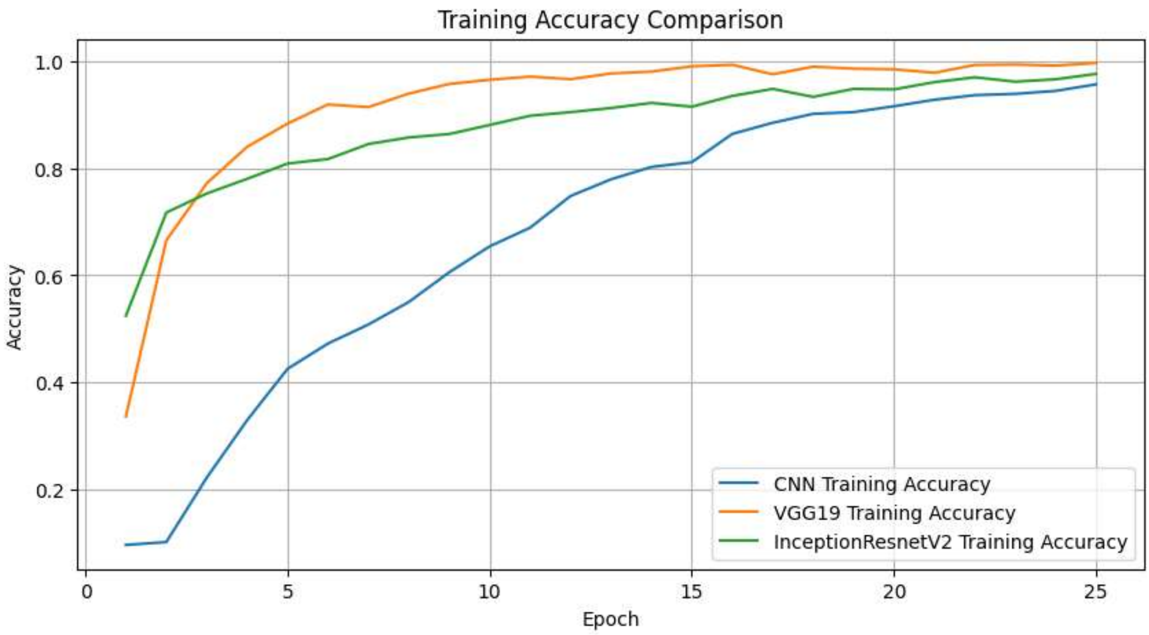

Fig. 7. Validation Accuracy of CNN, VGG19 and InceptionResnetv2

Fig. 8. Output on testing image with its correctly Predicted class

In the image depicted in Figure 8, the left side displays the test image alongside its predicted name, while the right side showcases the original image whose name was predicted by our model. Remarkably, our model demonstrates exceptional performance, accurately predicting the drug name with high precision. This proficient identification underscores the robustness and efficacy of our model in pharmaceutical drug recognition tasks. The side-by-side comparison of the predicted and original images vividly illustrates our model’s exceptional predictive prowess, confirming its dependability and accuracy in real-world situations. Table II summarizes the numerical accuracies achieved by some of the similar studies performed for drug image recognition task [5] [6] [4] [7].

Table 2. Summary Table of Existing work

Used DL model	Used dataset	Performance
ResNet50 [5]	NLM Pill Image	95.3%
MobileNet [5]		94.4%
Squeezenet [5]		83.2%
InceptionV3 [5]		94.8%
MLP [6]	NLM Pill Image	99.3% Average
Inception-Resnet [6]		
ResNet50 [7]	NLM Pill Image	90.8%
ResNet50 [4]	Unknown	91.80%

V. CONCLUSION

In this paper, we classified the medical pill images using Deep learning models like CNN, VGG19, and inception- resnetv2. The model was trained over an actual pill image data set and delivered excellent accuracy. Our models have accurately predicted pharmaceutical drugs, demonstrating their exceptional performance. This success underscores the effectiveness of our models in handling complex pharmaceutical datasets, reaffirming their overall strong performance. Future extension of this work can include the integration of a camera system (e.g. mobile phone), to provide input images to the model and the model can classify the drug. Additionally, expanding the range of pill types considered allows for a more diverse dataset to be utilized, enabling the model to classify a wider variety of drugs effectively.

REFERENCES

[1] Nagaprasad, S., Padmaja, D. L., Qureshi, Y., Bangare, S. L., Mishra, M., & Mazumdar, B. D. (2021). Investigating the impact of machine learning in pharmaceutical industry. *Journal of Pharmaceutical Research International*, *33*(46A), 6-14.

[2] Sharif Razavian, A., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN features off-the-shelf: an astounding baseline for recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 806-813).

[3] Wang, Y., Ribera, J., Liu, C., Yarlagadda, S., & Zhu, F. (2017, April). Pill recognition using minimal labeled data. In *2017 IEEE Third International Conference on Multimedia Big Data (BigMM)* (pp. 346-353). IEEE.

[4] Ou, Y. Y., Tsai, A. C., Wang, J. F., & Lin, J. (2018, October). Automatic drug pills detection based on convolution neural network. In *2018 International Conference on Orange Technologies (ICOT)* (pp. 1-4). IEEE.

[5] Larios Delgado, N., Usuyama, N., Hall, A. K., Hazen, R. J., Ma, M., Sahu, S., & Lundin, J. (2019). Fast and accurate medication identification. *NPJ digital medicine*, *2*(1), 10.

[6] Cordeiro, L. S., Lima, J. S., Ribeiro, A. I. R., Bezerra, F. N., Rebouças Filho, P. P., & Neto, A. R. R. (2019, October). Pill image classification using machine learning. In *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)* (pp. 556-561). IEEE.

[7] Al-Hussaeni, K., Karamitsos, I., Adewumi, E., & Amawi, R. M. (2023). CNN-Based Pill Image Recognition for Retrieval Systems. *Applied Sciences*, *13*(8), 5050.

[8] Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., ... & Ghayvat, H. (2021). CNN variants for computer vision: History, architecture, application, challenges and future scope. *Electronics*, *10*(20), 2470.

[9] Wu, J. (2017). Introduction to convolutional neural networks. *National Key Lab for Novel Software Technology. Nanjing University. China*, *5*(23), 495.

[10] Hossain, M. A., & Sajib, M. S. A. (2019). Classification of image using convolutional neural network (CNN). *Global Journal of Computer Science and Technology*, *19*(2), 13-14.

[11] Bansal, M., Kumar, M., Sachdeva, M., & Mittal, A. (2023). Transfer learning for image classification using VGG19: Caltech-101 image data set. *Journal of ambient intelligence and humanized computing*, 1-12.

[12] Dey, N., Zhang, Y. D., Rajinikanth, V., Pugalenthi, R., & Raja, N. S. M. (2021). Customized VGG19 architecture for pneumonia detection in chest X-rays. *Pattern Recognition Letters*, *143*, 67-74.

[13] Ou, Y. Y., Tsai, A. C., Zhou, X. P., & Wang, J. F. (2020). Automatic drug pills detection based on enhanced feature pyramid network and convolution neural networks. *IET Computer Vision*, *14*(1), 9-17.

[14] Elmannai, H., Hamdi, M., & AlGarni, A. (2021). Deep learning models combining for breast cancer histopathology image classification. *International Journal of Computational Intelligence Systems*, *14*(1), 1003.
[15] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84-90.
[16] Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. *Scientific reports*, *12*(1), 5979.
[17] Drug image dataset is openly accessible on Kaggle: https://www.kaggle.com/datasets/gauravduttakiit/pharmaceutical-drug-recognition